UNIVERSIDAD
DE GRANADA

Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación

MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS E INGENIERÍA DE COMPUTADORES

TRABAJO DE FIN DE MÁSTER

Procesos Gaussianos para problemas MIL. Aplicación a la detección de hemorragias intracraneales.

Presentado por:
Francisco Miguel Castro Macías

Tutores:
Rafael Molina Soriano
Departamento de Ciencias de la Computación e Inteligencia Artificial

Pablo Morales Álvarez
Departamento de Estadística e Investigación Operativa

Curso académico 2021-2022

Procesos Gaussianos para problemas MIL. Aplicación a la detección de hemorragias intracraneales.

Francisco Miguel Castro Macías

Francisco Miguel Castro Macías *Procesos Gaussianos para problemas MIL. Aplicación a la detección de hemorragias intracraneales.*.

Trabajo de fin de Máster. Curso académico 2021-2022.

Responsable de tutorización

Rafael Molina Soriano
Departamento de Ciencias de la Computación e Inteligencia Artificial

Pablo Morales Álvarez
Departamento de Estadística e Investigación Operativa

Máster Universitario en Ciencia de Datos e Ingeniería de Computadores

Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación

Universidad de Granada

Declaración de originalidad

D. Francisco Miguel Castro Macías

Declaro explícitamente que el trabajo presentado como Trabajo de Fin de Máster (TFM), correspondiente al curso académico 2021-2022, es original, entendida esta, en el sentido de que no ha utilizado para la elaboración del trabajo fuentes sin citarlas debidamente.

En Granada a 4 de septiembre de 2022

Fdo: Francisco Miguel Castro Macías

Firmado por CASTRO MACIAS FRANCISCO MIGUEL - ***9217** el día 04/09/2022 con un certificado emitido por AC FNMT Usuarios

A mis padres, por creer siempre en mí.

Índice general

Resumen

Un hematoma o hemorragia intracraneal (*IntraCranial Hemorrhage*, ICH) es una acumulación de sangre dentro del cráneo que provoca daños graves en el cerebro o incluso la muerte. Es esencial que su diagnóstico sea rápido y efectivo, ya que casi la mitad de las muertes se producen en las primeras 24 horas. Su detección se realiza a través de las tomografías computarizadas (*Computed Tomography*, CT) o TAC, que obtienen imágenes ordenadas del cráneo llamadas cortes, de forma que todas juntas componen un escáner global. Estudios recientes muestran que tras varias horas analizando estas pruebas, la fatiga lleva a los radiólogos a equivocarse al valorarlos.

Para crear métodos de triaje masivos y fiables, en la actualidad se están desarrollando Sistemas de Diagnóstico Asistido por Ordenador (Computer Aided Diagnosis Systems, CADSs) mediante técnicas de Aprendizaje Automático, lo cual requiere grandes bases de datos etiquetadas. Anotar cada uno de los cortes de un TAC es una tarea muy costosa, por lo que es impensable solicitar el etiquetado de un gran volumen de ellos. Sin embargo, las etiquetas a nivel del escáner completo se almacenan de forma automática en el historial médico, por lo que obtenerlas no requiere ningún esfuerzo. ¿Es posible desarrollar modelos de diagnóstico usando sólo las etiquetas globales (a nivel del escáner completo) en lugar de las etiquetas de cada corte? Este es el reto al que nos enfrentamos en este trabajo.

Nuestro objetivo es tratar la detección de hemorragias intracraneales como un problema de Aprendizaje a partir de Múltiples Instancias (*Multiple Instance Learning*, MIL). Los ejemplos de entrenamiento, llamados instancias, se encuentran particionados en bolsas de instancias, y sólo conocemos la etiqueta global de cada bolsa. La etiqueta de cada instancia es desconocida y tiene asociada una gran incertidumbre. En nuestro caso, cada corte de un escáner es una instancia, y el escáner completo es la bolsa. Un escáner se clasifica como positivo si al menos uno de sus cortes es positivo (contiene evidencias de la lesión), y negativo en caso contrario.

Los Procesos Gaussianos (*Gaussian Processes*, GPs) son la herramienta que vamos a emplear para tal fin. Son modelos bayesianos que permiten aproximar funciones complejas y cuya formulación probabilística permite modelar la incertidumbre presente en los problemas de tipo MIL. El modelo *Variational Gaussian Process Multiple Instance Learning* (VGPMIL) es una de las aproximaciones más interesantes de la literatura para este problema. Emplea la *desigualdad de Jaakkola* para calcular distribuciones variacionales que estiman la probabilidad de que una bolsa o una instancia sean positivos, permitiendo obtener expresiones analíticas, pero puede conducir a soluciones subóptimas.

En este trabajo se realizan tres contribuciones dentro de esta línea de investigación.

1. La primera de ellas recibe el nombre de *Pólya-Gamma Variational Gaussian Process Multiple Instance Learning* (PG-VGPMIL) y propone, por primera vez en la literatura, el uso de variables aleatorias Pólya-Gamma en el contexto de MIL. En este trabajo se prueba su equivalencia, en el sentido *Mean Field*, con VGPMIL.

2. La segunda recibe el nombre de *Gamma Variational Gaussian Process Multiple Instance Learning* (G-VGPMIL) y propone, por primera vez en la literatura, sustituir las variables Pólya-Gamma por variables Gamma, siendo validado experimentalmente en tres problemas distintos.

3. La tercera contribución, en este caso de tipo práctico, consiste en el uso de G-VGPMIL junto con redes neuronales y mecanismos de atención para resolver el problema de la detección de hemorragias intracraneales. El modelo G-VGPMIL obtiene los mejores resultados de la literatura usando sólo etiquetas a nivel de escáner completo, lo cual es muy prometedor y establece el camino a seguir para desplegar esta solución en centros médicos.

Palabras clave: Hemorragia intracraneal, Tomografía Computarizada, Aprendizaje a partir de Múltiples Instancias, Procesos Gaussianos, Inferencia Variacional, desigualdad de Jaakkola, Pólya-Gamma, Gamma, Aprendizaje Automático, Aprendizaje Profundo, Redes Neuronales, Atención.

English summary

An IntraCranial Hemorrhage (ICH) is a collection of blood inside the skull that results in severe brain damage or even death. Early and effective diagnosis is essential, as almost half of the deaths occur within the first 24 hours. Its detection is done through computed tomography (CT) scans, which obtain several ordered images of the skull called slices, so that all together they make up an overall CT scan. Recents studies show that after several hours analyzing these tests, fatigue leads radiologists to make mistakes when evaluating them.

To create massive and reliable triage methods, Computer Aided Diagnosis Systems (CADSs) are currently being developed using Machine Learning techniques, which requires large labeled databases. Annotating every single CT slice is a very expensive task, so it is not possible to request the labeling of a large volume of CT slices. However, the label of the entire scan is automatically stored in the medical record, so obtaining them is effortless. Is it possible to develop diagnostic models using only the scan-level labels instead of the slice-level labels? This is the challenge we face in this work.

Our goal is to treat ICH detection as a Multiple Instance Learning (MIL) problem. The training examples, called instances, are partitioned into bags of instances, and we only know the global label of each bag. The label of each instance is unknown and has a large uncertainty associated with it. In our case, each slice of a scan is an instance, and the entire scan is the bag. A scan is classified as positive if at least one of its slices is positive (contains evidence of the lesion), and negative otherwise.

Gaussian Processes (GPs) are the tool we are going to use for this purpose. They are Bayesian models that allow us to approximate complex functions and whose probabilistic formulation allows us to model the uncertainty present in MIL type problems. The Variational Gaussian Process Multiple Instance Learning (VGPMIL) model is one of the most interesting approaches in the literature to solve this problem. It uses Jaakkola's inequality to compute variational distributions that estimate the probability that a bag or instance will be positive, which allows analytical expressions to be obtained, but may lead to suboptimal solutions.

This paper makes three contributions within this line of research.

1. The first one is called *Pólya-Gamma Variational Gaussian Process Multiple Instance Learning* (PG-VGPMIL) and proposes, for the first time in the literature, the use of Pólya-Gamma random variables in the context of MIL. In this paper we prove its equivalence, in the sense of *Mean Field*, with VGPMIL.

2. The second one is called *Gamma Variational Gaussian Variational Process Multiple Instance Learning* (G-VGPMIL) and proposes, for the first time in the literature, to replace the Pólya-Gamma variables by Gamma variables, being validated experimentally in three different problems.

3. The third one consists of the use of G-VGPMIL together with neural networks and attention mechanisms to solve the problem of ICH detection. G-VGPMIL model obtains the best results in the literature using only labels at scan-level, which is very promising and sets the way forward for deploying this solution in medical centers.

Keywords: Intracranial hemorrhage, Computed Tomography, Multiple Instance Learning, Gaussian Processes, Variational Inference, Jaakkola Inequality, Pólya-Gamma, Gamma, Machine Learning, Deep Learning, Neural Networks, Attention.

1. Introducción y objetivos

El objetivo de este capítulo es ofrecer una visión general del propósito del trabajo y aclarar los objetivos inicialmente planteados y el grado en que estos se han cumplido. Primero ofrecemos una introducción sobre los problemas que tratamos y después planteamos los objetivos, conectando cada uno con la parte de la memoria que lo cubre.

1.1. Introducción

Un hematoma o hemorragia intracraneal (*IntraCranial Hemorrhage*, ICH) es una acumulación de sangre dentro del cráneo que normalmente se debe a la rotura de un vaso sanguíneo [35]. La sangre que entra al cráneo provoca que aumente la presión cerebral, lo que conduce a daños graves en el cerebro o incluso a la muerte. Alrededor del 40 % de los pacientes fallecen durante el primer mes, la mayoría en los dos primeros días, y sólo un 20 % son independientes al cabo de 6 meses [16]. De hecho, casi el 30 % de las muertes se producen en las primeras 24 horas [40], por lo que el diagnóstico rápido y eficaz de esta lesión es esencial.

El diagnóstico se suele realizar a través de las tomografías computarizadas (*Computed Tomography*, CT) o TAC [31]. Esta técnica obtiene varios escáneres del cráneo llamados cortes, de forma que todos juntos componen un escáner global. Sin embargo, diversos estudios muestran que tras varias horas analizando estas pruebas, la fatiga lleva a los radiólogos a equivocarse en el diagnóstico [41]. Para conseguir mejores métodos de triaje se están desarrollando Sistemas de Diagnóstico Asistido por Ordenador (Computer Aided Diagnosis Systems, CADSs) basados en técnicas de Aprendizaje Automático (*Machine Learning*, ML) [8].

Muchas de las soluciones que se están proponiendo se basan en modelos de Aprendizaje Profundo (*Deep Learning*, DL) [18] que son entrenados de forma directa, usando las etiquetas de todos los cortes. Sin embargo, para que los modelos sean fiables y obtengan buenos resultados, necesitamos grandes base de datos de cortes de TAC etiquetados. Anotar cada uno de los cortes de un TAC es una tarea muy costosa, por lo que, debido al escaso personal de radiología y al poco tiempo del que disponen, es impensable solicitar el etiquetado de un gran volumen de datos. Sin embargo, las etiquetas a nivel del escáner completo se almacenan de forma automática en el historial médico al diagnosticar a los pacientes, por lo que obtenerlas no requiere ningún esfuerzo adicional. La solución pasa por desarrollar modelos que usen sólo las etiquetas globales (a nivel del escáner completo) en lugar de las etiquetas de cada corte.

En este trabajo tratamos el problema de la detección de hemorragias intracraneales como un problema de Aprendizaje a partir de Múltiples Instancias (*Multiple Instance Learning*, MIL) [50]. Este es una variación del aprendizaje supervisado clásico que alivia el esfuerzo de anotación requerido. Los ejemplos de entrenamiento, llamados instancias, se encuentran particionados en bolsas de instancias, y sólo conocemos la etiqueta global de cada bolsa. La etiqueta de cada instancia es desconocida y tiene asociada una gran incertidumbre. En el problema al que nos enfrentamos cada corte de un escáner es una instancia, y el escáner completo es la bolsa. Un escáner se clasifica como positivo si al menos uno de sus cortes es positivo, y negativo en caso contrario.

Para construir una solución, nos serviremos de los Procesos Gaussianos (*Gaussian Processes*, GPs) [34]. Son modelos bayesianos que permiten aproximar funciones complejas y cuya formulación probabilística permite modelar la incertidumbre presente en los problemas de tipo MIL. El modelo *Variational Gaussian Process Multiple Instance Learning* (VGPMIL) [19] es una de las aproximaciones más fructíferas de la literatura. Utiliza el modelo de observación logístico y destaca por su forma de tratar la relación entre las etiquetas de las instancias y las bolsas. Aproxima la distribución a posteriori usando inferencia variacional, para lo que emplea la *desigualdad de Jaakkola* [22]. Esto le permite obtener expresiones analíticas para las actualizaciones de los parámetros, aunque puede conducir a soluciones subóptimas.

En este trabajo estamos interesados en el uso de técnicas de *data augmentation*[1] para realizar inferencia variacional en estos escenarios. Motivados por el buen comportamiento de las variables aleatorias Pólya-Gamma en problemas de clasificación clásicos [51], nos preguntamos qué ocurre al introducir estas variables en un modelo de GPs para MIL. La respuesta a esta pregunta es la primera contribución del trabajo, que recibe el nombre de *Pólya-Gamma Variational Gaussian Process Multiple Instance Learning* (PG-VGPMIL). Uno de los resultados más importantes de este trabajo es que es equivalente, en sentido *Mean Field*, a VGPMIL. Por otra parte, la sustitución de variables Pólya-Gamma por variables Gamma conduce al nuevo modelo *Gamma Variational Gaussian Process Multiple Instance Learning* (G-VGPMIL), la segunda contribución del trabajo.

Este último modelo es validado experimentalmente en tres problemas distintos antes de proponer una solución definitiva al problema de la detección de hemorragias intracraneales. Siguiendo la metodología de [53], proponemos una arquitectura que combina redes neuronales con mecanismos de atención para extraer características, junto con G-VGPMIL para realizar predicciones a nivel de corte y de escáner completo. Los resultados obtenidos se comparan con las aproximaciones existentes en la literatura.

1.2. Objetivos

Cuando se propuso el trabajo, inicialmente se plantearon los siguientes objetivos:

A. Estudiar los modelos basados en GPs y comprender los problemas derivados de utilizar datos masivos y qué soluciones existen.

B. Estudiar los problemas tipo MIL y cómo los GPs se pueden usar en ellos.

C. Desarrollar un modelo basado en GPs para detectar hemorragias intracraneales a partir de tomografías computarizadas.

Todos han sido completamente satisfechos, como justificamos a continuación.

- El Capítulo 2 cubre la primera parte del objetivo B. En él se recoge una definición general del concepto de MIL [50], así como un estudio de las hipótesis existentes para modelar la relación entre las instancias y las bolsas. También repasamos los esfuerzos realizados en la literatura para encontrar una solución. Por último, ofrecemos una visión general de las tareas o aplicaciones en las que estos problemas están presentes.

[1] En el contexto de este trabajo, *data augmentation* se refiere a una técnica consistente en añadir variables aleatorias para aumentar la probabilidad conjunta del modelo. Marginalizando estas variables recuperamos la distribución original. En nuestro caso, condicionar sobre estas variables nos permite calcular analíticamente la esperanza en el modelo de observación logístico.

- El estudio de GPs se aborda con total rigurosidad en el Capítulo 3. En él motivamos los GPs y presentamos su formulación en problemas de regresión y clasificación [34]. El cálculo de la distribución predictiva conlleva un coste computacional muy elevado, de donde se derivan todos los problemas de usar grandes volúmenes de datos con estos modelos. La aproximación más exitosa para solventar este asunto recibe el nombre de *Sparse GPs* y se estudia en profundidad en la Sección 3.5. Con esto, el objetivo A queda totalmente cubierto.

- La segunda parte del objetivo B se completa en el Capítulo 4. En él usamos a los GPs como herramienta para resolver los problemas de MIL a los que nos enfrentamos. Este es un capítulo muy importante ya que en él se presentan de forma teórica dos de nuestras contribuciones.

  1. Empezamos estudiando el trabajo VGPMIL [19], que es uno de los modelos que componen el estado del arte a la hora de usar GPs para resolver problemas de MIL. Como ya hemos aclarado, VGPMIL emplea la desigualdad de Jaakkola [22], que le permite llegar a expresiones analíticas de las mismas.
  2. Motivados por su buen comportamiento en problemas de clasificación clásicos [51], proponemos por primera vez el uso de variables Pólya-Gamma para poder realizar inferencia. Como resultado obtenemos PG-VGPMIL, del cual probamos su equivalencia, en el sentido *Mean Field*, con VGPMIL.
  3. La definición de las variables Pólya-Gamma nos conduce a considerar variables Gamma, obteniendo el nuevo modelo G-VGPMIL, que usa variables Gamma en lugar de Pólya-Gamma.

- El Capítulo 5 es el más práctico. En él se valida experimentalmente el nuevo modelo G-VGPMIL, comparándolo con VGPMIL en tres problemas distintos. El objetivo C se corresponde con la Sección 5.4, en la que presentamos una arquitectura inspirada en [53] para la detección de hemorragias intracraneales. Además, se hace un estudio exhaustivo de los resultados obtenidos.

2. Aprendizaje a partir de Múltiples Instancias

En este capítulo se ofrece una introducción general al concepto de *Aprendizaje a partir de Múltiples Instancias* (*Multiple Instance Learning*), el cual juega un papel clave en este trabajo. Empezaremos motivando este concepto y ofreciendo la intuición detrás del mismo. Definiremos rigurosamente qué es un problema de este tipo, revisaremos las aproximaciones más populares para resolverlo y señalaremos algunas de las aplicaciones en las que se emplea.

2.1. Motivación y orígenes

Uno de los mayores inconvenientes del aprendizaje supervisado es que no siempre podemos disponer de suficientes ejemplos etiquetados con los que entrenar un modelo. En muchos ámbitos, como por ejemplo la medicina, el proceso de etiquetado puede ser costoso y lento, por lo que es necesario explorar alternativas que alivien esta necesidad de anotación masiva.

El *aprendizaje a partir de múltiples instancias* (*Multiple Instance Learning*, *MIL*) es una variación del aprendizaje supervisado clásico que surge para dar respuesta a este problema. En lugar de etiquetar cada una de las instancias por separado, se etiquetan por grupos o *bolsas*, de forma que se ofrece una sóla etiqueta para cada bolsa. A partir de esta información, se pretende aprender a etiquetar correctamente nuevas instancias y bolsas. De esta forma, el esfuerzo de anotación se reduce notablemente.

El término MIL fue usado por primera vez en [10] mientras se investigaba un problema relacionado con el desarrollo de nuevos fármacos. Se pretendía predecir si una molécula de un fármaco podría o no vincularse a una proteína relacionada con una enfermedad. Se disponían de ejemplos de moléculas que sí consiguen vincularse y de moléculas que no lo consiguen. El factor más importante para que esto ocurra es la forma (como ocurre con una llave y una cerradura). Así pues, cada bolsa representa a una molécula, y las instancias dentro de ella son todas las formas que puede adoptar. Una etiqueta positiva en la bolsa quiere decir que existe una forma con la que la molécula se adhiere a la proteína, mientras que una etiqueta negativa quiere decir que no hay ninguna forma con la que ocurra esto.

El problema al que nos enfrentamos en este trabajo encaja de forma natural en este paradigma. Cada escáner o tomografía completa es una bolsa de instancias. Una instancia se corresponde con un corte. El experto (en este caso un especialista médico) etiqueta el escáner completo, lo cual consiste en determinar si hay presencia o no de hemorragia. De forma natural, si un paciente sufre una hemorragia, al menos uno de los cortes así lo evidenciará. Obsérvese que el etiquetado del escáner completo (bolsa) es mucho más fácil que el de los cortes y que, además, la información sobre la presencia o no de hemorragia puede encontrarse en el historial clínico del paciente.

2.2. Definición formal

Presentamos a continuación la definición de MIL proporcionada en [50]. Denotamos por $\mathcal{X}$ al espacio de instancias y por $\mathcal{T}$ al espacio de etiquetas de las bolsas, respectivamente. Una *bolsa*

es un elemento de $\mathcal{P}(\mathcal{X})$[1]. Un *concepto* es una función ideal

$$v\colon \mathcal{P}(\mathcal{X}) \to \mathcal{T},$$

que asigna una etiqueta a cada bolsa. El objetivo de un problema MIL es aproximar esta función, en caso de que exista. Para ello, dispondremos de un conjunto de entrenamiento finito compuesto por bolsas etiquetadas, es decir, pares de la forma $(B, T) \in \mathcal{P}(\mathcal{X}) \times \mathcal{T}$. Las bolsas tienen intersección vacía y se supone que $T = v(B)$.

En la mayoría de problemas de MIL las instancias también tienen una etiqueta, aunque esta es desconocida. Es el caso de la detección de hemorragias intracráneales a partir de escáneres de TAC, donde cada corte puede presentar o no evidencias de la enfermedad, pero sólo conocemos la etiqueta del escáner completo. En estos casos existe un espacio $\mathcal{Y}$ con las etiquetas de las instancias y cada instancia x tiene una etiqueta $y_x \in \mathcal{Y}$. Diremos que una bolsa B contiene un *concepto* $c \in \mathcal{Y}$ si existe $x \in B$ tal que $y_x = c$.

2.2.1. Algunos tipos de hipótesis

La forma en que se relacionan las bolsas y las instancias individuales se denomina *supuesto*, *suposición* o *hipótesis* y es fundamental en el problema. Vamos a enumerar algunas de las más populares.

Hipótesis normal. La hipótesis más común, y con la que nosotros trabajaremos en el problema al que nos enfrentamos, ocurre en problemas de clasificación binarios, en los que $\mathcal{Y} = \mathcal{T} = \{0, 1\}$. Consiste en tomar $T = \max_{x \in B} \{y_x\}$. Es decir, una bolsa se etiqueta como positiva si y sólamente si al menos una de las instancias tiene etiqueta positiva. Recordemos que la etiqueta de las instancias es desconocida.

Jerarquía de Weidman et al. En [50] se propusieron tres paradigmas de complejidad incremental para modelar problemas de clasificación en los que $\mathcal{T} = \{0, 1\}$ y el espacio $\mathcal{Y}$ es discreto.

- **Hipótesis basada en la presencia de instancias.** Existe un subconjunto de conceptos requeridos $\mathcal{C} \subset \mathcal{Y}$. Una bolsa se etiqueta como positiva si y sólo si contiene al menos uno de estos conceptos. Esto es,

$$T = \begin{cases} 1 & \text{si } |\{y_x\colon x \in B\} \cap \mathcal{C}| \geq 1, \\ 0 & \text{en otro caso} \end{cases}$$

  Tomando $\mathcal{C} = \{1\}$ estamos en la hipótesis normal.

- **Hipótesis basada en un umbral.** Como en el anterior, se define un subconjunto de conceptos requeridos $\mathcal{C} \subset \mathcal{Y}$ y un número entero positivo n. Una bolsa B se etiqueta como positiva si y sólo si contiene al menos n de estos conceptos. Es decir,

$$T = \begin{cases} 1 & \text{si } |\{y_x\colon x \in B\} \cap \mathcal{C}| \geq n, \\ 0 & \text{en otro caso} \end{cases}$$

  Tomando $n = 1$ y $\mathcal{C} = \{1\}$ estamos en el caso anterior.

[1] $\mathcal{P}(\mathcal{X})$ denota las partes de $\mathcal{X}$, es decir, el conjunto de todos los subconjuntos de $\mathcal{X}$.

- **Hipótesis basada en dos umbrales.** Nuevamente existe un subconjunto de conceptos requeridos $\mathcal{C} \subset \mathcal{Y}$. Se definen dos enteros positivos n_1 y n_2, con $n_1 < n_2$. Una bolsa B se etiqueta como positiva si y sólo si contiene al menos n_1 y como máximo n_2 de estos conceptos. Esto es,

$$T = \begin{cases} 1 & \text{si } n_1 \leq |\{y_x \colon x \in B\} \cap \mathcal{C}| \leq n_2, \\ 0 & \text{en otro caso} \end{cases}$$

  Tomando n_2 suficientemente grande se tiene el caso anterior.

Hipótesis colectiva. En lugar de ver a las bolsas como subconjuntos fijos, se modela una bolsa B como una distribución de probabilidad con densidad $p(x \mid B)$. También se modelan las etiquetas de las instancias y como una distribución con densidad $p(y \mid x)$. El objetivo es modelar la distribución $p(y \mid B) = \int_{\mathcal{X}} p(y \mid x)p(x \mid B)\mathrm{d}x$.

2.3. Aproximaciones al problema

A partir de ahora vamos a centrarnos en problemas de clasificación binarios donde se cumple la hipótesis normal. En ellos disponemos de dos tipos de etiquetas (positiva o negativa), tanto a nivel de bolsa como de instancia. La presencia o no de una instancia positiva determina la etiqueta de la bolsa. A continuación, usando [23], revisamos algunas de las soluciones que se han propuesto para solventar los problemas MIL de este tipo.

Las primeras aproximaciones se centraron en diseñar algoritmos a medida para el dominio del problema que trataban. Dos ejemplos de ello son las soluciones que se presentan en [10] y en [26]. Posteriormente se comenzaron a diseñar técnicas que pretendían predecir las etiquetas de las bolsas empleando núcleos o *kernels* que medían la similitud entre las bolsas. También se comenzó a tratar como una aproximación para modelar el problema de la falta de etiquetas o la presencia de ruido en ellas. Surgieron así soluciones que modifican la formulación original de los modelos SVM (*Support Vector Machine*) para adaptarlos a MIL, como por ejemplo mi-SVM [1]. Sin embargo, requieren técnicas de optimización costosas que impiden usarlos con conjuntos grandes de datos.

Una de las tendencias más fructíferas consiste en encontrar el ejemplo "más positivo"dentro de una bolsa positiva, considerando que este es el responsable de la etiqueta de la bolsa. A este ejemplo se le suele denominar "testigo". MI-SVM [1] adapta la formulación de SVM con el objetivo de maximizar la distancia de tal instancia al hiperplano separador. En MICA [25] se emplea programación lineal y se construye un "testigo" como una combinación convexa de los ejemplos de una bolsa. En [58] se generaliza el trabajo de [26] incluyendo esta idea.

En los últimos años, muchos de los trabajos se centran en adaptar modelos de Aprendizaje Profundo (*Deep Learning*, *DL*). Este tipo de modelos son muy populares en todos los problemas de aprendizaje debido a su alta flexibilidad y su capacidad para aproximar funciones complejas. Por ello, los esfuerzos se centran en encontrar nuevas arquitecturas y mecanismos que permitan entrenar redes neuronales complejas usando bolsas de instancias [54, 48]. Dentro de esta línea, la aproximación que constituye el estado del arte emplea mecanismos de atención [45] para ponderar cada una de las instancias [20].

Junto a estos mecanismos de atención, los modelos probabilísticos constituyen el estado del arte para MIL. Los Procesos Gaussianos (*Gaussian Processes*, *GPs*) [34] son modelos bayesianos

que aproximan funciones muy complejas a la vez que otorgan una estimación de la incertidumbre presente en las predicciones. Su formulación se puede adaptar a los problemas MIL para diseñar potentes modelos [23, 19, 47] que posteriormente se combinan con mecanismos de atención, y que constituyen el estado del arte en los problemas MIL.

2.4. Algunas aplicaciones

El enfoque MIL es muy potente ya que reduce drásticamente la necesidad de datos completamente etiquetados. Por ello se ha intentado aplicar en diversos campos, algunos de los cuales comentamos a continuación. Puede consultarse más información en [7].

Biología y química. A raíz de la ya comentada tarea de clasificación de moléculas han surgido numerosos problemas del ámbito de la biología en los que se puede aplicar este enfoque. Las entidades biológicas complejas, como compuestos, moléculas o genes, son modeladas como bolsas cuyas partes pueden inducir un efecto u otro. Es mucho más difícil observar cada una de las partes de estas entidades que observarlas como un conjunto y ofrecer una sóla etiqueta. Algunos ejemplos concretos son la predicción de funciones de genes [13], la predicción de la afinidad de péptidos [56] (ejemplo de un problema de regresión) y el descubrimiento de lugares de enlace en la expresión de nuevos genes [3].

Visión por computador. En este campo el enfoque MIL permite caracterizar patrones complejos a partir de otros más simples que los componen. Además, permite entrenar modelos complejos reduciendo el esfuerzo de anotación, lo que lo está convirtiendo en un paradigma popular dentro del diagnóstico asistido por ordenador.

- La *recuperación de imágenes basada en el contenido* (*Content Based Image Retrieval, CBIR*) [24] pretende categorizar las imágenes basándose en los objetos y conceptos que contiene. La imagen completa se considera como una bolsa que dentro contiene diversos objetos, cuya localización exacta no es importante [58]. Así, las imágenes se particionan mediante alguna técnica (parches, *keypoints*, segmentación, ...) para conformar las instancias [49]. En la Figura 2.1 se muestra una imagen de un oso que se divide en varias regiones. De estas regiones sólo la que contiene al animal tiene etiqueta positiva, mientras que las demás tienen etiqueta negativa.

- En la *localización de objetos* [43] y *segmentación* [15] típicamente se han requerido imágenes fuertemente anotadas. Es decir, en el caso de la localización, era necesario dar la posición de una caja que englobase al objeto, mientras que en segmentación se necesitaban etiquetas a nivel de píxel. Estas tareas requieren un tiempo y esfuerzo considerables, por lo que la investigación se centra actualmente usar imágenes débilmente etiquetadas [57]. La aparición de MIL ha conseguido aliviar estos requerimientos y la comunidad investigadora pretende desarrollar soluciones para estos problemas usando sólo etiquetas a nivel de la imagen completa [2, 46].

- En tareas de *diagnóstico asistido por computador* (*Computer Aided Diagnosis, CAD*) [11] el paradigma MIL está adquiriendo una grandísima popularidad. Hay que tener en cuenta que el etiquetado de imágenes médicas requiere un grandísimo esfuerzo por parte de los médicos, por lo que usar soluciones basadas en imágenes fuertemente etiquetadas es impracticable. El enfoque MIL es muy apropiado en este tipo de situaciones ya que en

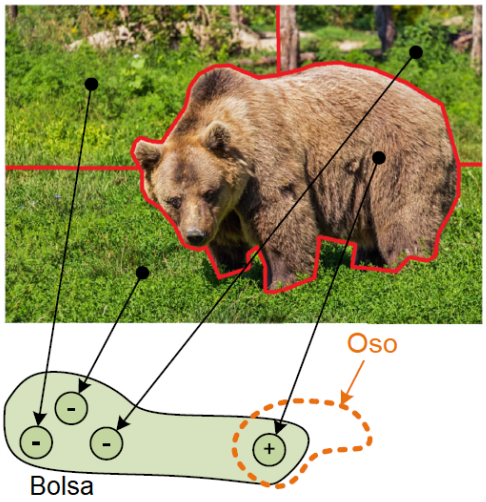


Figura 2.1.: La imagen se segmenta para conformar las instancias. Cada una de estas instancias tiene una etiqueta oculta. La etiqueta de la bolsa (imagen completa) indica la presencia de un oso. Imagen adaptada de [7].

un paciente existen regiones anormales y sanas, mientras que en una persona sana sólo existen regiones sanas. Algunas aplicaciones concretas son la clasificación de imágenes histológicas [36] o la detección de demencia a partir de resonancias magnéticas del cerebro [44].

Como vemos, el paradigma MIL no sólo sirve para formular nuevos problemas de aprendizaje, sino también para reformular los problemas ya existentes de una forma alternativa y natural. La definición de un problema MIL sólo exige el etiquetado de grupos de instancias, lo cual reduce el esfuerzo de anotación y permite utilizar conjuntos de datos débilmente anotados.

En este capítulo hemos presentado qué son los problemas MIL y la motivación detrás de ellos. Como ya hemos señalado, el objetivo de este trabajo es resolver un problema concreto de este tipo. Los siguientes capítulos preparan el camino para ello. En particular, el capítulo (Capítulo 3) presenta los Procesos Gaussianos, que será la herramienta que emplearemos para hacer frente al problema. Tras haberlos comprendido, podremos adaptarlos a los problemas MIL en el Capítulo 4, en donde señalaremos nuestras contribuciones al estado del arte.

3. Procesos Gaussianos

Un Proceso Gaussiano (*Gaussian Process*, *GP*) puede entenderse como una distribución de probabilidad normal sobre el espacio infinito dimensional de todas las funciones [34]. Numerosos problemas dentro del aprendizaje automático y estadístico consisten en aproximar funciones desconocidas. Para ello, es necesario realizar alguna suposición *a priori* sobre la naturaleza de la función. Estas suposiciones se codifican en un GP a priori que, tras un procedimiento de optimización, dará lugar a un proceso *a posteriori* que se ajusta a los datos [9].

El lector puede preguntarse, de forma natural, por qué hemos decidido usar GPs para solucionar los problemas MIL. Un aspecto clave en el escenario MIL es la gran incertidumbre presente en las etiquetas de las instancias. Por ello, los métodos bayesianos son una buena elección ya que permiten tenerla en cuenta. Como muestra este capítulo y el siguiente, los GPs poseen una formulación probabilística en la que la hipótesis de los problemas MIL se puede incluir directamente y que permite cuantificar esa incertidumbre.

En este capítulo presentamos formalmente los GPs y aquellos elementos teóricos necesarios para modelizar el problema al que nos enfrentamos. Comenzaremos motivando y definiendo los GPs para después analizar su uso en problemas de regresión y clasificación. Posteriormente presentaremos una aproximación conocida como *Procesos Gaussianos Dispersos* (*Sparse GPs*, *SGPs*) y que posibilita el uso de GPs con grandes conjuntos de datos. Este capítulo está construido en torno al contenido de [34, 9, 4].

3.1. De distribuciones gaussianas a procesos gaussianos

Consideremos una colección de variables aleatorias reales $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n]$ tales que su distribución conjunta es una normal multivariante, también llamada distribución gaussiana multivariante. Una importante propiedad es que la marginalización de $\boldsymbol{X}$ a cualquier subconjunto de variables es también una distribución gaussiana. Nos hacemos la siguiente pregunta: ¿y si $\boldsymbol{X}$ fuese la marginalización de una colección (posiblemente infinita) de variables? Surge así la siguiente definición:

Definición 3.1 (Proceso Gaussiano, GP)**.** Un Proceso Gaussiano (GP) es una colección de variables aleatorias tales que cualquier subconjunto finito suyo sigue una distribución gaussiana. Si f es un GP en $\mathcal{X}$, escribiremos $f \sim \mathcal{GP}(m, k)$, donde

$$\begin{aligned} m\colon \mathcal{X} \to \mathbb{R}, \quad & m(x) = \mathbb{E}\left[f(x)\right], \\ k\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}, \quad & k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \mathbb{E}\left[\left(f(\boldsymbol{x}_1) - m(\boldsymbol{x}_1)\right)\left(f(\boldsymbol{x}_2) - m(\boldsymbol{x}_2)\right)\right]. \end{aligned}$$

Las aplicaciones m y k reciben el nombre de *media* y *covarianza* o *núcleo*, respectivamente.

Observación 3.1. De la definición deducimos que un GP induce una distribución de probabilidad sobre el espacio de todas las funciones. Esto es, si $f\colon \mathbb{R}^D \to \mathbb{R}$ es tal que $f \sim \mathcal{GP}(m, k)$ y $\boldsymbol{X} = \left[\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_N^\top\right] \in \mathbb{R}^{D \times N}$ entonces $\boldsymbol{f} = f(\boldsymbol{X}) = [f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)]^T \in \mathbb{R}^N$ sigue una distribución gaussiana de media $m(\boldsymbol{X}) = [m(\boldsymbol{x}_1), \ldots, m(\boldsymbol{x}_N)]^\top$ y matriz de covarianza

$(k(\boldsymbol{X},\boldsymbol{X}))_{ij} = k(\boldsymbol{X},\boldsymbol{X})_{ij} = k(\boldsymbol{x}_i,\boldsymbol{x}_j)$. Es común escribir $\boldsymbol{f} \sim \mathcal{N}(m(\boldsymbol{X}), k(\boldsymbol{X},\boldsymbol{X}))$, omitiendo así la dependencia del conjunto de entrenamiento.

Como veremos a lo largo de este trabajo, no perdemos generalidad al elegir $m(\boldsymbol{x}) = 0$ para cada $\boldsymbol{x} \in \mathcal{X}$. El núcleo suele elegirse de entre una familia paramétrica de los mismos. Los parámetros del núcleo reciben el nombre de *hiperparámetros*. Un ejemplo ampliamente usado es el núcleo de funciones de base radial, también llamado núcleo gaussiano.

Definición 3.2 (Núcleo de funciones de base radial)**.** Sea $l \in \mathbb{R} \setminus \{0\}$. El núcleo de funciones de base radial (*Radial Basis Function kernel, RBF*), se define como

$$k(\boldsymbol{x}_1,\boldsymbol{x}_2) = \exp\left(-\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2}{2l^2}\right), \quad \forall \boldsymbol{x}_1,\boldsymbol{x}_2 \in \mathbb{R}^D$$

El hiperparámetro l se denomina *lengthscale*.

3.2. Procesos gaussianos para regresión

Supongamos que disponemos de un conjunto de entrenamiento con N observaciones, $\mathcal{D} = \{(\boldsymbol{x}_i, y_i) : i \in \{1,\dots,N\}\} \subset \mathbb{R}^D \times \mathbb{R}$, donde $\boldsymbol{x}$ denota el vector de las variables observadas de un ejemplo e y denota la variable objetivo. Podemos recoger los ejemplos en una matriz $\boldsymbol{X} = [\boldsymbol{x}_1,\dots,\boldsymbol{x}_N] \in \mathbb{R}^{D\times N}$ y en un vector $\boldsymbol{y} = [y_1,\dots,y_N]^\top \in \mathbb{R}^N$. Para modelizar la relación entre $\boldsymbol{x}$ e y suponemos

$$y(\boldsymbol{x}) = f(\boldsymbol{x}) + \epsilon, \tag{3.1}$$

donde $f : \mathbb{R}^D \to \mathbb{R}$ es una función desconocida y $\epsilon \sim \mathcal{N}(0, \sigma_r^2)$. Suponemos la existencia de cierto ruido ϵ que hace diferir los valores observados de los que produce f. Cuando tomamos $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{w}$ nos encontramos en el escenario de la regresión lineal.

En lugar de suponer una forma funcional para f, podemos inducir una distribución sobre el espacio de las funciones usando un proceso gaussiano. Tomamos así $f \sim \mathcal{GP}(0,k)$, por lo que $\boldsymbol{f} \mid \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, k(\boldsymbol{X},\boldsymbol{X}))$. De (3.1) tenemos $\boldsymbol{y} \mid \boldsymbol{f} \sim \mathcal{N}(\boldsymbol{f}, \sigma_r^2\boldsymbol{I})$. Es importante conocer la siguiente definición.

Definición 3.3 (Verosimilitud marginal)**.** En las condiciones anteriores, la verosimilitud marginal es

$$p(\boldsymbol{y} \mid \boldsymbol{X}) = \int p(\boldsymbol{y},\boldsymbol{f} \mid \boldsymbol{X})\mathrm{d}\boldsymbol{f} = \int p(\boldsymbol{y} \mid \boldsymbol{f},\boldsymbol{X})p(\boldsymbol{f} \mid \boldsymbol{X})\mathrm{d}\boldsymbol{f}.$$

Observación 3.2. Se cumple que

$$\log p(\boldsymbol{f} \mid \boldsymbol{X}) = -\frac{1}{2}\boldsymbol{y}^T k(\boldsymbol{X},\boldsymbol{X})^{-1}\boldsymbol{y} - \frac{1}{2}\log|k(\boldsymbol{X},\boldsymbol{X})| - \frac{N}{2}\log(2\pi),$$

de donde

$$\log p(\boldsymbol{y} \mid \boldsymbol{X}) = -\frac{1}{2}\boldsymbol{y}^T\left(k(\boldsymbol{X},\boldsymbol{X}) + \sigma_r^2\boldsymbol{I}\right)^{-1}\boldsymbol{y} - \frac{1}{2}\log\left|k(\boldsymbol{X},\boldsymbol{X}) + \sigma_r^2\boldsymbol{I}\right| - \frac{N}{2}\log(2\pi). \tag{3.2}$$

Podemos deducir entonces que $\boldsymbol{y} \mid \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, k(\boldsymbol{X},\boldsymbol{X}) + \sigma_r^2\boldsymbol{I})$. Atendiendo a (3.1) esto prueba que y puede ser considerado otro GP con media 0 y covarianza dada, para cada par $\boldsymbol{x}_1,\boldsymbol{x}_2$, por $k(\boldsymbol{x}_1,\boldsymbol{x}_2) + \sigma_r^2\delta(\boldsymbol{x}_1,\boldsymbol{x}_2)$, donde $\delta(\boldsymbol{x}_1,\boldsymbol{x}_2) = 1$ si $\boldsymbol{x}_1 = \boldsymbol{x}_2$ y $\delta(\boldsymbol{x}_1,\boldsymbol{x}_2) = 0$ en otro caso.

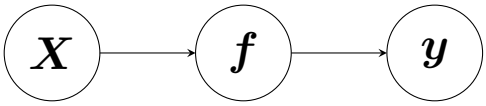


Figura 3.1.: Diagrama del modelo probabilístico para un problema de regresión.

El modelo que hemos construido se ilustra en la Figura 3.1 y queda determinado como sigue,

$$p(\boldsymbol{f} \mid \boldsymbol{X}) = \mathcal{N}\left(\boldsymbol{f} \mid \boldsymbol{0}, k\left(\boldsymbol{X}, \boldsymbol{X}\right)\right),$$
$$p(\boldsymbol{y} \mid \boldsymbol{f}) = \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{f}, \sigma_r^2 \boldsymbol{I}\right).$$

El objetivo en un problema de regresión es hacer predicciones para nuevos ejemplos basándonos en el conocimiento obtenido del conjunto de entrenamiento. Denotemos por $\boldsymbol{x}_*$ a un nuevo ejemplo (ejemplo de *test*) y por $y_* = y(\boldsymbol{x}_*)$ al valor de la función objetivo que deseamos predecir. Para poder hacer esto necesitamos calcular la distribución predictiva $p(y_* \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}_*)$. De la Observación 3.2 deducimos

$$p\left(\begin{bmatrix}\boldsymbol{f}\\ f_*\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{f}\\ f_*\end{bmatrix} \mid \boldsymbol{0}, \begin{bmatrix}k\left(\boldsymbol{X}, \boldsymbol{X}\right) & k(\boldsymbol{X}, \boldsymbol{x}_*)\\ k(\boldsymbol{x}_*, \boldsymbol{X}) & k(\boldsymbol{x}_*, \boldsymbol{x}_*)\end{bmatrix}\right)$$
$$p\left(\begin{bmatrix}\boldsymbol{y}\\ y_*\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{y}\\ y_*\end{bmatrix} \mid \boldsymbol{0}, \begin{bmatrix}k\left(\boldsymbol{X}, \boldsymbol{X}\right) + \sigma_r^2 \boldsymbol{I} & k(\boldsymbol{X}, \boldsymbol{x}_*)\\ k(\boldsymbol{x}_*, \boldsymbol{X}) & k(\boldsymbol{x}_*, \boldsymbol{x}_*) + \sigma_r^2\end{bmatrix}\right).$$

A partir de una distribución gaussiana podemos calcular fácilmente las distribuciones condicionadas a los valores de una o varias variables. Esto da lugar a los siguientes resultados:

Proposición 3.1. *La distribución predictiva de un GP sin considerar ruido, cumple*

$$p(f_* \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}_*) = \mathcal{N}\left(f_* \mid \mu_*, \sigma_*\right),$$

donde

$$\mu_* = k(\boldsymbol{x}_*, \boldsymbol{X}) k\left(\boldsymbol{X}, \boldsymbol{X}\right)^{-1} \boldsymbol{y},$$
$$\sigma_* = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{x}_*, \boldsymbol{X}) k(\boldsymbol{X}, \boldsymbol{X})^{-1} k(\boldsymbol{X}, \boldsymbol{x}_*).$$

Proposición 3.2. *La distribución predictiva de un GP considerando ruido, cumple*

$$p(y_* \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}_*) = \mathcal{N}\left(y_* \mid \mu_*, \sigma_*\right),$$

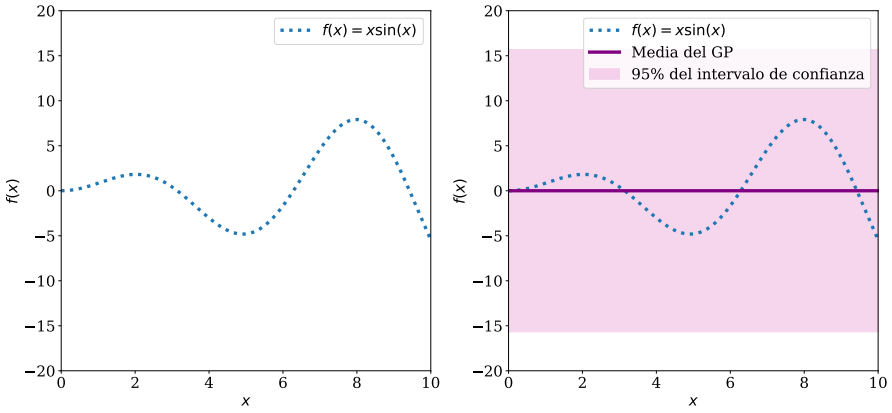
donde

$$\mu_* = k(\boldsymbol{x}_*, \boldsymbol{X}) \left[k\left(\boldsymbol{X}, \boldsymbol{X}\right) + \sigma_r^2 \boldsymbol{I}\right]^{-1} \boldsymbol{y},$$
$$\sigma_* = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{x}_*, \boldsymbol{X}) \left[k(\boldsymbol{X}, \boldsymbol{X}) + \sigma_r^2 \boldsymbol{I}\right]^{-1} k(\boldsymbol{X}, \boldsymbol{x}_*).$$

La distribución predictiva depende del ejemplo $\boldsymbol{x}_*$ y el cálculo de la misma conlleva la inversión de una matriz de dimensiones $N \times N$, lo cual tiene una complejidad $O(N^3)$. Esto es

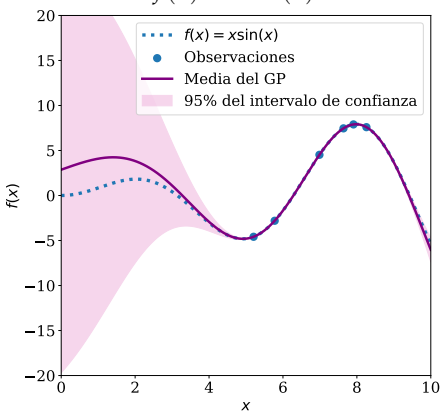
una gran limitación para los GPs ya que, en su formulación original, no pueden trabajar con conjuntos datos grandes. Sabemos que cualquier modelo predictivo aumenta su capacidad de generalizar a nuevos ejemplos conforme mayor es el conjunto de entrenamiento. Para solucionar esto, se utilizan técnicas consistentes en evaluar la covarianza en un conjunto de M puntos (con M fijo) que resumen el conjunto de entrenamiento.

En la Figura 3.2 se muestra un GP unidimensional. En la subfigura (a) se muestra la verdadera función que queremos aproximar. En la subfigura (b) se muestra la distribución a priori, antes de incorporar el conocimiento sobre los ejemplos observados. Una vez que se observa el valor de la función para algunos valores, la distribución a posteriori toma la forma de la subfigura (b). En la zona donde se han observado los ejemplos la varianza es menor, mientras que donde no hay ejemplos observados hay una gran incertidumbre. Si consideramos que hay cierto ruido a la hora de generar los ejemplos, se obtiene la distribución predictiva de la subfigura (d).

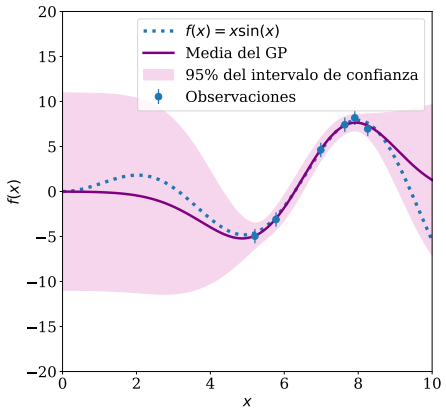


(a) La función que genera los ejemplos es $f(x) = x\sin(x)$.

(b) Distribución a priori.



(c) Distribución a posteriori sin considerar ruido.



(d) Distribución a posteriori considerando ruido.

Figura 3.2.: Visualización de un GP para regresión unidimensional.

3.3. Procesos gaussianos para clasificación

En un problema de clasificación supervisado, las etiquetas de los ejemplos no son valores reales, sino que pertenecen a algún conjunto discreto. La modelización probabilística de este problema pretende modelar la probabilidad a posteriori de que un nuevo ejemplo pertenezca a una clase o a otra, una vez observado el conjunto de entrenamiento. Para poder aplicar GPs en este escenario debemos transformar la salida del GP mediante alguna función de activación no lineal que nos permita realizar esa interpretación probabilística.

Fijemos en primer lugar la notación. Nuevamente, sea $\mathcal{D} = \{(\boldsymbol{x}_i, y_i) : i \in \{1, \ldots, N\}\} \subset \mathbb{R}^D \times \mathcal{Y}$ el conjunto de entrenamiento, donde $\mathcal{Y}$ es el espacio de las etiquetas. Consideremos el caso de un problema binario, $\mathcal{Y} = \{0, 1\}$. Modelamos la relación entre x e y suponiendo

$$p(y \mid f) = \sigma(f)^y \left(1 - \sigma(f)\right)^{1-y}$$

donde $\sigma : \mathbb{R} \to [0, 1]$ es una función sigmoidal y $f \sim \mathcal{GP}(0, k)$. La función f juega un papel peculiar ya que no la observamos directamente (sólo a través de las etiquetas) y tampoco estamos interesados en los valores que toma. Simplemente nos sirve para modelar el problema usando un GP. El modelo de GP para clasificación se puede representar gráficamente como en la Figura 3.1. Es común suponer que los ejemplos de entrenamiento son independientes, por lo que la distribución conjunta queda completamente determinada por

$$p(\boldsymbol{f} \mid \boldsymbol{X}) = \mathcal{N}\left(\boldsymbol{f} \mid \boldsymbol{0}, k(\boldsymbol{X}, \boldsymbol{X})\right),$$
$$p(\boldsymbol{y} \mid \boldsymbol{f}) = \prod_{n=1}^{N} \text{Bernouilli}(y_n \mid \sigma(f_n)).$$

Sin embargo, esta última suposición no tiene por qué darse en problemas MIL. Obsérvese que estamos suponiendo que las etiquetas de las instancias son independientes, lo cual no tiene por qué ocurrir en todos los problemas MIL. Un ejemplo de ello es el problema del cáncer de piel [36]. El modelo que hemos construido se ilustra en la Figura 3.1.

Para clasificar un ejemplo $\boldsymbol{x}_*$ necesitamos calcular la distribución predictiva $p(y_* \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}_*)$. Llamando $f_* = f(\boldsymbol{x}_*)$, se tiene

$$p(y_* \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}_*) = \int p(y_* \mid f_*) p(f_* \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}_*) \mathrm{d}f_*.$$

En un problema binario es suficiente calcular $p(y_* = 1 \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}_*)$ ya que $p(y_* = 0 \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}_*) = 1 - p(y_* = 1 \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}_*)$, luego basta concentrarse en

$$p(y_* = 1 \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}_*) = \int \sigma(f_*) p(f_* \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}_*) \mathrm{d}f_*.$$

Tanto esta expresión como la que define a $p(f_* \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}_*)$ son intratables de forma analítica. En el caso de la regresión sí que podíamos obtener expresiones cerradas para la distribución predictiva ya que los términos involucrados se correspondían a distribuciones gaussianas. Para salvar esto, debemos aproximar dichas integrales. Algunas aproximaciones comunes son la *aproximación de Laplace* [52], la conocida como *propagación de la esperanza* (*Expectance Propagation*, *EP*) [27], o aquellas basadas en *inferencia variacional* [17], aunque se han propuesto muchas otras [37, 21]. Todos estos métodos tienen complejidad $O(N^3)$, siendo N el número de ejemplos en entrenamiento. Para poder utilizar conjuntos grandes, al igual que en el caso

de los GPs para regresión, debemos integrar estas técnicas con otras que permitan reducir el coste computacional.

3.4. Ajuste de hiperparámetros

La distribución predictiva de un GP es la que nos permite predecir la variable objetivo para un ejemplo nuevo. Sin embargo, el comportamiento de esta distribución y la capacidad de un GP viene determinado por el núcleo que elijamos. Este núcleo, la mayoría de las ocasiones, dependerá de la elección de varios hiperparámetros, los cuales debemos ajustar cuidadosamente.

Para tal fin, recurrimos a la verosimilitud marginal, cuya expresión habíamos calculado para el problema de regresión en la ecuación (3.2) y volvemos a escribir a continuación haciendo explícita la dependencia de los hiperparámetros $\boldsymbol{\theta}$,

$$\log p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{y}^T \left(k_{\boldsymbol{\theta}}(\boldsymbol{X}, \boldsymbol{X}) + \sigma_r^2 \boldsymbol{I}\right)^{-1} \boldsymbol{y} - \frac{1}{2}\log \left|k_{\boldsymbol{\theta}}(\boldsymbol{X}, \boldsymbol{X}) + \sigma_r^2 \boldsymbol{I}\right| - \frac{N}{2}\log(2\pi).$$

Si la expresión que define al núcleo, $k_{\boldsymbol{\theta}}$ es diferenciable, podremos aplicar métodos de gradiente para optimizar la verosimilitud marginal. Esto se traducirá en encontrar aquellos hiperparámetros que maximizan la verosimilitud de la muestra.

El caso de la clasificación es más complicado ya que no disponemos de una expresión cerrada para la verosimilitud marginal. En la práctica, esta cantidad se aproxima de alguna forma y se efectúa la optimización sobre la aproximación obtenida.

3.5. Sparse GP

La solución más popular para salvar el coste computacional de los GPs recibe el nombre de *Procesos Gaussianos Dispersos* (*Sparse GPs*, *SGPs*). Consiste en asumir la existencia de un conjunto de *inducing points* $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M\} \subset \mathbb{R}^D$ procedentes del mismo espacio que los ejemplos de entrenamiento. El número de puntos considerados, M, se suele fijar mucho más pequeño que N.

La idea consiste en considerar este conjunto, $\boldsymbol{Z} = [\boldsymbol{z}_1^\top, \ldots, \boldsymbol{z}_M^\top] \in \mathbb{R}^{D \times M}$ y los valores que producen a través del GP f, $\boldsymbol{u} = f(\boldsymbol{Z}) = [f(\boldsymbol{z}_1), \ldots, f(\boldsymbol{z}_M)]^\top$. A través de estas variables pretendemos sustituir la matriz $k(\boldsymbol{X}, \boldsymbol{X})$ por otra de rango M, cuya inversión sea computacionalmente menos costosa.

Las relaciones que se supongan entre $\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{u}$ y $\boldsymbol{f}$ dan lugar a diferentes aproximaciones. Siguiendo [9], suponemos que $\boldsymbol{Z}$ y $\boldsymbol{u}$ se relacionan de la misma forma que lo hacen $\boldsymbol{X}$ y $\boldsymbol{f}$. Esto nos permite escribir

$$p\left(\begin{bmatrix}\boldsymbol{f}\\ \boldsymbol{u}\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{f}\\ \boldsymbol{u}\end{bmatrix}, \mid \boldsymbol{0}, \begin{bmatrix}k(\boldsymbol{X}, \boldsymbol{X}) & k(\boldsymbol{X}, \boldsymbol{Z})\\ k(\boldsymbol{Z}, \boldsymbol{X}) & k(\boldsymbol{Z}, \boldsymbol{Z})\end{bmatrix}\right),$$

lo cual da lugar a que las distribuciones involucradas tomen la siguiente forma,

$$\begin{aligned} p(\boldsymbol{u} \mid \boldsymbol{Z}) &= \mathcal{N}(\boldsymbol{u} \mid \boldsymbol{0}, k(\boldsymbol{Z}, \boldsymbol{Z})) \\ p(\boldsymbol{f} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{u}) &= \mathcal{N}(\boldsymbol{f} \mid \boldsymbol{a}, \widetilde{\boldsymbol{K}}), \quad \text{donde} \\ \boldsymbol{a} &= k(\boldsymbol{X}, \boldsymbol{Z})\, k(\boldsymbol{Z}, \boldsymbol{Z})^{-1}\, \boldsymbol{u}, \\ \widetilde{\boldsymbol{K}} &= k(\boldsymbol{X}, \boldsymbol{X}) - k(\boldsymbol{X}, \boldsymbol{Z})\, k(\boldsymbol{Z}, \boldsymbol{Z})^{-1}\, k(\boldsymbol{Z}, \boldsymbol{X}). \end{aligned} \tag{3.3}$$

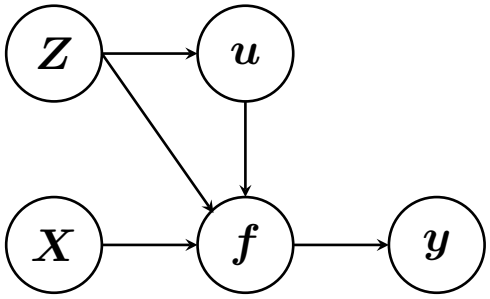


Figura 3.3.: Diagrama del modelo probabilístico SGPs.

Este modelo se representa gráficamente en la Figura 3.3. Obsérvese que las variables de $\boldsymbol{u}$ se introdujeron como variables auxiliares, luego las podemos marginalizar para volver al espacio de probabilidad de partida. Siguiendo [33], nos damos cuenta de que para aproximar $k(\boldsymbol{X}, \boldsymbol{X})$ basta considerar una aproximación $q(\boldsymbol{f} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{u})$ de la verdadera $p(\boldsymbol{f} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{u})$. Las diferentes aproximaciones a los SGPs se engloban dentro de esta interpretación sin más que cambiar las formas funcionales de $q(\boldsymbol{f} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{u})$. En general, se considera una distribución a posteriori aproximada de la forma

$$q(\boldsymbol{f} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{u}) = \mathcal{N}\left(\boldsymbol{f} \mid \boldsymbol{a}, \widetilde{\boldsymbol{Q}}\right),$$

siendo $\widetilde{\boldsymbol{Q}}$ una matriz aún por elegir. Podemos entonces marginalizar las variables $\boldsymbol{u}$ para obtener

$$q(\boldsymbol{f} \mid \boldsymbol{X}, \boldsymbol{Z}) = \int q(\boldsymbol{f} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{u}) p(\boldsymbol{u} \mid \boldsymbol{Z}) \mathrm{d}\boldsymbol{u} = \mathcal{N}\left(\boldsymbol{f} \mid \boldsymbol{0}, \widetilde{\boldsymbol{Q}} + \underbrace{k(\boldsymbol{X}, \boldsymbol{Z})\, k(\boldsymbol{Z}, \boldsymbol{Z})^{-1}\, k(\boldsymbol{Z}, \boldsymbol{X})}_{\boldsymbol{Q}_{\boldsymbol{X}}}\right),$$

donde, en la última igualdad, se ha usado la ecuación (3.3). La matriz $\boldsymbol{Q}_{\boldsymbol{X}}$ se conoce como la *aproximación de Nyström* de $k(\boldsymbol{X}, \boldsymbol{X})$. Para un problema de regresión podemos obtener la verosimilitud marginal aproximada como

$$q(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{Z}) = \int p(\boldsymbol{y} \mid \boldsymbol{f}, \boldsymbol{X}, \boldsymbol{Z}) q(\boldsymbol{f} \mid \boldsymbol{X}, \boldsymbol{Z}) = \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \widetilde{\boldsymbol{Q}} + \boldsymbol{Q}_{\boldsymbol{X}} + \sigma_r^2 \boldsymbol{I}\right)$$

Aún no hemos especificado la matriz $\widetilde{\boldsymbol{Q}}$, la cual tiene las mismas dimensiones que $k(\boldsymbol{X}, \boldsymbol{X})$. Para algunas elecciones de la misma podemos expresar su inversa de forma que involucre sólo la inversión de una matriz $M \times M$. Por ejemplo, el enfoque *deterministic training condicional (DTC)* [38] emplea $\widetilde{\boldsymbol{Q}} = \boldsymbol{0}$. La aproximación *fully independent training conditional (FITC)* [39] usa $\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}_{\boldsymbol{X}} - \operatorname{diag}(\boldsymbol{Q}_{\boldsymbol{X}} - k(\boldsymbol{X}, \boldsymbol{X}))$. Nosotros vamos a centrarnos en la propuesta de [42],

la cual introduce la inferencia variacional en el contexto de los GPs.

3.6. Aproximación variacional

La metodología que hemos descrito anteriormente se basa, esencialmente, en modificar la distribución a priori. De esta forma, los *inducing points* actúan como hiperparámetros del núcleo que hay que optimizar, lo cual puede llevar al modelo al sobreajuste. En lugar de esto, en [42] se propone aproximar la distribución a posteriori exacta empleando inferencia variacional. Así, los puntos $\boldsymbol{z}_n$ son parámetros que hay que elegir cuidadosamente.

En lo que sigue vamos a omitir la dependencia respecto de $\boldsymbol{X}$ y $\boldsymbol{Z}$. Obsérvese que para calcular la distribución predictiva para un nuevo ejemplo $\boldsymbol{x}_*$ podemos escribir

$$\begin{aligned} p(f_* \mid \boldsymbol{y}, \boldsymbol{x}_*) &= \int p(f_*, \boldsymbol{f}, \boldsymbol{u} \mid \boldsymbol{y}, \boldsymbol{x}_*) \mathrm{d}(\boldsymbol{f}, \boldsymbol{u}) = \\ &= \int p(f_* \mid \boldsymbol{f}, \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{x}_*) p(\boldsymbol{f} \mid \boldsymbol{u}) p(\boldsymbol{u} \mid \boldsymbol{y}) \mathrm{d}(\boldsymbol{f}, \boldsymbol{u}). \end{aligned}$$

Si suponemos que $\boldsymbol{u}$ es un estadístico suficiente para $\boldsymbol{f}$, esto es, $p(f_* \mid \boldsymbol{f}, \boldsymbol{u}) = p(f_* \mid \boldsymbol{u})$, entonces

$$\begin{aligned} p(f_* \mid \boldsymbol{y}, \boldsymbol{x}_*) &= \int p(f_* \mid \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{x}_*) p(\boldsymbol{f} \mid \boldsymbol{u}) p(\boldsymbol{u} \mid \boldsymbol{y}) \mathrm{d}(\boldsymbol{f}, \boldsymbol{u}) = \\ &= \int p(f_* \mid \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{x}_*) p(\boldsymbol{u} \mid \boldsymbol{y}) \mathrm{d}\boldsymbol{u}. \end{aligned}$$

En la práctica será difícil encontrar $\boldsymbol{u}$ que cumpla esta propiedad, por lo que la anterior será sólo una aproximación a la verdadera distribución predictiva. En tal caso podemos aproximar la intratable $p(\boldsymbol{u} \mid \boldsymbol{y})$ por $q(\boldsymbol{u})$. Si elegimos $q(\boldsymbol{u}) = \mathcal{N}(\boldsymbol{m}, \boldsymbol{S})$ y hacemos uso de la definición de GP (como hacíamos en la proposición Proposición 3.2, pero ahora considerando también los *inducing points*), entonces ocurrirá que $p(f_* \mid \boldsymbol{y}, \boldsymbol{x}_*) \approx \mathcal{N}(\overline{f_*}, c_*)$ siendo

$$\begin{aligned} \overline{f_*} &= k(\boldsymbol{x}_*, \boldsymbol{Z}) k(\boldsymbol{Z}, \boldsymbol{Z})^{-1} \boldsymbol{m} \\ c_* &= k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{x}_*, \boldsymbol{Z}) k(\boldsymbol{Z}, \boldsymbol{Z})^{-1} k(\boldsymbol{Z}, \boldsymbol{x}_*) + \\ &\quad + k(\boldsymbol{x}_*, \boldsymbol{Z}) k(\boldsymbol{Z}, \boldsymbol{Z})^{-1} \boldsymbol{S} k(\boldsymbol{Z}, \boldsymbol{Z})^{-1} k(\boldsymbol{Z}, \boldsymbol{x}_*). \end{aligned}$$

El cálculo de esta aproximación involucra la inversión de matrices de orden $M \times M$, de forma que la complejidad es $O(NM^2)$.

El problema ahora reside en cómo seleccionar los parámetros $(\boldsymbol{m}, \boldsymbol{S})$ y la localización de los *inducing points*. Empleando inferencia variacional, minimizamos la *divergencia de Kullback-Leibler* entre la verdadera distribución $p(\boldsymbol{f}, \boldsymbol{u} \mid \boldsymbol{y}) = p(\boldsymbol{f} \mid \boldsymbol{u}) p(\boldsymbol{u} \mid \boldsymbol{y})$ y la aproximación $Q(\boldsymbol{f}, \boldsymbol{u}) = p(\boldsymbol{f} \mid \boldsymbol{u}) q(\boldsymbol{u})$. Esto es equivalente a maximizar el ELBO, el cual actúa como una cota inferior al logaritmo de la verosimilitud marginal,

$$\begin{aligned} \log\left(p(\boldsymbol{y})\right) \geq ELBO(Q) &= \mathbb{E}_{Q(\boldsymbol{f}, \boldsymbol{u})}\left[\frac{p(\boldsymbol{f}, \boldsymbol{u}, \boldsymbol{y})}{Q(\boldsymbol{f}, \boldsymbol{u})}\right] = \\ &= \int p(\boldsymbol{f} \mid \boldsymbol{u}) q(\boldsymbol{u}) \log\left(\frac{p(\boldsymbol{y} \mid \boldsymbol{f}) p(\boldsymbol{u})}{q(\boldsymbol{u})}\right) \mathrm{d}(\boldsymbol{f}, \boldsymbol{u}). \end{aligned}$$

Como mostramos en la siguiente subsección, el ELBO es una aproximación inferior a la verosimilitud marginal, con la que podemos trabajar para ajustar los hiperparámetros. Cuanto mayor sea el ELBO, mejor aproximará a la verdadera verosimilitud marginal, por lo que su optimización tiene también un significado intuitivo. Tanto el ELBO como la solución óptima al problema de su optimización tienen una expresión cerrada, que se deduce de (3.4) y que puede consultarse en [42].

En esta sección hemos presentado la formulación de los SGPs para un problema de regresión. En un problema de clasificación es necesario, además, emplear algún tipo de técnica con la que poder tratar la función sigmoidal. En el Capítulo 4 vamos a estudiar dos de estas técnicas para adaptar los GPs al problema MIL, por lo que no consideramos necesario mostrarlas otra vez aquí. Para más información puede consultarse [4, 9].

3.6.1. Inferencia variacional

La inferencia variacional [5] consiste en la aplicación de métodos de optimización variacionales para resolver el problema de la inferencia bayesiana. Esta subsección recoge los resultados principales para poder aplicarla con toda generalidad. Esta técnica es esencial en nuestro trabajo, ya que la solución que proporciona constituye el estado del arte en los SGPs. Pueden consultarse algunos ejemplos en [4].

La divergencia de Kullback-Leibler jugará un papel muy importante. Empezamos definiéndola cuidadosamente.

Definición 3.4 (Divergencia de Kullback-Leibler)**.** Sean P y Q dos medidas de probabilidad en el mismo espacio de las que se conocen funciones de densidad o funciones masa de probabilidad, p y q respectivamente. Se define la *divergencia Kullback-Leibler* de p y q como

$$\mathrm{KL}(p \,||\, q) = \mathbb{E}_{p(x)} \left[\log p(x) - \log q(x)\right].$$

La desigualdad de la información hace referencia a uno de los resultados más importantes sobre la divergencia de Kullback-Leibler. Para probarla podemos recurrir a la desigualdad de Jensen [12]. La enunciamos a continuación.

Teorema 3.1 (Desigualdad de la información)**.** *Sean P y Q dos medidas de probabilidad en el mismo espacio de las que se conocen funciones de densidad, p y q respectivamente. Entonces*

1. $\mathrm{KL}(p \,||\, q) \geq 0$
2. $\mathrm{KL}(p \,||\, q) = 0$ *si y sólamente si* $p = q$ *en casi todo punto.*

Gracias a este resultado podemos identificar que la divergencia de Kullback-Leibler es una medida del parecido entre dos distribuciones de probabilidad. Veamos cómo podemos usar esta idea en el contexto de la inferencia bayesiana.. Supongamos que tenemos un modelo bayesiano definido por unos parámetros a los que se les ha dado una distribución a priori. Supongamos un conjunto de variables observadas $\boldsymbol{X}$ y otro conjunto de variables latentes (no observadas) $\boldsymbol{Z}$, dentro de las cuales entran los parámetros. El modelo especifica una probabilidad conjunta $p(\boldsymbol{X}, \boldsymbol{Z})$, y nuestro objetivo es encontrar una aproximación para la distribución a posteriori $p(\boldsymbol{Z} \mid \boldsymbol{X})$ y para la verosimilitud $p(\boldsymbol{X})$.

Consideremos una distribución de $q(\boldsymbol{Z})$, en principio arbitraria, con la que queremos aproximar a $p(\boldsymbol{Z} \mid \boldsymbol{X})$. Para cualquier elección de la misma la verosimilitud se puede descomponer

como sigue,

$$p(\boldsymbol{X}) = \underbrace{\int q(\boldsymbol{Z}) \log\left(\frac{p(\boldsymbol{X},\boldsymbol{Z})}{q(\boldsymbol{Z})}\right) \mathrm{d}\boldsymbol{Z}}_{\mathrm{ELBO}(q(\boldsymbol{Z}))} + \underbrace{\int q(\boldsymbol{Z}) \log\left(\frac{q(\boldsymbol{Z})}{p(\boldsymbol{Z} \mid \boldsymbol{X})}\right) \mathrm{d}\boldsymbol{Z}}_{\mathrm{KL}(q(\boldsymbol{Z})||p(\boldsymbol{Z}|\boldsymbol{X}))}.$$

El primer sumando recibe el nombre de *Evidence Lower Bound* (ELBO) y jugará un papel muy importante a la hora de aproximar la verosimilitud. Su nombre se debe a la siguiente desigualdad, consecuencia de la desigualdad de la información,

$$p(\boldsymbol{X}) = \mathrm{ELBO}(q(\boldsymbol{Z})) + \mathrm{KL}\left(q(\boldsymbol{Z}) \mid\mid p(\boldsymbol{Z} \mid \boldsymbol{X})\right) \geq \mathrm{ELBO}(q(\boldsymbol{Z})).$$

Obsérvese que se da la igualdad si y sólo si las distribuciones $q(\boldsymbol{Z})$ y $p(\boldsymbol{Z} \mid \boldsymbol{X})$ coinciden. Minimizar la divergencia de Kullback-Leibler es equivalente a maximizar el ELBO, lo cual lo convierte en una aproximación a la verosimilitud, cuya calidad dependerá de cómo de bien $q(\boldsymbol{Z})$ aproxime a $p(\boldsymbol{Z} \mid \boldsymbol{X})$. Así pues, para cumplir nuestro objetivo, debemos minimizar la divergencia de Kullback-Leibler o, equivalentemente, maximizar el ELBO.

Para llevar a cabo esta tarea no podemos trabajar con todos los tipos de distribuciones existentes, sino que debemos de restringirlos para poder encontrar una solución cerrada. Una forma de trabajar es parametrizar la distribución variacional, de forma que el ELBO sea una función de estos parámetros. En ese caso podemos emplear técnicas de optimización no lineal (como por ejemplo gradiente descendente en caso de que la expresión sea diferenciable) para maximizarlo.

Alternativamente, vamos a considerar la aproximación *Mean Field* [30]. Supongamos que dividimos las variables latentes en grupos disjuntos, de forma que $\boldsymbol{Z} = \{\boldsymbol{Z}_i \colon i \in \{1, \ldots, M\}\}$. La teoría *Mean Field* consiste en suponer que estos grupos son independientes, de forma que la distribución variacional se escribe como

$$q(\boldsymbol{Z}) = \prod_{i=1}^{M} q_i(\boldsymbol{Z}_i).$$

Obsérvese que no se están imponiendo restricciones sobre la forma funcional de las distribuciones q_i. En su lugar, se está asumiendo que la distribución variacional conjunta factoriza de una forma concreta.

Para maximizar el ELBO podemos pensar en optimizar su expresión respecto de cada distribución q_i mientras que se dejan el resto fijas. Escribamos la expresión del ELBO haciendo

uso de esta factorización,

$$\begin{aligned}\text{ELBO}(q(\boldsymbol{Z})) &= \int \prod_{i=1}^{M} q_i(\boldsymbol{Z}_i) \left(\log p(\boldsymbol{X}, \boldsymbol{Z}) - \sum_{i=1}^{M} \log q_i(\boldsymbol{Z}_i) \right) \mathrm{d}\boldsymbol{Z} = \\ &= \int q_j(\boldsymbol{Z}_j) \underbrace{\left(\int \log p(\boldsymbol{X}, \boldsymbol{Z}) \prod_{i=1, i\neq j}^{M} q_i(\boldsymbol{Z}_i) \mathrm{d}\boldsymbol{Z}_i \right)}_{\log \widetilde{p}(\boldsymbol{X}, \boldsymbol{Z}_j)} \mathrm{d}\boldsymbol{Z}_j - \int q_j(\boldsymbol{Z}_j) \log q_j(\boldsymbol{Z}_j) \mathrm{d}\boldsymbol{Z}_j + \text{const} = \\ &= \int q_j(\boldsymbol{Z}_j) \log \widetilde{p}(\boldsymbol{X}, \boldsymbol{Z}_j) \mathrm{d}\boldsymbol{Z}_j - \int q_j(\boldsymbol{Z}_j) \log q_j(\boldsymbol{Z}_j) \mathrm{d}\boldsymbol{Z}_j + \text{const} = \\ &= -\,\mathrm{KL}\left(q_j(\boldsymbol{Z}_j) \,||\, \widetilde{p}(\boldsymbol{X}, \boldsymbol{Z}_j) \right) + \text{const.}\end{aligned}$$

donde const hace referencia a una constante que no depende de $\boldsymbol{Z}_j$ y hemos definido una nueva distribución mediante la relación

$$\log \widetilde{p}(\boldsymbol{X}, \boldsymbol{Z}_j) = \mathbb{E}_{q(\boldsymbol{Z})/q_j(\boldsymbol{Z}_j)} \left[\log p(\boldsymbol{X}, \boldsymbol{Z}) \right] + \text{const.}$$

Si queremos maximizar el ELBO respecto de $q_j(\boldsymbol{Z}_j)$ dejando el resto de las distribuciones fijas, a la vista de la igualdad obtenida basta minimizar la divergencia de Kullback-Leibler entre $q_j(\boldsymbol{Z}_j)$ y $\widetilde{p}(\boldsymbol{X}, \boldsymbol{Z}_j)$, lo cual ocurre cuando ambas coinciden. Por tanto, hemos obtenido una expresión para la solución óptima,

$$\log q_j^*(\boldsymbol{Z}_j) = \mathbb{E}_{q(\boldsymbol{Z})/q_j(\boldsymbol{Z}_j)} \left[\log p(\boldsymbol{X}, \boldsymbol{Z}) \right] + \text{const.} \tag{3.4}$$

Otra forma de expresar este resultado, sin involucrar a la constante aditiva, es

$$q_j^*(\boldsymbol{Z}_j) = \frac{\exp\left(\mathbb{E}_{q(\boldsymbol{Z})/q_j(\boldsymbol{Z}_j)} \left[\log p(\boldsymbol{X}, \boldsymbol{Z}) \right] \right)}{\int \exp\left(\mathbb{E}_{q(\boldsymbol{Z})/q_j(\boldsymbol{Z}_j)} \left[\log p(\boldsymbol{X}, \boldsymbol{Z}) \right] \mathrm{d}\boldsymbol{Z}_j \right)}.$$

Las ecuaciones dadas por (3.4) no representan una solución explícita para la distribución variacional porque cada una depende del resto. Por ello, para obtener una solución en la práctica, debemos aplicarlas iterativamente hasta llegar a la solución óptima. Puede probarse que el ELBO es convexo respecto de cada uno de los factores, por lo que la convergencia a la solución óptima está garantizada [6].

4. Procesos Gaussianos para Aprendizaje a partir de Múltiples Instancias

En este punto del trabajo ya hemos estudiado los fundamentos de MIL y de los GPs. Ahora, vamos a usar los GPs como herramienta para resolver el problema MIL de clasificación binaria bajo la hipótesis normal. Recuérdese que la elección de los GPs estaba justificada por su formulación probabilística, que nos permite tener en cuenta la incertidumbre a la hora de modelar el problema.

Para poder usar los GPs en este problema, primero estudiamos el estado del arte. Es decir, analizamos cuidadosamente qué enfoques son los que presentan mejores resultados en clasificación con GPs y en MIL. Las ideas obtenidas de este estudio dan lugar a dos grandes contribuciones, que reciben el nombre de PG-VGPMIL y G-VGPMIL. La justificación teórica de ambas se recogen en este capítulo, por lo que es uno de los más importantes. Por ello creemos necesario explicar la estructura del mismo.

1. En la primera sección se estudia el modelo *Variational Gaussian Process Multiple Instance Learning* (VGPMIL) [19], que constituye el estado del arte en GPs para problemas MIL bajo el modelo de observación logístico. Para poder calcular las distribuciones variacionales, VGPMIL usa la *desigualdad de Jaakkola* que proporciona una cota cuadrática de la función logística. La desventaja de usar una desigualdad es que podemos perder información, por lo que conviene estudiar alternativas.

2. Motivados por el buen comportamiento de las variables aleatorias Pólya-Gamma en problemas de clasificación [51], nos preguntamos qué ocurre si las introducimos en la formulación de los GPs. Obtenemos así la primera constribución del trabajo, en la que, por primera vez en la literatura, se propone el uso de variables aleatorias Pólya-Gamma para realizar inferencia variacional en un nuevo modelo basado en GPs para MIL. Este modelo recibirá el nombre de *Pólya-Gamma Variational Gaussian Process Multiple Instance Learning* (PG-VGPMIL) y probaremos que es equivalente, en el *sentido Mean Field*, a VGPMIL.

3. PG-VGPMIL es un *modelo aumentado*, en el que se introducen ciertas variables que facilitan el cálculo de las distribuciones variacionales. La definición de las variables Pólya-Gamma motiva el uso de variables Gamma. Llegamos así a la segunda contribución, la cual propone por primera vez en la literatura el uso de variables aleatorias Gamma para realizar inferencia variacional en un nuevo modelo basado en GPs para MIL. Este modelo recibirá el nombre de *Gamma Variational Gaussian Process Multiple Instance Learning* (G-VGPMIL).

La primera sección de este capítulo fija la notación y el ambiente en el que nos vamos a mover. Hemos extraído aquellos elementos comunes a todos los modelos y los hemos apartado en esta sección introductoria.

4.1. Notación y ambiente

Antes de entrar en materia debemos fijar la notación y el ambiente en el que nos vamos a mover. Consideramos un problema MIL de clasificación binaria. Sea el conjunto de entrenamiento,

$$\mathcal{D} = \{(\boldsymbol{x}_n, y_n) \colon n \in \{1, \dots, N\}\} \subset \mathbb{R}^D \times \{0, 1\}\,,$$

formado por las variables observadas $\boldsymbol{x}_n$ y las etiquetas de las instancias que no han sido observadas, y_n. Este conjunto se recoge en una matriz $\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_N] \in \mathbb{R}^{D \times N}$ y en un vector $\boldsymbol{y} = [y_1, \dots, y_N]^\top \in \{0, 1\}^N$. El conjunto de entrenamiento se particiona en B bolsas $\{\text{Bag}_1, \dots, \text{Bag}_B\}$ con intersección vacía y se observa una etiqueta $T_b \in \{0, 1\}$ para cada bolsa. Usamos el operador $\{\cdot\}_b$ para referirnos a los elementos de la bolsa b. Por ejemplo, $\{y\}_b = \{y_i \colon i \in \text{Bag}_b\}$ y $\{y\}_{b \setminus n} = \{y\}_b \setminus \{y_n\}$ Asumimos que se cumple $T_b = \max\{y\}_b$. Dado una aplicación núcleo $k \colon \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ y dos conjuntos de datos $\boldsymbol{X} \in \mathbb{R}^{N \times D}, \boldsymbol{Z} \in \mathbb{R}^{M \times D}$, llamamos $\boldsymbol{K}_{\boldsymbol{XZ}} \in \mathbb{R}^{N \times M}$ a la matriz dada por $(\boldsymbol{K}_{\boldsymbol{XZ}})_{ij} = k(\boldsymbol{x}_i, \boldsymbol{z}_j)$. Recordemos que la clasificación binaria mediante GPs se efectúa tomando

$$\begin{aligned} p(\boldsymbol{f} \mid \boldsymbol{X}) &= \mathcal{N}(\boldsymbol{f} \mid \boldsymbol{0}, \boldsymbol{K}_{\boldsymbol{XX}})\,, \\ p(\boldsymbol{y} \mid \boldsymbol{f}) &= \prod_{n=1}^{N} p(y_n \mid f_n) = \prod_{n=1}^{N} \text{Bernouilli}\,(y_n \mid \sigma(f_n))\,, \end{aligned} \tag{4.5}$$

donde $\sigma \colon \mathbb{R} \to]0, 1[$ es la función logística. Así, el signo de f determina la clase de la predicción y el valor absoluto determina la confianza que se tiene en ella. Obsérvese, y esto es muy importante, que se está asumiendo independencia entre todas las instancias y, particularmente, independencia entre las instancias de una bolsa. Para poder trabajar con grandes cantidades de datos adoptamos la aproximación *fully independent training conditional (FITC,* [39]*)*, a través de la cual tomamos un conjunto de *inducing points* $\boldsymbol{Z} = [\boldsymbol{z}_1, \dots, \boldsymbol{z}_N]^\top \in \mathbb{R}^{D \times M}$ y su evaluación mediante f, $\boldsymbol{u} = f(\boldsymbol{Z}) = [f(\boldsymbol{z}_1), \dots, f(\boldsymbol{z}_M)]^\top \in \mathbb{R}^M$. Suponemos que $\boldsymbol{Z}$ y $\boldsymbol{u}$ siguen la misma relación que $\boldsymbol{X}$ y $\boldsymbol{f}$. Esto es,

$$p(\boldsymbol{u} \mid \boldsymbol{Z}) = \mathcal{N}(\boldsymbol{u} \mid \boldsymbol{0}, k(\boldsymbol{Z}, \boldsymbol{Z})) \tag{4.6}$$

$$p(\boldsymbol{f} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{u}) = \mathcal{N}(\boldsymbol{f} \mid \boldsymbol{a}, \widetilde{\boldsymbol{K}}), \quad \text{donde} \tag{4.7}$$

$$\begin{aligned} \boldsymbol{a} &= k(\boldsymbol{X}, \boldsymbol{Z})\, k(\boldsymbol{Z}, \boldsymbol{Z})^{-1}\, \boldsymbol{u}, \\ \widetilde{\boldsymbol{K}} &= k(\boldsymbol{X}, \boldsymbol{X}) - k(\boldsymbol{X}, \boldsymbol{Z})\, k(\boldsymbol{Z}, \boldsymbol{Z})^{-1}\, k(\boldsymbol{Z}, \boldsymbol{X})\,. \end{aligned}$$

En un problema MIL no conocemos las etiquetas de las instancias, por lo que debemos extender el modelo probabilístico para tener en cuenta la relación existente entre las instancias y las bolsas. Esto es lo que hacemos a continuación.

4.2. VGPMIL: modelo basado en la desigualdad de Jaakkola

El modelo *Variational Gaussian Process Multiple Instance Learning* (VGPMIL) fue presentado en [19] y muestra, por primera vez, que es posible predecir las etiquetas a nivel de instancia de un clasificador basado en GPs usando inferencia variacional. Para ello, aproxima la hipótesis normal de los problemas MIL mediante una versión ruidosa. Esto, junto con la desigualdad de Jaakkola, permite obtener una expresión analítica de las distribuciones variacionales, lo

que asegura que el modelo es escalable y las distribuciones convergen rápidamente al óptimo.

4.2.1. Relación entre las instancias y las bolsas

Recordemos que en el escenario MIL las etiquetas $\boldsymbol{y}$ no son conocidas. En su lugar, hemos observado las etiquetas a nivel de bolsa, $\boldsymbol{T} = [T_1, \ldots, T_B]^\top$. La clave para adaptar la formulación de un GP de clasificación a MIL es cómo modelamos la relación entre estas etiquetas y las instancias. Por ejemplo, en [23] se toma $p(T_b \mid \{\boldsymbol{f}\}_b) = \text{Bernouilli}\left(T_b \mid \sigma\left(\max\{\boldsymbol{f}\}_b\right)\right)$, suponiendo que la instancia con mayor confianza es la responsable de la clase de la bolsa.

En su lugar, para modelizar la relación entre la etiqueta de la bolsa y las etiquetas de las instancias en VGPMIL se propone usar

$$p(T_b \mid \{\boldsymbol{y}\}_b) = \left(\frac{H}{H+1}\right)^{G_b}\left(1 - \frac{H}{H+1}\right)^{1-G_b} = \frac{H^{G_b}}{H+1}, \tag{4.8}$$

siendo $H \in \mathbb{R}^+$ y $G_b = T_b \max\{\boldsymbol{y}\}_b + (1 - T_b)\left(1 - \max\{\boldsymbol{y}\}_b\right)$. Obsérvese que G_b es igual a uno si se cumple la hipótesis MIL, en cuyo caso se asigna una probabilidad cercana a 1. Si no se cumple, G_b es cero y se asigna una probabilidad cercana a cero. La constante H determina cómo de cercanas serán estas probabilidades a uno o a cero, respectivamente. Obsérvese que

$$\lim_{H \to +\infty} p(T_b \mid \{\boldsymbol{y}\}_b) = \begin{cases} 1 & \text{si } T_b = \max\{\boldsymbol{y}\}_b, \\ 0 & \text{en otro caso,} \end{cases}$$

lo cual convierte a $p(T_b \mid \{\boldsymbol{y}\}_b)$ en una versión ruidosa de la hipótesis MIL. Esto justifica que la constante H deba fijarse en torno a valores altos. Por último, suponemos independencia entre las bolsas, lo que da lugar a

$$p(\boldsymbol{T} \mid \boldsymbol{y}) = \prod_{b=1}^{B} \frac{H^{G_b}}{H+1}.$$

El modelo completo se obtiene tomando el método de observación logístico para las etiquetas de las instancias (ecuación (4.5)) junto con la aproximación FITC (ecuaciones (4.6) y (4.7)) y la versión ruidosa de la hipótesis normal que acabamos de presentar (ecuación (4.8)). Se ilustra en la Figura 4.1 y queda definido por las siguientes ecuaciones,

$$\begin{aligned} p\left(\boldsymbol{u} \mid \boldsymbol{Z}\right) &= \mathcal{N}\left(\boldsymbol{u} \mid \boldsymbol{0}, \boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}\right) \\ p\left(\boldsymbol{f} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{u}\right) &= \mathcal{N}(\boldsymbol{f} \mid \boldsymbol{a}, \widetilde{\boldsymbol{K}}), \\ p\left(\boldsymbol{y} \mid \boldsymbol{f}\right) &= \prod_{n=1}^{N} \text{Bernouilli}(y_n \mid \sigma(f_n)), \\ p(\boldsymbol{T} \mid \boldsymbol{y}) &= \prod_{b=1}^{B} \frac{H^{G_b}}{H+1}. \end{aligned}$$

Para poder realizar predicciones con este modelo, debemos calcular las distribuciones predictivas de las etiquetas de las instancias y de las bolsas. Veamos cómo hacerlo a continuación.

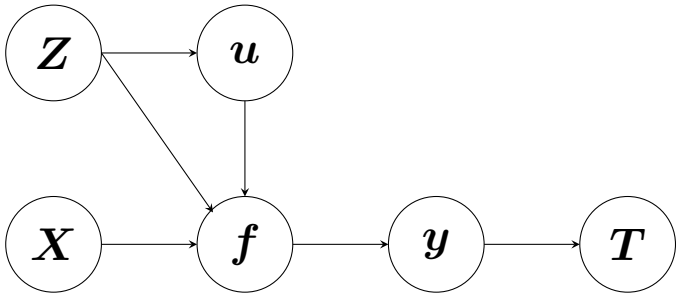


Figura 4.1.: Diagrama del modelo probabilístico VGPMIL.

4.2.2. Distribuciones predictivas

Para obtener la distribución predictiva empleamos un procedimiento análogo al que se utiliza en los SGPs. Dado un nuevo ejemplo $\boldsymbol{x}_*$, se tiene

$$\begin{aligned} p(f_* \mid \boldsymbol{T}) &= \int p(f_* \mid \boldsymbol{f}, \boldsymbol{u}) p(\boldsymbol{f}, \boldsymbol{u} \mid \boldsymbol{T}) \mathrm{d}\boldsymbol{f} \mathrm{d}\boldsymbol{u} \approx \\ &\approx \int p(f_* \mid \boldsymbol{f}, \boldsymbol{u}) p(\boldsymbol{f} \mid \boldsymbol{u}) q(\boldsymbol{u}) \mathrm{d}\boldsymbol{f} \mathrm{d}\boldsymbol{u} = \\ &= \int p(f_* \mid \boldsymbol{u}) q(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} = \mathcal{N}\left(f_* \mid \mu_*, \sigma_*^2\right) \end{aligned} \tag{4.9}$$

con $\mu_* = \boldsymbol{K}_{\boldsymbol{x}_*\boldsymbol{Z}} \boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1} \boldsymbol{m}$ y $\sigma_*^2 = \boldsymbol{K}_{\boldsymbol{x}_*\boldsymbol{x}_*} + \boldsymbol{K}_{\boldsymbol{x}_*\boldsymbol{Z}} \boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1} \left(\boldsymbol{S}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1} - \boldsymbol{I}\right) \boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{x}_*}$. La distribución de las etiquetas se calcula mediante

$$p(y_* = 1 \mid \boldsymbol{T}) = \int \sigma(f_*) p(f_* \mid \boldsymbol{T}) \mathrm{d}f_*,$$

la cual es una integral que no admite una expresión analítica pero que se puede calcular mediante métodos numéricos de forma muy rápida. Para predecir la etiqueta de una bolsa $\left\{\boldsymbol{x}_*^1, \ldots, \boldsymbol{x}_*^{N_*}\right\}$ con etiquetas desconocidas $\boldsymbol{y}_* = \left\{y_*^1, \ldots, y_*^{N_*}\right\}$, calculamos la distribución de la etiqueta de la bolsa mediante

$$\begin{aligned} p(T_* = 1 \mid \boldsymbol{T}) &= \prod_{\boldsymbol{y}_* \in \{0,1\}^{N_*}} p(T_* = 1 \mid \boldsymbol{y}_*) p(\boldsymbol{y}_* \mid \boldsymbol{T}) = \\ &= \prod_{\boldsymbol{y}_* \in \{0,1\}^{N_*}} \frac{H^{\max \boldsymbol{y}_*}}{H+1} \left(\prod_{i=1}^{N_*} p(y_*^i \mid \boldsymbol{T})\right). \end{aligned}$$

Llegados a este punto ya hemos establecido el modelo probabilístico subyacente y la forma de realizar predicciones a partir de las distribuciones predictivas. Ahora pasamos a estudiar el proceso de inferencia de VGPMIL.

4.2.3. Inferencia

El proceso de inferencia consiste en calcular la distribución a posteriori de las variables no observadas dadas las variables observadas. Esta distribución es $p(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y} \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Z})$ y es

esencial para poder calcular las distribuciones predictivas. Como esta distribución no admite una expresión analítica, se decide aproximarla empleando inferencia variacional. El objetivo de esta subsección es obtener una expresión analítica para tal aproximación.

En lo que sigue omitiremos la dependencia respecto de $\boldsymbol{X}$ y $\boldsymbol{Z}$. Para obtener la aproximación a posteriori $p(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y} \mid \boldsymbol{T})$ se emplea inferencia variacional y el enfoque Mean Field, considerando así la distribución variacional dada por

$$\begin{aligned} Q(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}) &= q(\boldsymbol{u}) p(\boldsymbol{f} \mid \boldsymbol{u}) q(\boldsymbol{y}), \quad \text{donde} \\ q(\boldsymbol{u}) &= \mathcal{N}(\boldsymbol{u} \mid \boldsymbol{m}, \boldsymbol{S}), \\ q(\boldsymbol{y}) &= \prod_{n=1}^{N} q(y_n) = \prod_{n=1}^{N} \pi_n^{y_n} (1-\pi_n)^{1-y_n}. \end{aligned}$$

Buscamos minimizar la divergencia de Kullback-Leibler,

$$Q^*(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}) = q^*(\boldsymbol{u}) p(\boldsymbol{f} \mid \boldsymbol{u}) q^*(\boldsymbol{y}) = \arg\min_{Q} \mathrm{KL}\left(Q(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}) \mid\mid p(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y} \mid \boldsymbol{T})\right).$$

Como sabemos, la solución a este problema viene dada, para cada factor q^* de Q^*, por (3.4), que en nuestro caso se traduce en

$$\log q^* = \mathbb{E}_{Q^*/q^*}\left[\log p(\boldsymbol{T}, \boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y})\right] + \text{const},$$

donde Q^*/q^* se refiere al cociente entre la distribución Q^* y el factor q^*. Es decir, de Q^* eliminamos el factor q^*. La clave de VGPMIL reside en la *desigualdad de Jaakkola*, que nos permite aproximar $\mathbb{E}_{p(\boldsymbol{f}\mid\boldsymbol{u})q(\boldsymbol{y})}\left[p(\boldsymbol{y} \mid \boldsymbol{f})\right]$.

Lema 4.1 (Desigualdad de Jaakkola [22].)**.** *Para todo $x \in \mathbb{R}$ y $\xi \in \mathbb{R}^+$ se cumple*

$$\sigma(x) \geq \sigma(\xi) \exp\left(\frac{x-\xi}{2} - \lambda(\xi)(x^2 - \xi^2)\right),$$

donde $\lambda(\xi) = (\sigma(\xi) - 1/2)\,/2$ y σ denota a la función logística.

Obsérvese que se da la igualdad si tomamos $x = \xi$, luego la bondad de esta aproximación vendrá dada por cómo se elija el valor de ξ. De ella deducimos

$$p(y_n \mid f_n) \geq \exp\left(-\frac{f_n + \xi_n}{2} - \lambda(\xi)(f_n^2 - \xi_n^2)\right) \exp\left(y_n f_n\right) \sigma(\xi_n).$$

Esta última desigualdad conduce a una cota inferior del ELBO, que es la que debemos maximizar para elegir los parámetros ξ_n. Esto es equivalente a minimizar la cota inferior de la siguiente expresión,

$$\mathbb{E}\left[\log p(y_n \mid f_n)\right] \geq \mathbb{E}\left[f_n\right](\pi_n - 1/2) - \mathbb{E}\left[f_n^2\right]\lambda(\xi_n) - \frac{\xi_n}{2} + \lambda(\xi_n)\xi_n^2 - \log\left(1 + e^{-\xi_n}\right),$$

donde la esperanza se toma respecto a la distribución $p(f_n \mid \boldsymbol{u})q(\boldsymbol{u})$. Es fácil comprobar, y así lo hacemos en la prueba del resultado principal de esta sección, que el máximo se alcanza cuando $\xi_n = \sqrt{\mathbb{E}[f_n^2]}$. El resultado clave para poder realizar inferencia es el siguiente.

Resultado 4.1. *Sea $p(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}, \boldsymbol{T})$ la distribución conjunta dada por el modelo VGPMIL. Entonces la solución óptima al problema*

$$\min_{q(\boldsymbol{u})q(\boldsymbol{y})} \mathrm{KL}\left(q(\boldsymbol{u})p(\boldsymbol{f} \mid \boldsymbol{u})q(\boldsymbol{y}) \mid\mid p(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y} \mid \boldsymbol{T})\right)$$

viene dada por las siguientes distribuciones:

$$\begin{aligned} q^*(\boldsymbol{u}) &\approx \mathcal{N}\left(\boldsymbol{u} \mid \boldsymbol{m}, \boldsymbol{S}\right), \\ q^*(y_n) &\approx \text{Bernouilli}\left(y_n \mid \pi_n\right) \end{aligned}$$

siendo

$$\begin{aligned} \pi_n &= \sigma\left[\boldsymbol{\kappa}_n \boldsymbol{m} + \log H\left(2T_b - 1\right)\left(1 - \mathbb{E}\left[\max\left\{\boldsymbol{y}\right\}_{b\setminus n}\right]\right)\right], \\ \boldsymbol{m} &= \boldsymbol{S}\boldsymbol{K}_{\boldsymbol{ZZ}}^{-1}\boldsymbol{K}_{\boldsymbol{ZX}}\left(\boldsymbol{\pi} - \frac{1}{2}\right), \\ \boldsymbol{S} &= \left(\boldsymbol{K}_{\boldsymbol{ZZ}}^{-1}\boldsymbol{K}_{\boldsymbol{ZX}}\boldsymbol{\Lambda}\boldsymbol{K}_{\boldsymbol{XZ}}\boldsymbol{K}_{\boldsymbol{ZZ}}^{-1} + \boldsymbol{K}_{\boldsymbol{ZZ}}^{-1}\right)^{-1}, \\ \xi_n &= \sqrt{\mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}[f_n^2]} = \sqrt{\widetilde{\boldsymbol{K}}_{nn} + \boldsymbol{\kappa}_n\left(\boldsymbol{m}\boldsymbol{m}^\top + \boldsymbol{S}\right)\boldsymbol{\kappa}_n^\top}, \end{aligned}$$

donde $\boldsymbol{\kappa}_n = \boldsymbol{K}_{\boldsymbol{x}_n\boldsymbol{Z}}\boldsymbol{K}_{\boldsymbol{ZZ}}^{-1} = \mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}\left[f_n\right]$ y $\boldsymbol{\Lambda} = \mathrm{diag}\left(2\lambda(\xi_1), \ldots, 2\lambda(\xi_N)\right)$.

Antes de proceder a la demostración, debemos enunciar y probar dos lemas que serán muy útiles en lo que sigue.

Lema 4.2. *Consideremos una instancia n y sea b la bolsa donde se encuentra esta instancia. Entonces*

$$\max\left\{\boldsymbol{y}\right\}_b = y_n + \max\left\{\boldsymbol{y}\right\}_{b\setminus n} - y_n \max\left\{\boldsymbol{y}\right\}_{b\setminus n}.$$

Demostración. Repetimos el proceso de [19],

$$\begin{aligned} \max\left\{\boldsymbol{y}\right\}_b &= 1 - \prod_{i\in\mathrm{Bag}_b}\left(1 - y_i\right) = 1 - \left(1 - y_n\right)\prod_{i\in\mathrm{Bag}_b\setminus\{n\}}\left(1 - y_i\right) \\ &= 1 - \prod_{i\in\mathrm{Bag}_b\setminus\{n\}}\left(1 - y_i\right) + y_n - y_n + y_n\prod_{i\in\mathrm{Bag}_b\setminus\{n\}}\left(1 - y_i\right) \\ &= \max\left\{\boldsymbol{y}\right\}_{b\setminus n} + y_n - y_n\left(1 - \prod_{i\in\mathrm{Bag}_b\setminus\{n\}}\left(1 - y_i\right)\right) \\ &= y_n + \max\left\{\boldsymbol{y}\right\}_{b\setminus n} - y_n \max\left\{\boldsymbol{y}\right\}_{b\setminus n} \end{aligned}$$

□

Lema 4.3. *Sean $a \in \mathbb{R}^+$ y $b \in \mathbb{R}$. Consideramos la función $f\colon \mathbb{R}^+ \to \mathbb{R}$ dada por*

$$f(x) = b + \lambda(x)\left(-a + x^2\right) - \frac{x}{2} - \log\left(1 + e^{-x}\right).$$

Entonces f alcanza un máximo absoluto cuando $x = \sqrt{a}$.

Demostración. La función f es derivable con

$$f'(x) = \lambda'(x)\left(-a + x^2\right).$$

Como λ es estrictamente decreciente en $\mathbb{R}^+$, la función f' sólo se anula en $x = \sqrt{a}$ y este es un máximo absoluto. □

Demostración. Apliquemos la desigualdad de Jaakkola,

$$\begin{aligned} p(\boldsymbol{y} \mid \boldsymbol{f}) &= \prod_{n=1}^{N} p(y_n \mid f_n) \geq \\ &\geq \exp\left(\sum_{n=1}^{N} f_n\,(y_n - 1/2) - f_n^2 \lambda(\xi_n)\right) \underbrace{\exp\left(\sum_{n=1}^{N} -\frac{\xi_n}{2} + \lambda(\xi_n)\xi_n^2\right) \prod_{n=1}^{N} \sigma(\xi_n)}_{C(\boldsymbol{\xi})} = \\ &= \exp\left[(\boldsymbol{y} - 1/2)^\top \boldsymbol{f} - \frac{1}{2}\boldsymbol{f}^\top \boldsymbol{\Lambda} \boldsymbol{f}\right] C(\boldsymbol{\xi}). \end{aligned}$$

Actualización de $q(y_n)$. Denotamos $q(y_{j\neq n}) = \prod_{j\neq n} \pi_j^{y_j}(1-\pi_j)^{y_j}$. Fijado n, sea b el índice de la bolsa en la que se encuentra la instancia n. Se tiene:

$$\begin{aligned} \log q^*(y_n) = &\underbrace{\mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})q(y_{j\neq n})}\left[\log p(\boldsymbol{y} \mid \boldsymbol{f})\right]}_{A_1} + \\ &+ \underbrace{\mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})q(y_{j\neq n})}\left[\log p(\boldsymbol{T} \mid \boldsymbol{y})\right]}_{A_2} + \text{const}. \end{aligned}$$

Analizamos cada sumando por separado,

$$\begin{aligned} A_1 &\geq \mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})q(y_{j\neq n})}\left[(\boldsymbol{y} - 1/2)^\top \boldsymbol{f}\right] + \text{const} = \\ &= y_n \underbrace{\boldsymbol{K}_{\boldsymbol{x}_i \boldsymbol{Z}} \boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}}_{\boldsymbol{\kappa}_n} \boldsymbol{m} + \text{const} \\ A_2 &= \log H \mathbb{E}_{q(y_{j\neq n})}\left[G_b\right] + \text{const} = \\ &= y_n \log H(2T_b - 1)\left(1 - \mathbb{E}\left[\max\left\{\boldsymbol{y}\right\}_{b\setminus n}\right]\right) + \text{const}. \end{aligned}$$

Juntando todo nos queda,

$$\log q^*(y_n) \geq y_n\left[\boldsymbol{\kappa}_n \boldsymbol{m} + \log H(2T_b - 1)\left(1 - \mathbb{E}\left[\max\left\{\boldsymbol{y}\right\}_{b\setminus n}\right]\right)\right] + \text{const},$$

de donde deducimos que $q(y_n)$ debe aproximarse a una Bernouilli $(y_n \mid \pi_n)$ con

$$\pi_n = \sigma\left\{\boldsymbol{\kappa}_n \boldsymbol{m} + \log H(2T_b - 1)\left(1 - \mathbb{E}\left[\max\left\{\boldsymbol{y}\right\}_{b\setminus n}\right]\right)\right\}$$

Actualización de $q(\boldsymbol{u})$.

$$\log q^*(\boldsymbol{u}) = \underbrace{\log p(\boldsymbol{u})}_{B_1} + \underbrace{E_{p(\boldsymbol{f}|\boldsymbol{u})}\left[\log p(\boldsymbol{f} \mid \boldsymbol{u})\right]}_{\text{const}} + \underbrace{\mathbb{E}_{p(\boldsymbol{f}|\boldsymbol{u})q(\boldsymbol{y})}\left[\log p(\boldsymbol{y}|\boldsymbol{f})\right]}_{B_2} + \text{const}$$

Analizamos cada sumando por separado,

$$\begin{aligned} B_1 &= -\frac{1}{2}\boldsymbol{u}^\top \boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}\boldsymbol{u} + \text{const}, \\ B_2 &\geq \mathbb{E}_{q(\boldsymbol{y})p(\boldsymbol{f}|\boldsymbol{u})}\left[(\boldsymbol{y}-1/2)^\top \boldsymbol{f} - \frac{1}{2}\boldsymbol{f}^\top \boldsymbol{\Lambda} \boldsymbol{f}\right] + \text{const} = \\ &= (\boldsymbol{\pi}-1/2)^\top \boldsymbol{K}_{\boldsymbol{X}\boldsymbol{Z}}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}\boldsymbol{u} - \frac{1}{2}\left(\text{tr}(\boldsymbol{\Lambda}\widetilde{\boldsymbol{K}}) + \boldsymbol{u}^\top \boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{X}}\boldsymbol{\Lambda}\boldsymbol{K}_{\boldsymbol{X}\boldsymbol{Z}}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}\boldsymbol{u}\right) + \text{const}. \end{aligned}$$

Juntando todo:

$$\log q^*(\boldsymbol{u}) \geq (\boldsymbol{\pi}-1/2)^\top \boldsymbol{K}_{\boldsymbol{X}\boldsymbol{Z}}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}^\top\left(\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{X}}\boldsymbol{\Lambda}\boldsymbol{K}_{\boldsymbol{X}\boldsymbol{Z}}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1} + \boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}\right)\boldsymbol{u} + \text{const},$$

de donde $q(\boldsymbol{u})$ debe aproximarse a una $\mathcal{N}(\boldsymbol{u} \mid \boldsymbol{m}, \boldsymbol{S})$ con

$$\begin{aligned} \boldsymbol{m} &= \boldsymbol{S}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{X}}(\boldsymbol{\pi}-1/2) \\ \boldsymbol{S} &= \left(\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{X}}\boldsymbol{\Lambda}\boldsymbol{K}_{\boldsymbol{X}\boldsymbol{Z}}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1} + \boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}\right)^{-1} \end{aligned}$$

Actualización de ξ_n. Para obtener la expresión de ξ_n debemos de maximizar la cota inferior del ELBO que produce la desigualdad de Jaakkola. Esto es equivalente a maximizar la cota inferior de la siguiente expresión,

$$\mathbb{E}\left[\log p(y_n \mid f_n)\right] \geq \mathbb{E}\left[f_n\right](\pi_n - 1/2) - \mathbb{E}\left[f_n^2\right]\lambda(\xi_n) - \frac{\xi_n}{2} + \lambda(\xi_n)\xi_n^2 - \log\left(1+e^{-\xi_n}\right),$$

en la que la esperanza se toma respecto a la distribución $p(f_n \mid \boldsymbol{u})q(\boldsymbol{u})$. Aplicando el Lema 4.3 se tiene que el máximo se alcanza cuando $\xi = \sqrt{\mathbb{E}\left[f_n^2\right]}$. □

Obsérvese que no hemos escrito una igualdad al dar la forma de las distribuciones óptimas. Esto se debe al uso que se hace de la desigualdad de Jaakkola. Es natural pensar que al usar una desigualdad estamos perdiendo información, lo cual puede conducir a que las distribuciones variacionales no sean óptimas. La pregunta es, ¿cómo podemos realizar inferencia sin emplear una desigualdad para tratar la función logística? Este trabajo encontró la respuesta a tal pregunta en las variables Pólya-Gamma, que permiten expresar la función σ de forma cerrada. Así es como surge la primera contribución de este trabajo, que presentamos a continuación.

4.3. PG-VGPMIL: modelo basado en variables Pólya-Gamma

Como hemos comentado, el principal inconveniente de la desigualdad de Jaakkola es que puede conducir a distribuciones que no son óptimas. En [51] se usan variables aleatorias Pólya-Gamma para derivar un modelo de GPs para clasificación binaria. La técnica usada

recibe el nombre de *data augmentation* y consiste en modificar la distribución conjunta incluyendo nuevas variables. Al marginalizarlas, recuperamos el modelo original. En el caso de las variables Pólya-Gamma, estas facilitan el tratamiento de la función logística a la hora de llevar a cabo inferencia variacional ya que permiten expresarla como una integral sobre ellas.

La pregunta que surge es: ¿qué ocurre si empleamos este procedimiento en la formulación de un GP para MIL? La respuesta a esta pregunta es el modelo *Pólya-Gamma Variational Gaussian Process Multiple Instance Learning* PG-VGPMIL. Por primera vez en la literatura, empleamos variables aleatorias Pólya-Gamma para aumentar la probabilidad conjunta y construir un nuevo modelo. Estas variables permiten obtener las distribuciones variacionales sin emplear ninguna desigualdad.

Definición 4.1 (Pólya-Gamma, [32])**.** Decimos que una variable aleatoria ω sigue una distribución Pólya-Gamma con parámetros $b \in \mathbb{R}^+$ y $c \in \mathbb{R}$, y lo denotamos por $\omega \sim \mathrm{PG}(b, c)$, si

$$\omega \overset{D}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k-\frac{1}{2})^2 + \frac{c^2}{4\pi^2}},$$

donde $g_k \sim \mathrm{Gamma}(b, 1)$ son variables aleatorias independientes que siguen una distribución Gamma, y $\overset{D}{=}$ indica que las distribuciones de ambas variables aleatorias son iguales.

Obsérvese que hemos definido las nuevas variables exactamente como se hizo originalmente. Para este trabajo estamos interesados en las propiedades que se enuncian en el siguiente resultado. La prueba de las mismas puede consultarse en [32].

Teorema 4.1 (Propiedades de la distribución Pólya-Gamma, [32])**.**

1. *Sea* $\omega \sim \mathrm{PG}(b, 0)$. *Entonces*

$$\mathbb{E}_{\mathrm{PG}(\omega|b,0)}[\exp(-\omega t)] = \frac{1}{\cosh^b(\sqrt{t/2})}.$$

2. *Se cumple la siguiente igualdad,*

$$\mathrm{PG}(\omega \mid b, c) = \cosh^b\left(\frac{c}{2}\right) \exp\left(-\frac{c^2}{2}\omega\right) \mathrm{PG}(\omega \mid b, 0)$$

3. *Si denotamos a la función logística por* σ, *para cada* $x \in \mathbb{R}$ *se cumple*

$$\sigma(x) = \frac{1}{2}\mathbb{E}_{\mathrm{PG}(\omega|1,0)}\left[\exp\left(\frac{x}{2} - \frac{x^2}{2}\omega\right)\right]. \tag{4.10}$$

La igualdad (4.10) nos permitirá incorporar las variables aleatorias Pólya-Gamma en nuestro modelo de observación. Para ello, basta escribir $p(y_n \mid f_n) = \sigma\left((2y_n - 1)f_n\right)$ y aplicar (4.10) considerando $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_N]$ con $\omega_n \sim \mathrm{PG}(1, 0)$,

$$p(\boldsymbol{y} \mid \boldsymbol{f}) = \prod_n \frac{1}{2}\mathbb{E}_{p(\omega_n)}\left[\exp\left(\frac{(2y_n - 1)f_n}{2} - \frac{f_n^2}{2}\omega_n\right)\right].$$

Condicionando a $\boldsymbol{\omega}$ se obtiene

$$p(\boldsymbol{y} \mid \boldsymbol{f}, \boldsymbol{\omega}) \propto \exp\left(\sum_n (y_n - 1/2) f_n - \frac{f_n^2}{2}\omega_n\right) = \exp\left((\boldsymbol{y} - 1/2)^\top \boldsymbol{f} - \frac{1}{2}\boldsymbol{f}^\top \boldsymbol{\Omega} \boldsymbol{f}\right),$$

siendo $\boldsymbol{\Omega} = \text{diag}(\omega_1, \ldots, \omega_N)$. Llegamos así al nuevo modelo PG-VGPMIL,

$$\begin{aligned}
p(\boldsymbol{u} \mid \boldsymbol{Z}) &= \mathcal{N}(\boldsymbol{u} \mid \boldsymbol{0}, \boldsymbol{K}_{\boldsymbol{ZZ}}) \\
p(\boldsymbol{f} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{u}) &= \mathcal{N}(\boldsymbol{f} \mid \boldsymbol{a}, \widetilde{\boldsymbol{K}}), \\
p(\boldsymbol{y} \mid \boldsymbol{f}, \boldsymbol{\omega}) &\propto \exp\left((\boldsymbol{y} - 1/2)^\top \boldsymbol{f} - \frac{1}{2}\boldsymbol{f}^\top \boldsymbol{\Omega} \boldsymbol{f}\right), \\
p(\boldsymbol{\omega}) &= \prod_{n=1}^{N} p(\omega_n), \quad \omega_n \sim \text{PG}(1, 0) \\
p(\boldsymbol{T} \mid \boldsymbol{y}) &= \prod_{b=1}^{B} \frac{H^{G_b}}{H+1}.
\end{aligned}$$

Una representación gráfica del mismo se encuentra en la Figura 4.2. Obsérvese que el modelo de observación no ha cambiado ya que

$$p(y_n \mid f_n) = \int p(y_n \mid f_n, \omega_n) p(\omega_n) \mathrm{d}\omega_n = \sigma\left((2y_n - 1) f_n\right).$$

Además, al marginalizar en la distribución conjunta sobre $\boldsymbol{\omega}$ recuperamos el modelo VGPMIL.

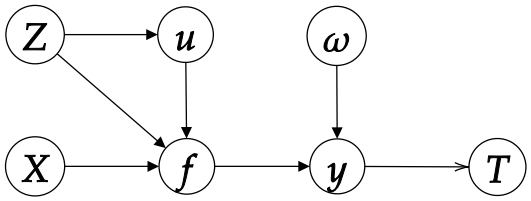


Figura 4.2.: Diagrama del modelo probabilístico PG-VGPMIL.

4.3.1. Inferencia

A diferencia de VGPMIL, ahora la distribución a posteriori debe tener en cuenta las nuevas variables $\boldsymbol{\omega}$. En lo que sigue omitiremos la dependencia respecto de $\boldsymbol{X}$ y $\boldsymbol{Z}$. Para obtener la aproximación a posteriori $p(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}, \boldsymbol{\omega} \mid \boldsymbol{T})$ vamos a emplear inferencia variacional. Empleamos el enfoque *Mean Field* considerando la distribución variacional

$$\begin{aligned}
Q(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}, \boldsymbol{\omega}) &= q(\boldsymbol{u}) p(\boldsymbol{f} \mid \boldsymbol{u}) q(\boldsymbol{y}) q(\boldsymbol{\omega}), \quad \text{donde} \\
q(\boldsymbol{y}) &= \prod_{n=1}^{N} q(y_n), \\
q(\boldsymbol{\omega}) &= \prod_{n=1}^{N} q(\omega_n).
\end{aligned}$$

Buscamos minimizar la divergencia de Kullback-Leibler,

$$\begin{aligned} Q^*(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}, \boldsymbol{\omega}) &= q^*(\boldsymbol{u}) p(\boldsymbol{f} \mid \boldsymbol{u}) q^*(\boldsymbol{y}) q^*(\boldsymbol{\omega}) = \\ &= \arg\min_{Q} \mathrm{KL}\left(Q(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}, \boldsymbol{\omega}) \mid\mid p(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}, \boldsymbol{\omega} \mid \boldsymbol{T})\right). \end{aligned}$$

Como sabemos, la solución a este problema viene dada, para cada factor q^* de Q^*, por (3.4), que en nuestro caso se traduce en

$$\log q^* = \mathbb{E}_{Q^*/q^*}\left[\log p(\boldsymbol{T}, \boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}, \boldsymbol{\omega})\right] + \text{const}, \tag{4.11}$$

donde Q^*/q^* se refiere al cociente entre la distribución Q^* y el factor q^*. Es decir, de Q^* eliminamos el factor q^*. En la práctica, esto se lleva a cabo iterativamente, actualizando los parámetros de un factor manteniendo el resto fijos. El siguiente resultado recoge la forma de las distribuciones y es consecuencia de aplicar (4.11).

Resultado 4.2. *Sea $p(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}, \boldsymbol{\omega}, \boldsymbol{T})$ la distribución conjunta dada por el modelo PG-VGPMIL. Entonces la solución óptima al problema*

$$\min_{q(\boldsymbol{u})q(\boldsymbol{y})q(\boldsymbol{\omega})} \mathrm{KL}\left(q(\boldsymbol{u})p(\boldsymbol{f} \mid \boldsymbol{u})q(\boldsymbol{y})q(\boldsymbol{\omega}) \mid\mid p(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}, \boldsymbol{\omega} \mid \boldsymbol{T})\right)$$

viene dada por las siguientes distribuciones:

$$\begin{aligned} q^*(\boldsymbol{u}) &= \mathcal{N}\left(\boldsymbol{u} \mid \boldsymbol{m}, \boldsymbol{S}\right), \\ q^*(y_n) &= \text{Bernouilli}\left(y_n \mid \pi_n\right), \\ q^*(\omega_n) &= \text{PG}\left(\omega_n \mid 1, c_n\right), \end{aligned}$$

siendo

$$\begin{aligned} \pi_n &= \sigma\left[\boldsymbol{\kappa}_n \boldsymbol{m} + \log H\left(2T_b - 1\right)\left(1 - \mathbb{E}\left[\max\left\{\boldsymbol{y}\right\}_{b \setminus n}\right]\right)\right], \\ \boldsymbol{m} &= \boldsymbol{S}\boldsymbol{K}_{\boldsymbol{ZZ}}^{-1}\boldsymbol{K}_{\boldsymbol{ZX}}\left(\boldsymbol{\pi} - \frac{1}{2}\right), \\ \boldsymbol{S} &= \left(\boldsymbol{K}_{\boldsymbol{ZZ}}^{-1}\boldsymbol{K}_{\boldsymbol{ZX}}\boldsymbol{\Theta}\boldsymbol{K}_{\boldsymbol{XZ}}\boldsymbol{K}_{\boldsymbol{ZZ}}^{-1} + \boldsymbol{K}_{\boldsymbol{ZZ}}^{-1}\right)^{-1}, \\ c_n &= \sqrt{\widetilde{\boldsymbol{K}}_{nn} + \boldsymbol{\kappa}_n\left(\boldsymbol{m}\boldsymbol{m}^\top + \boldsymbol{S}\right)\boldsymbol{\kappa}_n^\top}, \end{aligned}$$

donde $\boldsymbol{\kappa}_n = \boldsymbol{K}_{\boldsymbol{x}_n\boldsymbol{Z}}\boldsymbol{K}_{\boldsymbol{ZZ}}^{-1} = \mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}\left[f_n\right]$, $\boldsymbol{\Theta} = \text{diag}\left(\theta(c_1), \ldots, \theta(c_N)\right)$ *y* $\theta(c) = \tanh(c/2)/(2c)$.

Demostración. Calculamos la solución óptima para cada una de las distribuciones fijando el resto,

$$\log q^* = \mathbb{E}_{Q^*/q^*}\left[\log p(\boldsymbol{T}, \boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}, \boldsymbol{\omega})\right] + \text{const}.$$

Actualización de $q(y_n)$. Denotamos $q(y_{j\neq n}) = \prod_{j\neq n} \pi_j^{y_j}(1-\pi_j)^{y_j}$. Fijado n, sea b el índice de

la bolsa en la que se encuentra la instancia n. Se tiene:

$$\log q^*(y_n) = \underbrace{\mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})q(y_{j\neq n})q(\boldsymbol{\omega})}\left[\log p(\boldsymbol{y} \mid \boldsymbol{f}, \boldsymbol{\omega})\right]}_{A_1} + \\ + \underbrace{\mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})q(y_{j\neq n})q(\boldsymbol{\omega})}\left[\log p(\boldsymbol{T} \mid \boldsymbol{y})\right]}_{A_2} + \text{const}.$$

Analizamos cada sumando por separado,

$$\begin{aligned} A_1 &= \mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})q(y_{j\neq n})}\left[(\boldsymbol{y} - 1/2)^\top \boldsymbol{f}\right] + \text{const} = \\ &= \mathbb{E}_{q(y_{j\neq n})}\left[\boldsymbol{y}\right]^\top \mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}\left[\boldsymbol{f}\right] + \text{const} \\ &= y_n \mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}\left[f_n\right] + \text{const} = \\ &= y_n \underbrace{\boldsymbol{K}_{\boldsymbol{x}_i\boldsymbol{Z}}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}}_{\boldsymbol{\kappa}_n} \boldsymbol{m} + \text{const} \\ A_2 &= \log H \mathbb{E}_{q(y_{j\neq n})}\left[G_b\right] + \text{const} = \\ &= \log H(2T_b - 1)\mathbb{E}_{q(y_{j\neq n})}\left[\max\left\{\boldsymbol{y}\right\}_b\right] + \text{const} = \\ &= y_n \log H(2T_b - 1)\left(1 - \mathbb{E}\left[\max\left\{\boldsymbol{y}\right\}_{b\setminus n}\right]\right) + \text{const}, \end{aligned}$$

donde hemos empleado el Lema 4.2 para calcular la esperanza involucrada y hemos aproximado $\max \mathbb{E}\left[x\right] \approx \mathbb{E}\left[\max x\right]$. Juntando todo nos queda,

$$\log q^*(y_n) = y_n\left[\boldsymbol{\kappa}_n\boldsymbol{m} + \log H(2T_b - 1)\left(1 - \mathbb{E}\left[\max\left\{\boldsymbol{y}\right\}_{b\setminus n}\right]\right)\right] + \text{const},$$

de donde deducimos que $q(y_n) = \text{Bernouilli}\left(y_n \mid \pi_n\right)$ con

$$\pi_n = \sigma\left\{\boldsymbol{\kappa}_n\boldsymbol{m} + \log H(2T_b - 1)\left(1 - \mathbb{E}\left[\max\left\{\boldsymbol{y}\right\}_{b\setminus n}\right]\right)\right\}$$

Actualización de $q(\boldsymbol{u})$.

$$\log q^*(\boldsymbol{u}) = \underbrace{\log p(\boldsymbol{u})}_{B_1} + \underbrace{E_{p(\boldsymbol{f}|\boldsymbol{u})}\left[\log p(\boldsymbol{f} \mid \boldsymbol{u})\right]}_{\text{const}} + \underbrace{\mathbb{E}_{p(\boldsymbol{f}|\boldsymbol{u})q(\boldsymbol{y})q(\boldsymbol{\omega})}\left[\log p(\boldsymbol{y}|\boldsymbol{f}, \boldsymbol{\omega})\right]}_{B_2} + \text{const}$$

Analizamos cada sumando por separado,

$$B_1 = -\frac{1}{2}\boldsymbol{u}^\top \boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}\boldsymbol{u} + \text{const}$$

$$
\begin{aligned}
B_2 &= \mathbb{E}_{q(\boldsymbol{y})p(\boldsymbol{f}|\boldsymbol{u})q(\boldsymbol{\omega})}\left[(\boldsymbol{y}-1/2)^\top \boldsymbol{f} - \frac{1}{2}\boldsymbol{f}^\top \boldsymbol{\Omega}\boldsymbol{f}\right] + \text{const} = \\
&= \mathbb{E}_{p(\boldsymbol{f}|\boldsymbol{u})q(\boldsymbol{\omega})}\left[(\boldsymbol{\pi}-1/2)^\top \boldsymbol{f} - \frac{1}{2}\boldsymbol{f}^\top \boldsymbol{\Omega}\boldsymbol{f}\right] + \text{const} = \\
&= \mathbb{E}_{q(\boldsymbol{\omega})}\left[(\boldsymbol{\pi}-1/2)^\top \boldsymbol{K_{XZ}}\boldsymbol{K_{ZZ}^{-1}}\boldsymbol{u} - \frac{1}{2}\left(\text{tr}(\boldsymbol{\Omega}\widetilde{\boldsymbol{K}}) + \boldsymbol{u}^\top \boldsymbol{K_{ZZ}^{-1}}\boldsymbol{K_{ZX}}\boldsymbol{\Omega}\boldsymbol{K}_{X\boldsymbol{Z}}\boldsymbol{K_{ZZ}^{-1}}\boldsymbol{u}\right)\right] + \text{const} = \\
&= (\boldsymbol{\pi}-1/2)^\top \boldsymbol{K_{XZ}}\boldsymbol{K_{ZZ}^{-1}}\boldsymbol{u} - \frac{1}{2}\left(\text{tr}(\boldsymbol{\Theta}\widetilde{\boldsymbol{K}}) + \boldsymbol{u}^\top \boldsymbol{K_{ZZ}^{-1}}\boldsymbol{K_{ZX}}\boldsymbol{\Theta}\boldsymbol{K_{XZ}}\boldsymbol{K_{ZZ}^{-1}}\boldsymbol{u}\right) + \text{const} \\
&= (\boldsymbol{\pi}-1/2)^\top \boldsymbol{K_{XZ}}\boldsymbol{K_{ZZ}^{-1}}\boldsymbol{u} - \frac{1}{2}\left(\text{tr}(\boldsymbol{\Theta}\widetilde{\boldsymbol{K}}) + \boldsymbol{u}^\top \boldsymbol{K_{ZZ}^{-1}}\boldsymbol{K_{ZX}}\boldsymbol{\Theta}\boldsymbol{K_{XZ}}\boldsymbol{K_{ZZ}^{-1}}\boldsymbol{u}\right) + \text{const},
\end{aligned}
$$

donde $\boldsymbol{\Theta} = \text{diag}(\theta_1, \ldots, \theta_N)$, siendo $\theta_n = \mathbb{E}_{q(\omega_n)}[\omega_n]$. Juntando todo:

$$
\log q^*(\boldsymbol{u}) = (\boldsymbol{\pi}-1/2)^\top \boldsymbol{K_{XZ}}\boldsymbol{K_{ZZ}^{-1}}\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}^\top \left(\boldsymbol{K_{ZZ}^{-1}}\boldsymbol{K_{ZX}}\boldsymbol{\Theta}\boldsymbol{K_{XZ}}\boldsymbol{K_{ZZ}^{-1}} + \boldsymbol{K_{ZZ}^{-1}}\right)\boldsymbol{u} + \text{const}
$$

Esto prueba que $q(\boldsymbol{u}) = N(\boldsymbol{u} \mid \boldsymbol{m}, \boldsymbol{S})$ con

$$
\begin{aligned}
\boldsymbol{m} &= \boldsymbol{S}\boldsymbol{K_{ZZ}^{-1}}\boldsymbol{K_{ZX}}(\boldsymbol{\pi}-1/2) \\
\boldsymbol{S} &= \left(\boldsymbol{K_{ZZ}^{-1}}\boldsymbol{K_{ZX}}\boldsymbol{\Theta}\boldsymbol{K_{XZ}}\boldsymbol{K_{ZZ}^{-1}} + \boldsymbol{K_{ZZ}^{-1}}\right)^{-1}
\end{aligned}
$$

Actualización de $q(\omega_n)$. Fijamos n. Denotamos $q(\omega_{j\neq n}) = \prod_{j\neq n} q(\omega_j)$.

$$
\log q^*(\omega) = \underbrace{\mathbb{E}_{q(\omega_{j\neq n})}[\log p(\boldsymbol{\omega})]}_{C_1} + \underbrace{\mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})q(\boldsymbol{y})q(\omega_{j\neq n})}[\log p(\boldsymbol{y} \mid \boldsymbol{f}, \boldsymbol{\omega})]}_{C_2}
$$

$$
\begin{aligned}
C_1 &= \int \prod_{j\neq n} q(\omega_j)\left(\sum_i \log p(\omega_i)\right) = \log p(\omega_n) + \sum_{j\neq n}\mathbb{E}_{q(\omega_j)}[\log p(\omega_j)] \\
C_2 &= \mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})q(\boldsymbol{y})q(\omega_{j\neq n})}\left[\sum_i \log p(y_i \mid \boldsymbol{f}_i, \omega_i)\right] + \text{const} = \\
&= \mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}[\log p(y_n \mid f_n, \omega_n)] + \text{const} = \\
&= \mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}\left[-\frac{1}{2}\omega_n f_n^2\right] + \text{const} = \\
&= -\frac{1}{2}\omega_n \mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}\left[f_n^2\right] + \text{const} \\
&= -\frac{1}{2}\omega_n\left[\widetilde{\boldsymbol{K}}_{nn} - \boldsymbol{\kappa}_n\left(\boldsymbol{S} + \boldsymbol{m}\boldsymbol{m}^\top\right)\boldsymbol{\kappa}_n^\top\right] + \text{const}
\end{aligned}
$$

Por tanto nos queda:

$$
\log q^*(\omega_n) = -\frac{1}{2}\omega_n\left[\widetilde{\boldsymbol{K}}_{nn} + \boldsymbol{\kappa}_n\left(\boldsymbol{S} + \boldsymbol{m}\boldsymbol{m}^\top\right)\boldsymbol{\kappa}_n^\top\right] + \log p(\omega_n) + \text{const},
$$

de donde deducimos que $q(\omega_n) = \text{PG}(\omega_n \mid 1, c_n)$ con $c_n = \sqrt{\widetilde{\boldsymbol{K}}_{nn} + \boldsymbol{\kappa}_n \left(\boldsymbol{S} + \boldsymbol{m}\boldsymbol{m}^\top\right) \boldsymbol{\kappa}_n^\top}$. Como consecuencia $\theta_n = \tanh\left(c_n/2\right)/2c_n$. □

Si uno revisa la prueba del Resultado 4.2 se da cuenta que las nuevas variables intervienen sólo a la hora de calcular la distribución variacional de $q(\boldsymbol{u})$. Estas variables se introducen para sustituir el cálculo de $\mathbb{E}_{p(\boldsymbol{f}|\boldsymbol{u})q(\boldsymbol{y})}\left[p(\boldsymbol{y} \mid \boldsymbol{f})\right]$ por el de $\mathbb{E}_{p(\boldsymbol{f}|\boldsymbol{u})q(\boldsymbol{y})}\left[p(\boldsymbol{y} \mid \boldsymbol{f}, \boldsymbol{\omega})\right]$. Alternativamente, VGPMIL aproxima la función logística, que es la responsable de que $\mathbb{E}_{p(\boldsymbol{f}|\boldsymbol{u})q(\boldsymbol{y})}\left[p(\boldsymbol{y} \mid \boldsymbol{f})\right]$ no admita una expresión cerrada.

4.3.2. Equivalencia con VGPMIL

Si comparamos los resultados 4.2 y 4.1 nos damos cuenta de que ambos modelos usan actualizaciones muy parecidas y sólo se diferencian en las matrices $\boldsymbol{\Lambda}$ y $\boldsymbol{\Theta}$. Pero estas matrices, en realidad, son las mismas. Basta tener en cuenta que $2\sigma(x) - 1 = \tanh\left(x/2\right)$, por lo que $\boldsymbol{\Lambda} = \boldsymbol{\Theta}$. Llegamos así al resultado clave de esta sección, que es la verdadera contribución de PG-VGPMIL.

Resultado 4.3. *VGPMIL y PG-VGPMIL son equivalentes, en el sentido de que las actualizaciones para obtener las distribuciones óptimas usando la aproximación Mean Field a la hora de emplear inferencia variacional son las mismas.*

Obsérvese que la equivalencia no se da en un sentido estricto. Lo que estamos afirmando es que ambos modelos son equivalentes si las distribuciones variacionales factorizan como indica la aproximación Mean Field. Pudiera ocurrir que la optimización de estas distribuciones usando otro método, como por ejemplo gradiente descendente, condujera a otras formas funcionales.

4.4. G-VGPMIL: modelo basado en variables Gamma

Hasta ahora hemos obtenido dos modelos equivalentes por caminos distintos. PG-VGPMIL emplea *data augmentation*, modificando la probabilidad conjunta para introducir las nuevas variables Pólya-Gamma. Intuitivamente, podemos decir que se pivota en estas nuevas variables para poder obtener expresiones cerradas de las distribuciones variacionales. Por su parte, VGPMIL emplea una aproximación cuadrática a la función logística. Obsérvese que gracias a ser cuadrática podemos calcular la esperanza de la cota obtenida respecto de una distribución normal. Lo que hemos obtenido es que esta aproximación es equivalente a introducir variables Pólya-Gamma.

Una pregunta natural, y es la que nosotros nos hacemos, es: ¿qué ocurre si cambiamos la distribución de las variables $\boldsymbol{\omega}$? De ella surgen muchas otras cuestiones, como ¿qué distribución elegimos? o ¿podríamos interpretar que estamos aproximando la función logística de otra forma? Estas preguntas conducen a la segunda contribución del trabajo.

Motivados por la Definición 4.1 de las variables Pólya-Gamma, en principio vamos a pedirle a las variables ω_n que sigan una distribución Gamma. La justificación sigue a continuación. Obsérvese que una variable $\omega \sim \text{PG}(1, 0)$ se define como una suma ponderada de variables $\text{Gamma}(1, 1)$. Esto es,

$$\omega = \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k - \frac{1}{2})^2},$$

con $g_k \sim \text{Gamma}(1,1)$. Si todas las variables fueran iguales a $g \sim \text{Gamma}(1,1)$, ocurriría

$$\omega = \frac{g}{2\pi^2} \underbrace{\sum_{k=1}^{\infty} \frac{1}{(k-\frac{1}{2})^2}}_{\pi^2/2} = \frac{g}{4},$$

de donde $\omega \sim \text{Gamma}(1,4)$. De forma general, supondremos $\omega \sim \text{Gamma}(1,k)$ y estudiaremos el valor de k. El nuevo modelo, que llamamos G-VGPMIL, se especifica como sigue,

$$\begin{aligned}
p(\boldsymbol{u} \mid \boldsymbol{Z}) &= \mathcal{N}(\boldsymbol{u} \mid \boldsymbol{0}, \boldsymbol{K_{ZZ}}) \\
p(\boldsymbol{f} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{u}) &= \mathcal{N}(\boldsymbol{f} \mid \boldsymbol{a}, \widetilde{\boldsymbol{K}}), \\
p(\boldsymbol{y} \mid \boldsymbol{f}, \boldsymbol{\omega}) &\propto \exp\left((\boldsymbol{y} - 1/2)^\top \boldsymbol{f} - \frac{1}{2} \boldsymbol{f}^\top \boldsymbol{\Omega} \boldsymbol{f} \right), \\
p(\boldsymbol{\omega}) &= \prod_{n=1}^{N} p(\omega_n), \quad \omega_n \sim \text{Gamma}(1,k) \\
p(\boldsymbol{T} \mid \boldsymbol{y}) &= \prod_{b=1}^{B} \frac{H^{G_b}}{H+1}.
\end{aligned}$$

Por primera vez en la literatura, usamos variables aleatorias Gamma para aumentar el modelo probabilístico y poder realizar inferencia con GPs en el escenario MIL. A continuación veremos que el proceso de inferencia, a pesar de ser análogo a los anteriores, conduce a un nuevo modelo con nuevas características.

4.4.1. Inferencia

El proceso de inferencia es completamente análogo al de PG-VGPMIL. No vamos a repetir toda la formulación, sino que especificamos directamente la forma que toma cada una de las distribuciones variacionales.

Resultado 4.4. *Sea $p(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}, \boldsymbol{\omega}, \boldsymbol{T})$ la distribución conjunta dada por el modelo G-VGPMIL. Entonces la solución óptima al problema*

$$\underset{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})q(\boldsymbol{y})q(\boldsymbol{\omega})}{\arg\min} \quad \text{KL}\left(q(\boldsymbol{u})p(\boldsymbol{f} \mid \boldsymbol{u})q(\boldsymbol{y})q(\boldsymbol{\omega}) \mid\mid p(\boldsymbol{u}, \boldsymbol{f}, \boldsymbol{y}, \boldsymbol{\omega} \mid \boldsymbol{T})\right)$$

viene dada por las siguientes distribuciones:

$$\begin{aligned}
q^*(\boldsymbol{u}) &= \mathcal{N}(\boldsymbol{u} \mid \boldsymbol{m}, \boldsymbol{S}), \\
q^*(y_n) &= \text{Bernouilli}(y_n \mid \pi_n), \\
q^*(\omega_n) &= \text{Gamma}(\omega_n \mid 1, \beta_n),
\end{aligned}$$

siendo

$$\pi_n = \sigma\left[\boldsymbol{\kappa}_n \boldsymbol{m} + \log H\left(2T_b - 1\right)\left(1 - \mathbb{E}\left[\max\left\{\boldsymbol{y}\right\}_{b\setminus n}\right]\right)\right],$$
$$\boldsymbol{m} = \boldsymbol{S}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{X}}\left(\boldsymbol{\pi} - \frac{1}{2}\right),$$
$$\boldsymbol{S} = \left(\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{X}}\boldsymbol{\Theta}\boldsymbol{K}_{\boldsymbol{X}\boldsymbol{Z}}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1} + \boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}\right)^{-1},$$
$$\beta_n = k + \frac{\mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}\left[f_n^2\right]}{2} = k + \frac{\widetilde{\boldsymbol{K}}_{nn} + \boldsymbol{\kappa}_n\left(\boldsymbol{m}\boldsymbol{m}^\top + \boldsymbol{S}\right)\boldsymbol{\kappa}_n^\top}{2},$$

donde $\boldsymbol{\kappa}_n = \mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}\left[f_n\right] = \boldsymbol{K}_{\boldsymbol{x}_n\boldsymbol{Z}}\boldsymbol{K}_{\boldsymbol{Z}\boldsymbol{Z}}^{-1}$, $\boldsymbol{\Theta} = \operatorname{diag}\left(\theta(\beta_1), \ldots, \theta(\beta_N)\right)$ *y* $\theta(\beta) = 1/\beta$.

Demostración. Las pruebas de las diferentes actualizaciones son muy similares a las de PG-VGPMIL. La de $q(y_n)$ es la misma y no la vamos a repetir. En la actualización de $q(\boldsymbol{u})$ debemos observar que la esperanza de ω_n cambia. La actualización de $q(\omega_n)$ es como sigue,

$$\log q^*(\omega_n) = \underbrace{\mathbb{E}_{q(\omega_j \neq n)}\left[\log p(\boldsymbol{\omega})\right]}_{A_1} + \underbrace{\mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})q(\boldsymbol{y})q(\omega_{j\neq n})}\left[\log p(y \mid \boldsymbol{f}, \omega)\right]}_{A_2} + \text{const}$$

Analizamos cada término por separado,

$$\begin{aligned}
A_1 &= \log p(\omega_n) + \sum_{j\neq n}\mathbb{E}_{q(\omega_j)}\left[\log p(\omega_j)\right] = \\
&= \log(k) - k\omega_n + \text{const} \\
A_2 &= \mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})q(\boldsymbol{y})q(\omega_{j\neq n})}\left[\sum_i \log p(y_i \mid f_i, \omega_i)\right] + \text{const} = \\
&= \mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}\left[\log p(y_n \mid f_n, \omega_n)\right] + \text{const} = \\
&= \mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}\left[-\frac{1}{2}\omega_n f_n^2\right] + \text{const} = \\
&= -\omega_n\frac{\mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}\left[f_n^2\right]}{2} + \text{const}
\end{aligned}$$

Por tanto nos queda:

$$\log q^*(\omega) = -\omega_n\left(k + \frac{\mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}\left[f_n^2\right]}{2}\right) + \text{const}$$

de donde deducimos que $q(\omega_n) = \text{Gamma}(\omega_n \mid 1, \beta_n)$ con

$$\beta_n = k + \frac{\mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}\left[f_n^2\right]}{2} = k + \frac{\widetilde{\boldsymbol{K}}_{nn} + \boldsymbol{\kappa}_n(\boldsymbol{S} + \boldsymbol{m}\boldsymbol{m}^\top)\boldsymbol{\kappa}_n^\top}{2}.$$

Por tanto $\theta_n = \mathbb{E}_{q(\omega_n)}\left[\omega_n\right] = 1/\beta_n$. □

El modelo G-VGPMIL guarda cierto parecido con sus predecesores. Dedicamos la siguiente subsección a comprender las diferencias entre ellos.

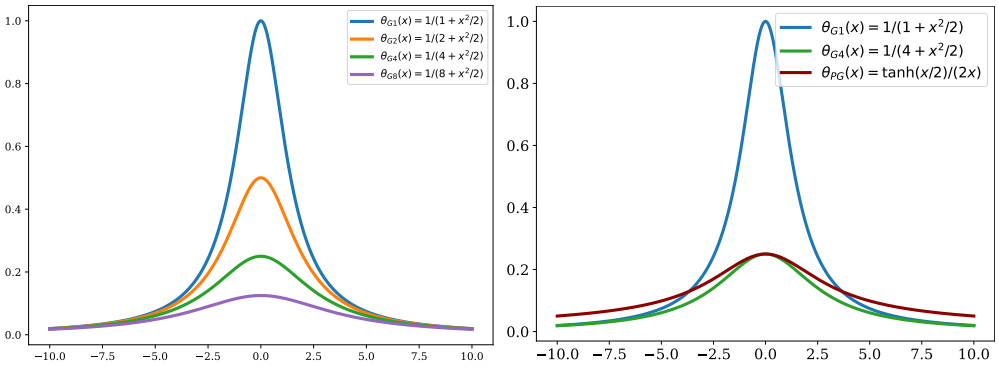
4.4.2. Sobre la función θ y el parámetro k

Al comparar las actualizaciones de PG-VGPMIL y G-VGPMIL es fácil darse cuenta de que sólo se producen cambios en la matriz diagonal $\boldsymbol{\Theta}$. Esta matriz se forma a partir de la evaluación de una función θ. En PG-VGPMIL esta función se evalúa sobre los parámetros c_n, mientras que en G-VGPMIL se evalúa sobre β_n. Sin embargo, en ambos casos esta es una función de $\sqrt{\mathbb{E}_{q(\boldsymbol{u})p(\boldsymbol{f}|\boldsymbol{u})}[f_n^2]}$. Podemos escribir

$$\theta_{\mathrm{PG}}(x) = \frac{\tanh(x/2)}{2x}, \tag{4.12}$$

$$\theta_{\mathrm{G}k}(x) = \frac{1}{k + x^2/2}, \tag{4.13}$$

para las funciones de PG-VGPMIL y G-VGPMIL (con parámetro k), respectivamente. Podemos analizar las diferencias entre ambos comprobando en qué se diferencian estas aplicaciones. En la Figura 4.3(a) se representa $\theta_{\mathrm{G}k}$ para distintos valores de k. Obsérvese que $\theta_{\mathrm{G}k}$ alcanza un máximo absoluto en cero con valor $1/k$ por lo que al variar el valor de k controlamos el valor máximo que alcanza esta función.



(a) Representación de la función $\theta_{\mathrm{G}k}$ para $k \in \{1, 2, 4, 8\}$.

(b) Representación de las funciones θ_{PG}, θ_{G1} y θ_{G4}.

Figura 4.3.: Representación de las funciones θ_{PG} y $\theta_{\mathrm{G}k}$.

En la Figura 4.3(b) representamos θ_{PG} junto a θ_{G4}. No nos sorprende que al tomar $k = 4$ ambas funciones alcancen el mismo valor máximo. Recuérdese lo que ocurría si en la definición de una variable Pólya-Gamma todas las variables Gamma eran iguales.

Debemos señalar una idea importante en relación al papel que juegan estas funciones. Si tomásemos $\theta = 0$, estaríamos imponiendo que la matriz de covarianza de $q(\boldsymbol{u})$ fuese igual que la de la distribución a priori $p(\boldsymbol{u})$. De alguna forma, a través de la función θ podemos cuantificar cuánto se diferenciarán las distribuciones a priori y a posteriori, o dicho de otra forma, la incertidumbre que tenemos sobre estas distribuciones. La distribución que le damos a la variable aleatoria $\boldsymbol{\omega}$ nos permite cuantificar esta incertidumbre.

Para acabar el capítulo, ofrecemos una visión general de lo que hemos conseguido hasta ahora. El estudio del modelo VGPMIL revela que el uso de la desigualdad de Jaakkola puede conducir a soluciones subóptimas para las distribuciones variacionales. Para intentar solucionar esto, hemos usado por primera vez las variables aleatorias Pólya-Gamma en un nuevo modelo basado en GPs para MIL. Esto nos ha llevado a la primera contribución de este trabajo,

el modelo PG-VGPMIL. El resultado clave es la prueba de la equivalencia en el sentido Mean Field con VGPMIL. Sustituir las variables Pólya-Gamma por variables Gamma nos ha conducido a la segunda contribución del trabajo, que recibe el nombre de G-VGPMIL, y propone por primera vez el uso de variables Gamma en un modelo basado en GPs para MIL.

En el siguiente capítulo vamos a validar experimentalmente el comportamiento de G-VGPMIL, comparándolo con el ya existente VGPMIL. Para ello, nos enfrentaremos a tres problemas de clasificación binaria de MIL, siendo el último de ellos la detección de hemorragias intracraneales a partir de escáneres TAC.

5. Aplicaciones

En el capítulo anterior hemos estudiado tres modelos basados en GPs para problemas MIL de clasificación: VGPMIL, PG-VGPMIL y G-VGPMIL. VGPMIL ya estaba propuesto en la literatura, siendo el estado del arte a la hora de emplear el modelo de observación logístico. PG-VGPMIL y G-VGPMIL son las dos contribuciones que presenta este trabajo. Ambas están basadas en la inclusión de nuevas variables en la probabilidad conjunta. PG-VGPMIL es el primer modelo para problemas MIL de clasificación propuesto en la literatura basado en variables Pólya-Gamma. Hemos probado que es equivalente, en el sentido Mean Field, a VGPMIL. Por su parte, G-VGPMIL usa por primera vez variables Gamma y no es equivalente a los anteriores.

El objetivo de este capítulo es doble. Por una parte, evaluamos G-VGPMIL experimentalmente. VGPMIL ya está validado por la comunidad investigadora, por lo que aquí se usará para compararlo con G-VGPMIL. Como VGPMIL y PG-VGPMIL son equivalentes, sólo nos referiremos al primero de ellos. Por otra parte, resolvemos el problema de detección de hemorragias intracraneales que motiva este trabajo y constituye la tercera contribución del mismo.

5.1. Sobre los conjuntos de datos y la metodología empleada

La primera parte de este capítulo presenta un experimento desarrollado sobre un conjunto de datos sintético. Partiendo del clásico conjunto MNIST, diseñamos un conjunto para MIL repartiendo las imágenes en bolsas. Evaluaremos el funcionamiento de VGPMIL y G-VGPMIL de forma cuantitativa y cualitativa (mostrando qué predicciones son capaces de obtener).

En la segunda parte usamos dos conjuntos de MIL clásicos como son musk1 y musk2 para comparar ambos algoritmos. Con estos conjuntos tratamos el problema de clasificación de moléculas que explicábamos en el primer capítulo.

La tercera parte presenta el caso real de detección de hemorragias intracraneales. Usaremos dos conjuntos de imágenes de TAC (RSNA y CQ500) para desarrollar un modelo capaz de predecir la presencia de hemorragia en una persona a partir de bolsas de imágenes. Este experimento muestra que el nuevo modelo propuesto, G-VGPMIL, encaja de forma natural en las arquitecturas de Aprendizaje Profundo y son esenciales para mejorar los resultados que estas obtienen.

La información de cada uno de los conjuntos utilizados se encuentra en la Tabla 5.1. Puede observarse que son de muy diversa naturaleza. Como es normal, este tipo de problemas son profundamente desbalanceados a nivel de instancia ya que contamos con muchas más instancias negativas que positivas. A nivel de bolsa se encuentra mucho más equilibrado, aunque en escenarios reales suelen predominar las bolsas negativas.

En todos los casos entrenamos los modelos VGPMIL y G-VGPMIL con el núcleo RBF. El número de iteraciones no es fijo, ya que el proceso de entrenamiento actualiza los parámetros de las distribuciones hasta que la función de pérdida (ELBO) desciende menos de 10^{-4} durante 10 iteraciones. Para el número de inducing points consideramos los valores $\{50, 100, 200\}$, mientras que para el parámetro *lengthscale* del núcleo consideramos $\left\{1, (1+\sqrt{D})/2, \sqrt{D}, (\sqrt{D}+D)/2, D\right\}$,

	Núm. instancias	Instancias positivas	Instancias negativas	Núm. bolsas	Bolsas positivas	Bolsas negativas
MNIST	70000	6903	63097	7778	4741	3037
musk1	476	?	?	92	47	45
musk2	6598	?	?	102	39	63
RSNA	39750	5782	33968	1150	483	667
CQ500	193317	?	?	490	205	285

Tabla 5.1.: Distribución de las etiquetas a nivel de instancia y bolsa para cada conjunto de datos.

siendo D el número de rasgos o características. Finalmente, para el parámetro k de G-VGPMIL se consideran los valores $\{0.5, 1.0, 2.0, 4.0\}$. Los resultados que se presentan, a menos que se indique lo contrario, corresponden a la media y a la desviación están obtenida tras realizar validación cruzada de cinco capas.

Por último, aclaramos qué métricas vamos a considerar. Como ya hemos explicado, los problemas MIL de clasificación binaria tienen características especiales que debemos tener en cuenta a la hora de evaluar los clasificadores obtenidos. A nivel de instancia hay un gran desequilibrio, por lo que la métrica F1 tendrá una gran importancia. Junto a esta evaluaremos también la tasa de acierto o *accuracy*, el *log loss* y el área bajo la curva ROC (AUC). A nivel de bolsa obtendremos las mismas métricas, pero ahora F1 no será tan determinante.

5.2. Un ejemplo ilustrativo: MNIST

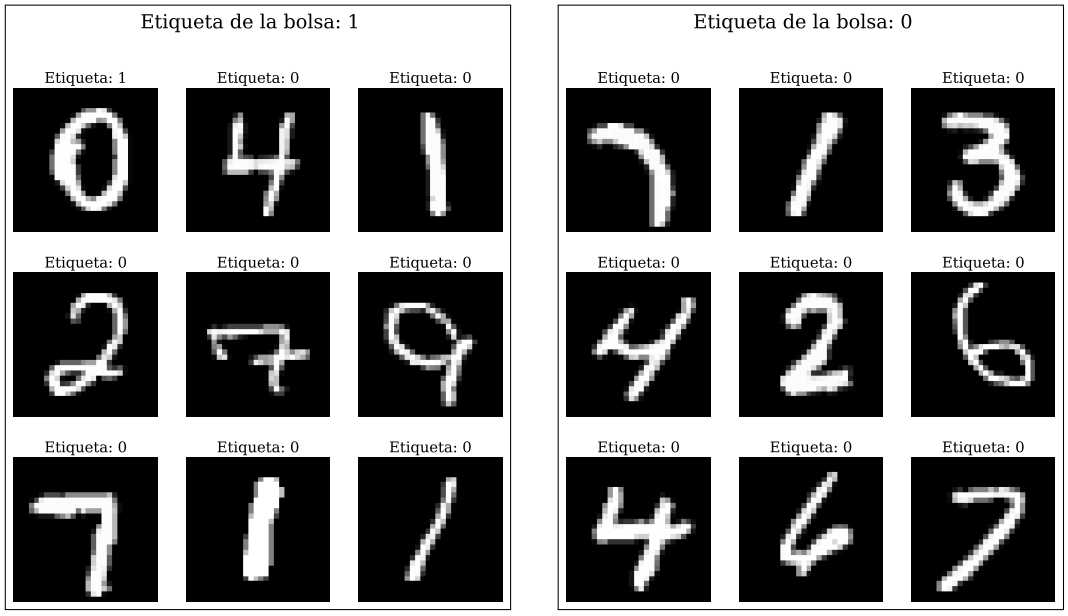
MNIST[1] es un conjunto clásico de Visión por Computador consistente en 70.000 imágenes de dígitos (del 0 al 9) manuscritos y etiquetados. Para obtener un conjunto apropiado para MIL repartimos aleatoriamente las imágenes en bolsas de 9 dígitos. Si una bolsa contiene un cero tendrá etiqueta positiva, y negativa en caso contrario. En la Figura 5.1 se muestran dos bolsas de ejemplo. Obtenemos así un conjunto con 7778 bolsas, de las cuales 4741 son positivas y 3037 son negativas. El objetivo de este experimento es ilustrar el correcto funcionamiento de VGPMIL y el nuevo modelo que proponemos, G-VGPMIL.

Dado que los ejemplos de MNIST son de alta dimensionalidad ($28 \times 28 = 784$) evaluamos el rendimiento de VGPMIL y G-VGPMIL sin aplicar ningún preprocesamiento (Tabla A.2) y aplicando PCA (*Principal Components Analysis*, [55]) para quedarnos con las 30 componentes principales (Tabla A.1). En las tablas mencionadas podemos comprobar que G-VGPMIL es un método muy competitivo con VGPMIL y que en algunas ocasiones mejora su rendimiento.

Los resultados obtenidos los hemos resumido en las tablas 5.2a y 5.2b para facilitar su discusión. Como esperábamos, G-VGPMIL presenta los mejores resultados cuando $k = 4$ (recuérdese que se obtenía $k = 4$ al sustituir las variables Gamma por una sóla variable Gamma en la definición de las variables Pólya-Gamma). Si no aplicamos PCA, G-VGPMIL obtiene mejores resultados en cuanto a la métrica AUC se refiere. Si aplicamos PCA, los papeles se intercambian y VGPMIL se comporta mejor. Por otra parte, en ambos modelos el rendimiento aumenta conforme crece el número de inducing points. Esto tiene sentido ya que cuantos más inducing points introduzcamos más información podemos extraer del conjunto de entrenamiento. En el caso de MNIST esto es especialmente beneficioso debido al elevado número de ejemplos de entrenamiento.

En la Figura 5.2 hemos representado las métricas del mejor clasificador de cada tipo. El

[1] http://yann.lecun.com/exdb/mnist/



(a) Ejemplo de una bolsa con etiqueta positiva. (b) Ejemplo de una bolsa con etiqueta negativa.

Figura 5.1.: Dos de las bolsas generadas a partir del conjunto MNIST.

mejor clasificador se elige teniendo en cuenta la métrica AUC a nivel de bolsa e instancia. Observamos que VGPMIL y G-VGPMIL obtienen resultados muy similares. Si no aplicamos PCA, G-VGPMIL se comporta mejor en las métricas AUC y Log Loss. Si aplicamos PCA ambos se encuentran muy igualados, aunque G-VGPMIL sigue presentando valores menores en Log Loss.

La Figura 5.3 muestra cómo los métodos estudiados permiten realizar predicciones a nivel de instancia proporcionando también el nivel de incertidumbre asociado a la predicción. Este es el tipo de comportamiento que podemos esperar de este tipo de clasificadores probabilísticos. Para cada bolsa que tengamos, podemos predecir la probabilidad de que cada una de sus instancias pertenezcan a la clase positiva o a la clase negativa. Junto a esta probabilidad obtendremos también la incertidumbre asociada a la predicción (4.9), la cual puede interpretarse como la seguridad del clasificador en su decisión. Obsérvese que en la subfigura 5.3a se asocia una incertidumbre mucho mayor al cero que al resto de instancias. Esto puede deberse a la forma extraña que presenta el dígito. Aquí se ve reflejada la utilidad de este tipo de clasificadores. Como las etiquetas a nivel de instancia suelen ser desconocidas, identificar las bolsas positivas de esta forma nos lleva a conocer qué instancias son positivas.

Los resultados obtenidos muestran que G-VGPMIL es un clasificador para MIL mucho más que aceptable y que es capaz de competir con el estado del arte. Hemos observado que presenta un comportamiento distinto a VGPMIL que aún debemos explorar en conjuntos más complejos. Además hemos ilustrado cómo interpretar las predicciones de estos modelos.

Num. Inducing	Lengthscale	k	Modelo	Inst. Acc.	Inst. F1	Inst. Log Loss	Inst. AUC	Bag Acc.	Bag F1	Bag Log Loss	Bag AUC
50	406.0	0.5	G-VGPMIL	0.9221 ± 0.0022	0.3471 ± 0.0269	0.6504 ± 0.0001	$\mathbf{0.976 \pm 0.0017}$	0.5624 ± 0.015	0.4397 ± 0.0303	0.6855 ± 0.0005	$\mathbf{0.9394 \pm 0.007}$
	28.0	-	VGPMIL	$\mathbf{0.958 \pm 0.0025}$	$\mathbf{0.7959 \pm 0.0132}$	$\mathbf{0.2749 \pm 0.0101}$	0.9679 ± 0.002	$\mathbf{0.8557 \pm 0.0103}$	$\mathbf{0.8842 \pm 0.0084}$	$\mathbf{0.426 \pm 0.0161}$	0.9338 ± 0.007
100	406.0	1.0	G-VGPMIL	0.9417 ± 0.0015	0.5813 ± 0.0211	0.6112 ± 0.0004	$\underline{\mathbf{0.9837 \pm 0.0015}}$	0.6999 ± 0.0146	0.6739 ± 0.0211	0.6698 ± 0.001	$\mathbf{0.9584 \pm 0.0036}$
	28.0	-	VGPMIL	$\mathbf{0.9684 \pm 0.0013}$	$\mathbf{0.8416 \pm 0.0086}$	$\mathbf{0.233 \pm 0.0071}$	0.9782 ± 0.0023	$\mathbf{0.8826 \pm 0.0044}$	$\mathbf{0.9043 \pm 0.0032}$	$\mathbf{0.3759 \pm 0.0132}$	0.9498 ± 0.0042
200	28.0	4.0	G-VGPMIL	$\underline{\mathbf{0.9757 \pm 0.0014}}$	$\underline{\mathbf{0.8694 \pm 0.0084}}$	$\underline{\mathbf{0.091 \pm 0.0046}}$	$\mathbf{0.9831 \pm 0.0025}$	0.9009 ± 0.0066	0.9151 ± 0.0069	$\underline{\mathbf{0.2617 \pm 0.0102}}$	$\underline{\mathbf{0.9687 \pm 0.0042}}$
	28.0	-	VGPMIL	0.9744 ± 0.0024	0.8691 ± 0.0104	0.1984 ± 0.004	0.983 ± 0.0034	$\underline{\mathbf{0.9038 \pm 0.0079}}$	$\underline{\mathbf{0.9205 \pm 0.0067}}$	0.3275 ± 0.0059	0.961 ± 0.0049

(a) Sin aplicar PCA

Num. Inducing	Lengthscale	k	Modelo	Inst. Acc.	Inst. F1	Inst. Log Loss	Inst. AUC	Bag Acc.	Bag F1	Bag Log Loss	Bag AUC
50	17.738	0.5	G-VGPMIL	0.9449 ± 0.0015	0.6124 ± 0.0124	0.6484 ± 0.0002	0.9923 ± 0.002	0.72 ± 0.0124	0.7017 ± 0.0174	0.6788 ± 0.0004	0.9835 ± 0.0026
	5.477	-	VGPMIL	$\mathbf{0.9855 \pm 0.0013}$	$\mathbf{0.9257 \pm 0.0077}$	$\mathbf{0.1414 \pm 0.0041}$	$\mathbf{0.993 \pm 0.0017}$	$\mathbf{0.9445 \pm 0.0072}$	$\mathbf{0.954 \pm 0.0058}$	$\mathbf{0.2353 \pm 0.0076}$	$\mathbf{0.9846 \pm 0.002}$
100	3.238	4.0	G-VGPMIL	0.9875 ± 0.0009	0.9333 ± 0.0051	$\mathbf{0.0509 \pm 0.0011}$	0.9961 ± 0.0006	0.9469 ± 0.0063	0.9547 ± 0.0056	$\mathbf{0.1598 \pm 0.0073}$	0.9902 ± 0.0011
	3.238	-	VGPMIL	$\mathbf{0.9909 \pm 0.0009}$	$\mathbf{0.9527 \pm 0.0059}$	0.1102 ± 0.0031	$\mathbf{0.9968 \pm 0.0007}$	$\mathbf{0.9645 \pm 0.0041}$	$\mathbf{0.9705 \pm 0.0034}$	0.1833 ± 0.0041	$\mathbf{0.9921 \pm 0.0011}$
200	3.238	4.0	G-VGPMIL	0.9897 ± 0.0005	0.9453 ± 0.0027	$\underline{\mathbf{0.0429 \pm 0.0008}}$	0.9974 ± 0.0003	0.9576 ± 0.0047	0.964 ± 0.0041	$\underline{\mathbf{0.1379 \pm 0.0068}}$	0.9934 ± 0.0012
	3.238	-	VGPMIL	$\underline{\mathbf{0.9924 \pm 0.0006}}$	$\underline{\mathbf{0.9607 \pm 0.0034}}$	0.0957 ± 0.0014	$\underline{\mathbf{0.9978 \pm 0.0004}}$	$\underline{\mathbf{0.9691 \pm 0.0032}}$	$\underline{\mathbf{0.9744 \pm 0.0027}}$	0.1597 ± 0.0025	$\underline{\mathbf{0.9944 \pm 0.0008}}$

(b) Aplicando PCA.

Tabla 5.2.: Resumen de los resultados en MNIST. Para cada valor del número de inducing points considerado, se recogen las métricas del mejor clasificador de cada tipo. Se señala en negrita el mejor valor de cada métrica dentro de cada grupo, y subrayado el mejor de cada columna.

5.3. Los conjuntos musk

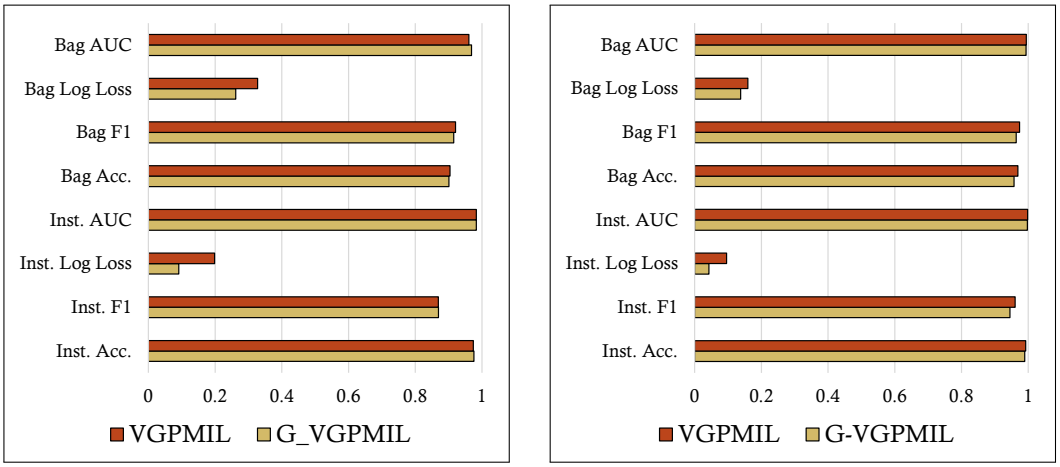
En esta sección vamos a trabajar en un escenario de MIL frecuentemente usado para evaluar modelos de MIL. Los conjuntos musk1[2] y musk2[3] están formados por datos tabulares. Cada fila (instancia) representa una de las posibles "formas" que puede adoptar una molécula. Las filas se agrupan en bolsas y cada bolsa conforma una molécula. El objetivo es predecir si esa molécula dará lugar a un almizcle o musco[4]. Esto queda determinado por las formas que puede adoptar esa molécula, por lo que habrá formas que den lugar a un musco (instancias positivas) y otras que no (instancias negativas). En estos conjuntos no se disponen de las etiquetas a nivel de instancia, por lo que sólo podemos evaluar los modelos a nivel de bolsa. Este ya no es un escenario sintético de MIL, que además tiene una gran utilidad. La identificación de bolsas positivas con modelos probabilísticos nos permite conocer qué tipo de formas pueden dar lugar a un musco.

Todos los resultados que vamos a discutir en esta sección se encuentra en las tablas A.3 y A.4. Nuevamente, nos centramos en los resultados del mejor clasificador (de acuerdo a las métricas AUC a nivel de instancia y de bolsa) de cada tipo para cada valor del número de *inducing points*:

- En musk1 (Tabla 5.3) ambos modelos son capaces de alcanzar el mismo rendimiento. Obsérvese de nuevo que predomina el valor de $k = 4$ para G-VGPMIL y que las métricas crecen conforme aumenta el número de *inducing points* considerado. Algo que llama la atención es que al fijar 200 *inducing points* el mejor valor de *lengthscale* es 1.0. Por otra parte, también debemos notar que aunque al usar 200 *inducing points* los clasificadores son mejores, los valores más bajos de Log Loss se consiguen al usar 100 *inducing points*.

- Si ahora nos fijamos en musk2 (Tabla 5.4) observamos un comportamiento totalmente distinto al que habíamos detectado hasta ahora. Ahora G-VGPMIL presenta mejores resultados cuando $k = 2$. Además, las métricas sólo mejoran al aumentar el número de

[2] https://archive.ics.uci.edu/ml/datasets/Musk+(Version+1)
[3] https://archive.ics.uci.edu/ml/datasets/Musk+(Version+2)
[4] https://es.wikipedia.org/wiki/Almizcle



(a) Sin aplicar PCA.

(b) Aplicando PCA.

Figura 5.2.: Gráficos de barras de los experimentos de MNIST. Estos gráficos recogen las métricas del mejor clasificador de cada tipo.

inducing points en el caso de VGPMIL. Puede verse que G-VGPMIL se muestra muy por encima de VGPMIL en este conjunto de datos.

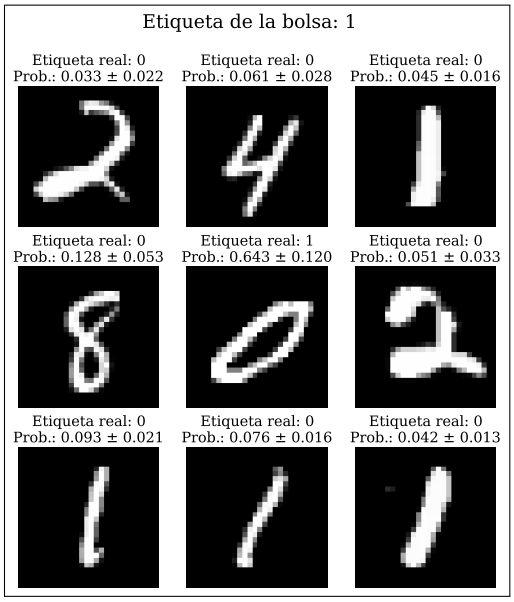
Num. Inducing	Lengthscale	k	Modelo	Inst. Acc.	Inst. F1	Inst. Log Loss	Inst. AUC	Bag Acc.	Bag F1	Bag Log Loss	Bag AUC
50	12.884	0.5	G-VGPMIL	**0.5968 ± 0.1239**	**0.2775 ± 0.14**	**0.6671 ± 0.0195**	**0.7975 ± 0.0499**	**0.6427 ± 0.0957**	**0.4449 ± 0.2241**	0.6643 ± 0.0101	**0.859 ± 0.0631**
	12.884	-	VGPMIL	0.5706 ± 0.1236	0.2118 ± 0.1242	0.7125 ± 0.1322	0.7148 ± 0.0953	0.6216 ± 0.1119	0.4162 ± 0.2313	**0.6568 ± 0.0656**	0.7985 ± 0.0493
100	12.884	0.5	G-VGPMIL	**0.5867 ± 0.1287**	**0.2643 ± 0.1293**	**0.6674 ± 0.0185**	**0.809 ± 0.0541**	**0.6316 ± 0.1077**	**0.4412 ± 0.2282**	0.665 ± 0.0089	**0.8615 ± 0.0644**
	12.884	-	VGPMIL	0.5683 ± 0.1238	0.2054 ± 0.1204	0.7086 ± 0.128	0.7191 ± 0.0941	0.6216 ± 0.1119	0.4162 ± 0.2313	**0.6522 ± 0.0629**	0.8081 ± 0.0457
200	12.884	0.5	G-VGPMIL	0.5917 ± 0.1236	0.2674 ± 0.1181	**0.6676 ± 0.018**	**0.8128 ± 0.0567**	0.6427 ± 0.0957	0.4449 ± 0.2241	**0.6651 ± 0.0079**	**0.8711 ± 0.0608**
	1.0	-	VGPMIL	**0.7432 ± 0.0786**	**0.6769 ± 0.081**	0.6931 ± 0.0	0.7889 ± 0.0369	**0.8053 ± 0.0606**	**0.7896 ± 0.0845**	0.6931 ± 0.0001	0.8567 ± 0.0292

Tabla 5.3.: Resumen de los resultados en musk1. Para cada valor del número de inducing points considerado, se recogen las métricas del mejor clasificador de cada tipo. Se señala en negrita el mejor valor de cada métrica dentro de cada grupo, y subrayado el mejor de cada columna.

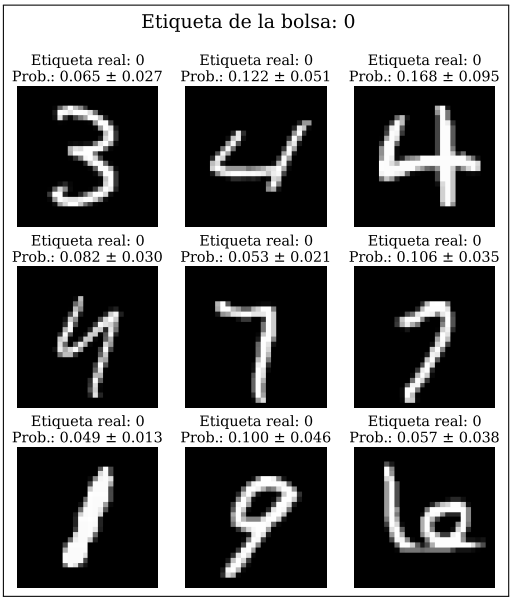
Si analizamos las tablas A.3 y A.4, donde se presentan todos los resultados, nos damos cuenta de que G-VGPMIL es muy superior a VGPMIL. Obsérvese que para cada valor del número de *inducing points* casi todas las métricas son mejores al usar alguno de los clasificadores G-VGPMIL. Al fijar este valor y el de *lengthscale* estamos comparando cómo se comportan los dos clasificadores partiendo de la misma información. De alguna forma, estamos evaluando cuánto conocimiento es capaz de extraer cada uno cuando se encuentran exactamente en las mismas condiciones. La Figura 5.4 representa la métrica AUC obtenida por ambos clasificadores en cada uno de los conjuntos. Obsérvese que la longitud de las barras amarillas (G-VGPMIL) es en todos los casos mayor o igual que las barras rojas (VGPMIL). Esto prueba la superioridad de G-VGPMIL en este conjunto de datos.

5.4. Detección de hemorragias intracraneales

En esta sección vamos a presentar una solución para el problema de detección de hemorragias intracraneales basada en los modelos que hemos desarrollado. En la introducción a este trabajo



(a) Predicciones en una bolsa con etiqueta positiva.



(b) Predicciones en una bolsa con etiqueta negativa.

Figura 5.3.: Predicciones de dos de las bolsas generadas a partir del conjunto MNIST. Todas las instancias se clasifican correctamente. Como consecuencia, la etiqueta que se predice de cada bolsa también es correcta.

ya se explicó la importancia de obtener un diagnóstico temprano de este tipo de lesiones. Además, también se justificó la idoneidad de los modelos probabilísticos para poder obtener predicciones interpretables y fiables, como las que se muestran en la Figura 5.3 para el caso de MNIST.

La detección de este tipo de lesiones se enmarca de forma natural dentro del paradigma de MIL. Cada escáner es una bolsa, y cada uno de los cortes que lo componen es una instancia. Si un paciente padece la lesión, su escáner se etiqueta como positivo. Pero, como se muestra en la Figura 5.5, no todos los cortes de un escáner positivo presentarán una etiqueta positiva, sino sólo unos pocos de ellos. Esto se debe a que el hematoma no se produce en todo el cráneo, sino sólo en una parte localizada de él. Obsérvese que en la subfigura 5.5a se aprecia una "mancha", correspondiente a la hemorragia.

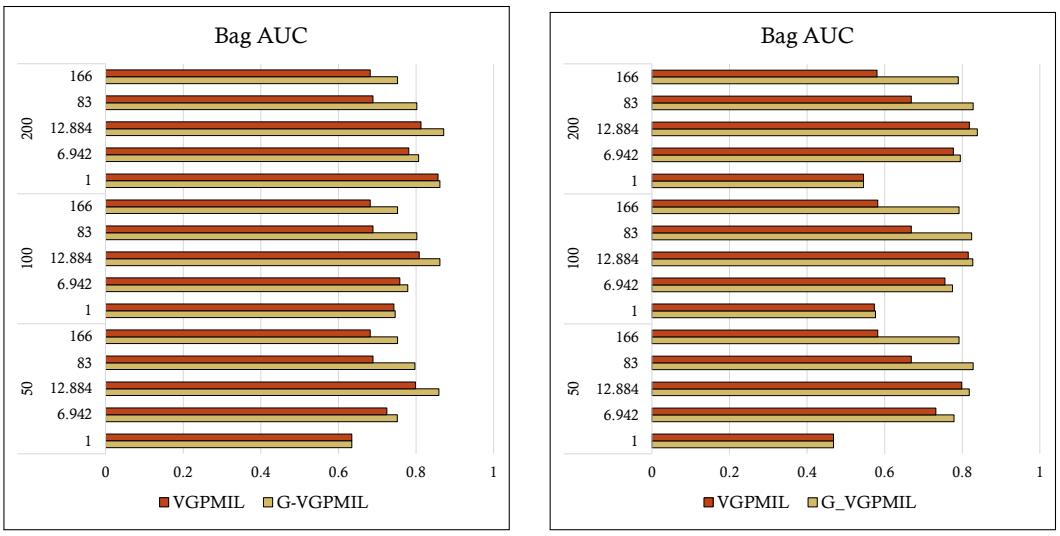
5.4.1. Conjuntos utilizados

Para poder entrenar y evaluar los modelos propuestos vamos a utilizar dos conjuntos de datos. El primero de ellos fue publicado por la Sociedad Norteamericana de Radiología (*Radiological Society of North America*, *RSNA*) [14] y se puede descargar desde Kaggle[5]. Está formado por 39750 cortes obtenidos de 1150 personas. De estas, se apartan 150 para test y 1000 para entrenamiento. El conjunto de entrenamiento contiene 589 escáneres normales (etiqueta negativa) y 411 escáneres con lesión (etiqueta positiva). El conjunto de test contiene 78 negativos y 72 positivos. El número de cortes en cada escáner varía entre desde 24 y 57, y cada uno tiene

[5] https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection

Num. Inducing	Lengthscale	k	Modelo	Inst. Acc.	Inst. F1	Inst. Log Loss	Inst. AUC	Bag Acc.	Bag F1	Bag Log Loss	Bag AUC
50	83.0	0.5	G-VGPMIL	0.8144 ± 0.0909	0.0183 ± 0.0366	0.6256 ± 0.0257	$\mathbf{0.7883 \pm 0.1846}$	0.6276 ± 0.0196	0.0444 ± 0.0889	0.6649 ± 0.0052	$\mathbf{0.8281 \pm 0.0871}$
	12.884	-	VGPMIL	$\mathbf{0.8262 \pm 0.1102}$	$\mathbf{0.2166 \pm 0.2576}$	$\mathbf{0.4003 \pm 0.1561}$	0.7425 ± 0.1759	$\mathbf{0.6662 \pm 0.0971}$	$\mathbf{0.338 \pm 0.2068}$	$\mathbf{0.5475 \pm 0.1046}$	0.7987 ± 0.107
100	12.884	2.0	G-VGPMIL	$\mathbf{0.8251 \pm 0.1105}$	$\underline{\mathbf{0.2973 \pm 0.2244}}$	0.4366 ± 0.1274	$\mathbf{0.7721 \pm 0.1503}$	$\mathbf{0.7167 \pm 0.0598}$	$\underline{\mathbf{0.5221 \pm 0.1362}}$	$\mathbf{0.5472 \pm 0.0397}$	$\mathbf{0.8271 \pm 0.0915}$
	12.884	-	VGPMIL	0.8224 ± 0.104	0.212 ± 0.232	$\mathbf{0.4057 \pm 0.1511}$	0.7374 ± 0.1713	0.6667 ± 0.0724	0.3453 ± 0.1301	0.5484 ± 0.1048	0.8157 ± 0.1014
200	12.884	1.0	G-VGPMIL	$\underline{\mathbf{0.8343 \pm 0.0894}}$	$\mathbf{0.2606 \pm 0.2131}$	0.5627 ± 0.0606	$\underline{\mathbf{0.7888 \pm 0.1704}}$	$\underline{\mathbf{0.7257 \pm 0.0577}}$	$\mathbf{0.4889 \pm 0.1438}$	0.6204 ± 0.0181	$\underline{\mathbf{0.8389 \pm 0.1011}}$
	12.884	-	VGPMIL	0.8226 ± 0.1101	0.2278 ± 0.2362	$\underline{\mathbf{0.3964 \pm 0.1476}}$	0.7497 ± 0.1668	0.6767 ± 0.0575	0.3756 ± 0.1034	$\underline{\mathbf{0.5421 \pm 0.1112}}$	0.8181 ± 0.1064

Tabla 5.4.: Resumen de los resultados en musk2. Para cada valor del número de inducing points considerado, se recogen las métricas del mejor clasificador de cada tipo. Se señala en negrita el mejor valor de cada métrica dentro de cada grupo, y subrayado el mejor de cada columna.



(a) musk1.

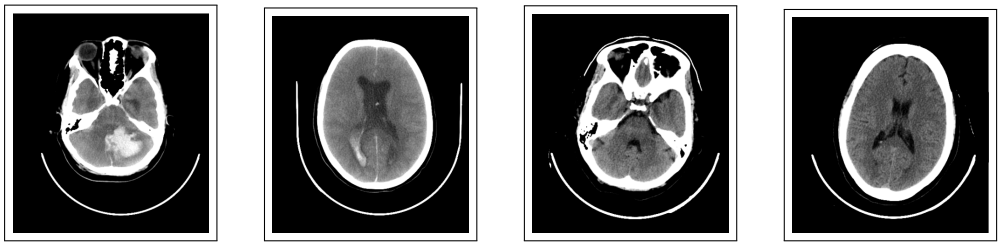
(b) musk2.

Figura 5.4.: Gráfico de barras de la métrica AUC en cada una de las configuraciones de los experimentos con los conjuntos musk1 y musk2. El eje vertical refleja el número de *inducing points* (50, 100 ó 200) seguido del parámetro *lengthscale* (1, 6.942, 12.884, 83 y 166). Obsérvese cómo la barra amarilla tiene en todos los casos una longitud mayor o igual a la roja.

dimensiones 512×512. En entrenamiento tenemos 29520 cortes negativos y 4976 cortes positivos. En test contamos con 4448 negativos y 806 positivos. Obsérvese que en este conjunto disponemos de las etiquetas tanto a nivel de instancia como a nivel de bolsa.

El otro conjunto que usaremos se denomina CQ500 y podemos acceder a él desde V7[6]. Está formado por escáneres de pacientes de varios centros médicos de India [8]. El número de cortes en cada escáner varía desde 16 hasta 128 cortes. En este conjunto sólo disponemos de etiquetas a nivel del escáner completo y lo usaremos para evaluación. En este tipo de trabajos es común evaluar los modelos desarrollados en este conjunto para poder compararlos con los de la comunidad.

[6] https://www.v7labs.com/open-datasets/cq500



(a) Corte con etiqueta positiva, procedente de un escáner con etiqueta positiva.

(b) Corte con etiqueta negativa, procedente de un escáner con etiqueta positiva.

(c) Corte con etiqueta negativa, procedente de un escáner con etiqueta negativa.

(d) Corte con etiqueta negativa, procedente de un escáner con etiqueta negativa.

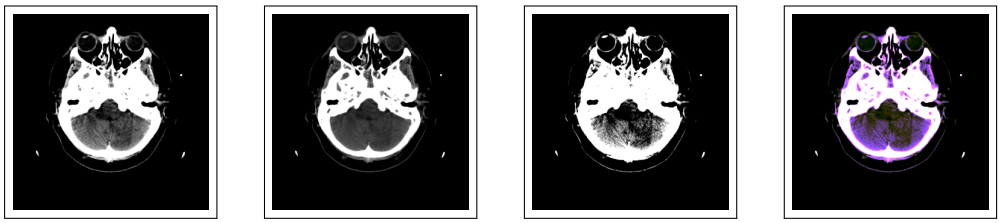
Figura 5.5.: Ejemplos de cortes del conjunto CQ500. Las subfiguras (a) y (b) muestran dos cortes de un escáner correspondiente a un paciente que presenta la lesión. Las subfiguras (c) y (d) muestran cortes de un escáner correspondiente a un peciente que no presenta la lesión.

5.4.2. Preprocesamiento

Para analizar los cortes de TAC, los radiólogos se fijan en diferentes intervalos de radiodensidad dentro de cada corte. Para imitar este comportamiento, a partir de cada imagen original I se extraen tres nuevas imágenes, I_1, I_2, I_3 de las mismas dimensiones. Cada una de estas imágenes I_n está definida mediante un par (c_n, w_n) donde c_n es el centro del intervalo y w_n es la anchura del intervalo,

$$I_n(i,j) = \begin{cases} c_n - w_n/2 & \text{si } I(i,j) < c_n - w_n/2, \\ I(i,j) & \text{si } c_n - w_n/2 \leq I(i,j) \leq c_n + w_n/2, \\ c_n + w_n/2 & \text{si } c_n + w_n/2 < I(i,j). \end{cases}$$

Los centros y anchuras que definen a cada una de las nuevas imágenes son $(40, 80)$, $(80, 200)$ y $(40, 380)$, todos ellos en unidades Hounsfield. Cada imagen captura un tipo de tejido diferente, de forma que la primera representa el tejido del cerebro, la segunda el subdural, y la tercera el hueso. Este proceso transforma cada matriz $I \in \mathbb{R}^{H \times W}$ (en escala de grises) en una matriz muldimensional o tensor $(I_1, I_2, I_3) \in \mathbb{R}^{H \times W \times 3}$. En la Figura 5.6 se muestra un ejemplo de este preprocesamiento.



(a) I_1, tejido cerebral.

(b) I_2, tejido subdural.

(c) I_3, tejido óseo.

(d) (I_1, I_2, I_3).

Figura 5.6.: Ejemplo de un corte de TAC preprocesado.

5.4.3. Metodología y arquitecturas

Para aumentar la capacidad predictiva de nuestros clasificadores, primero extraeremos características de cada uno de los cortes mediante una red convolucional. Con estas características, posteriormente se entrenarán los modelos G-VGPMIL y VGPMIL. De esta forma, siguiendo el modelo propuesto en [53], el entrenamiento se organiza en dos fases.

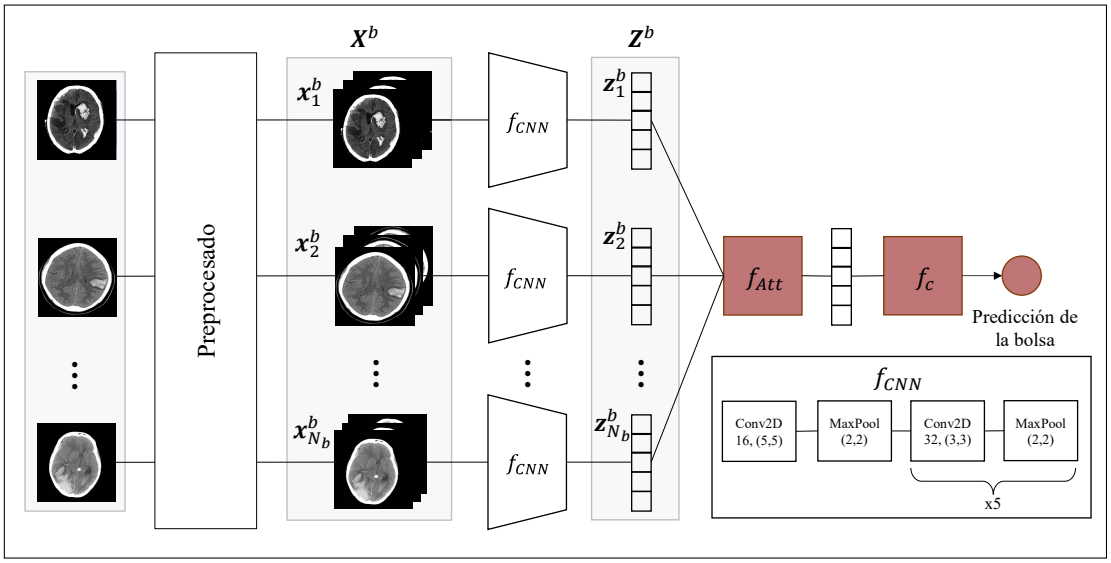


Figura 5.7.: Esquema de AttCNN. Imagen adaptada de [53].

Fase 1 (AttCNN). Usamos la arquitectura propuesta en [53], la cual está basada en capas convolucionales y una capa de atención. Un esquema de la misma puede observarse en la Figura 5.7. Este modelo recibe el nombre de AttCNN y responde a la composición de tres funciones, de forma que para cada bolsa $\boldsymbol{X}^b$ se aproxima $p(T_b = 1 \mid \boldsymbol{X}^b) \approx \text{AttCNN}\left(\boldsymbol{X}^b\right) = (f_c \circ f_{Att} \circ f_{CNN})\left(\boldsymbol{X}^b\right)$. A continuación describimos cada una de estas funciones.

- En primer lugar se aplica una red convolucional, que sirve como un primer extractor de características. Esta red se corresponde con la aplicación de $f_{CNN}\colon \mathbb{R}^{H\times W\times C} \to \mathbb{R}^D$ a cada instancia de la bolsa y la posterior concatenación de sus resultados. Esto es,

$$f_{CNN}(\boldsymbol{X}^b) = \left[f_{CNN}(\boldsymbol{x}_1^b), \ldots, f_{CNN}(\boldsymbol{x}_{N_b}^b)\right],$$

  donde N_b es el número de ejemplos en la bolsa b. Obsérvese que esta aplicación transforma cada bolsa $\boldsymbol{X}^b$ en otra $\boldsymbol{Z}^b = f_{CNN}(\boldsymbol{X}^b)$ que será la entrada a VGPMIL y G-VGPMIL. El valor de D determina el número de características que usarán estos clasificadores. Se mostrarán resultados con $D \in \{8, 32, 128\}$.

- A continuación se aplica una capa de atención. Esta capa se corresponde con la aplicación $f_{Att}\colon \mathbb{R}^{D\times N_b} \to \mathbb{R}^D$ y asigna un peso $\alpha_i \in \mathbb{R}$ a cada vector de entrada, de forma que

$$f_{Att}\left(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{N_b}\right) = \sum_{i=1}^{N_b} \alpha_i \boldsymbol{z}_i.$$

Los pesos se definen como

$$\alpha_i = \frac{\exp\left(\boldsymbol{w}^\top \tanh\left(V\boldsymbol{z}_i\right)\right)}{\sum_{j=1}^{N_b} \exp\left(\boldsymbol{w}^\top \tanh\left(V\boldsymbol{z}_j\right)\right)},$$

donde $\boldsymbol{w} \in \mathbb{R}^{50}$ y $V \in \mathbb{R}^{50\times D}$ son parámetros entrenables. Obsérvese que estos pesos siempre suman uno.

- Por último se aplica un clasificador $f_c\colon \mathbb{R}^D \to [0,1]$ cuya salida se puede interpretar como la probabilidad de que ese ejemplo pertenezca a la clase 1. Este clasificador está formado por una capa totalmente conectada con activación sigmoidal.

En la primera fase AttCNN se entrena para minimizar la entropía cruzada entre las verdaderas etiquetas de cada bolsa y las que ella misma predice. Tras este proceso f_{CNN} será capaz de extraer características apropiadas para el problema de MIL. Las características extraídas por AttCNN se encuentran disponibles en Github[7].

Fase 2 (AttCNN+(G-)VGPMIL). En la segunda fase AttCNN se modifica para poder usar los clasificadores basados en GPs. La nueva arquitectura se ilustra en la Figura 5.8. Como ya habíamos adelantado, cada bolsa $\boldsymbol{X}^b$ se transforma en $\boldsymbol{Z}^b = f_{CNN}(\boldsymbol{X}^b)$, y estas últimas se usan para aprender las distribuciones variacionales de los GPs. Se predice la etiqueta de cada instancia y, a partir de ellas, se obtiene una predicción de la etiqueta de la bolsa. Hay que recalcar que en esta segunda fase, f_{CNN} se encuentra "congelada", en el sentido de que sus parámetros no se modifican.

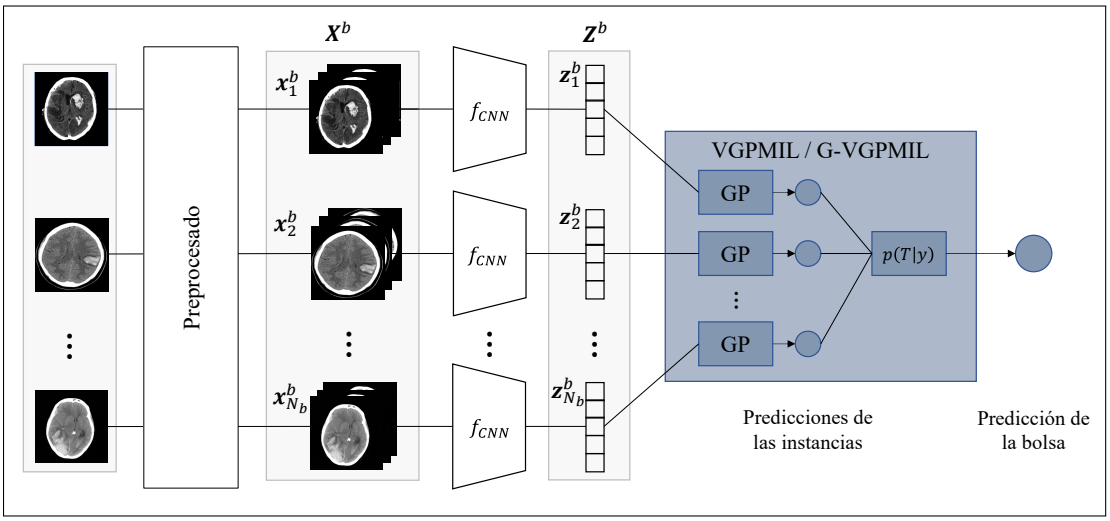


Figura 5.8.: Esquema de AttCNN+(G-)VGPMIL. Imagen adaptada de [53].

5.4.4. Resultados y discusión

En este problema contamos con tres clasificadores. El primero es AttCNN, y los otros dos son el resultado de combinar las características extraídas por AttCNN con VGPMIL y G-VGPMIL.

[7] https://github.com/arneschmidt/cnn_plus_vgpmil

Para cada valor de D se han repetido todos los experimentos, cuyos resultados se encuentran en las tablas A.5, A.6 y A.7.

5.4.4.1. AttCNN vs VGPMIL vs G-VGPMIL

Comencemos discutiendo los resultados sobre el conjunto de RSNA, que se resumen en la Tabla 5.5 y en la Figura 5.9 (subfiguras 5.9a y 5.9b). Lo primero que observamos es que VGPMIL y G-VGPMIL suponen un salto de calidad respecto a AttCNN. Obsérvese que G-VGPMIL ofrece resultados muy competitivos con los de VGPMIL. En las métricas Accuracy y F1 nuestro modelo se encuentra más de una décima por encima. Sólo a nivel de instancia AttCN y G-VGPMIL se encuentran más igualados. AttCNN usa los pesos de atención para aproximar la probabilidad a este nivel, lo cual produce predicciones de etiqueta acertadas al fijar el umbral en 0.5 (métricas Accuracy y F1). Sin embargo, estos pesos no funcionan mejor que G-VGPMIL a la hora de estimar las probabilidades (métricas Log Loss y AUC).

Num. características	Modelo	RSNA inst. Acc.	RSNA inst. F1	RSNA inst. Log Loss	RSNA inst. AUC	RSNA bag Acc.	RSNA bag F1	RSNA bag Log Loss	RSNA bag AUC
8	AttCNN	0.9228 ± 0.0042	0.7732 ± 0.0071	0.3347 ± 0.006	0.8956 ± 0.003	0.7813 ± 0.0208	0.8108 ± 0.0155	7.5526 ± 0.7194	0.7888 ± 0.0202
	G-VGPMIL	0.9211 ± 0.0024	0.6662 ± 0.0142	0.221 ± 0.0052	0.9613 ± 0.0008	0.884 ± 0.009	0.8654 ± 0.0115	0.2817 ± 0.0135	0.9596 ± 0.0072
	VGPMIL	$\underline{\mathbf{0.9397 \pm 0.0016}}$	$\underline{\mathbf{0.7778 \pm 0.0077}}$	$\underline{\mathbf{0.1689 \pm 0.0043}}$	$\underline{\mathbf{0.9649 \pm 0.002}}$	$\mathbf{0.8973 \pm 0.0161}$	$\mathbf{0.8914 \pm 0.0168}$	$\mathbf{0.2583 \pm 0.0204}$	$\mathbf{0.9604 \pm 0.007}$
32	AttCNN	0.9022 ± 0.0172	0.7359 ± 0.0293	0.3561 ± 0.001	0.8931 ± 0.0073	0.704 ± 0.065	0.7619 ± 0.0388	10.2237 ± 2.2448	0.7144 ± 0.0623
	G-VGPMIL	0.9243 ± 0.0088	0.7274 ± 0.0519	0.401 ± 0.0171	0.9414 ± 0.0066	0.868 ± 0.0314	0.8668 ± 0.022	0.4468 ± 0.0653	0.9568 ± 0.0029
	VGPMIL	$\mathbf{0.9316 \pm 0.0097}$	$\mathbf{0.7576 \pm 0.0342}$	$\mathbf{0.1895 \pm 0.0176}$	$\mathbf{0.9501 \pm 0.0058}$	$\mathbf{0.8893 \pm 0.0161}$	$\mathbf{0.8862 \pm 0.0158}$	$\mathbf{0.2791 \pm 0.0222}$	$\mathbf{0.9583 \pm 0.0048}$
128	AttCNN	0.9129 ± 0.0117	0.7557 ± 0.0223	0.3501 ± 0.003	0.897 ± 0.0038	0.7307 ± 0.0443	0.7777 ± 0.0262	9.3026 ± 1.5316	0.7401 ± 0.0423
	G-VGPMIL	0.928 ± 0.0036	0.7328 ± 0.012	0.4011 ± 0.0183	0.9415 ± 0.0051	0.8973 ± 0.0229	0.8902 ± 0.0216	0.4757 ± 0.0527	0.9657 ± 0.0083
	VGPMIL	$\mathbf{0.9344 \pm 0.0069}$	$\mathbf{0.7647 \pm 0.0243}$	$\mathbf{0.1797 \pm 0.0112}$	$\mathbf{0.9534 \pm 0.0049}$	$\underline{\mathbf{0.908 \pm 0.0186}}$	$\underline{\mathbf{0.9037 \pm 0.0196}}$	$\underline{\mathbf{0.246 \pm 0.0204}}$	$\underline{\mathbf{0.966 \pm 0.0062}}$

Tabla 5.5.: Resultados sobre el conjunto de RSNA. Para cada valor del número de características, D, se recogen las métricas obtenidas por los mejores modelos. Se señala en negrita el mejor valor de cada métrica para cada valor de D, y subrayado el mejor de cada columna.

Debemos recalcar la superioridad de VGPMIL en el conjunto de RSNA, que obtiene grandes resultados tanto a nivel de bolsa como a nivel de instancia. Para todos los valores de D este clasificador es, indiscutiblemente, el que mejor funciona. Llama la atención que el mejor modelo a nivel de bolsa no lo sea a nivel de instancia, y viceversa. Aunque el rendimiento a ambos niveles está claramente relacionado, esto muestra que no debemos concentrarnos en uno de ellos.

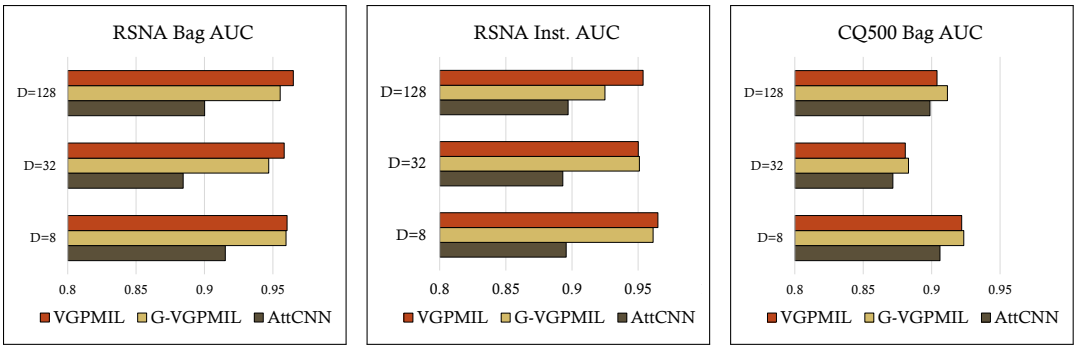
Vamos a centrarnos ahora en el conjunto CQ500, cuyos resultados se encuentran en la Tabla 5.6 y en la Figura 5.9c. Como era de esperar, todas la métricas son inferiores a las de RSNA. Esto ocurre porque los modelos han sido entrenados en un conjunto diferente, y al evaluarlos en CQ500 el rendimiento disminuye debido a la nueva naturaleza de las imágenes. Aún así, los resultados para VGPMIL y G-VGPMIL siguen siendo satisfactorios. AttCNN vuelve a quedarse por detrás de VGPMIL y G-VGPMIL, pero ahora este último ocupa el primer puesto con diferencia para todos los valores de D. En métricas como la tasa de acierto, G-VGPMIL se sitúa hasta dos centésimas por encima. Esto pone de manifiesto que G-VGPMIL presenta una capacidad de generalización mayor que la de VGPMIL.

5.4.4.2. Número de características

En la discusión anterior no hemos profundizado en el papel que juega el parámetro D. Recordemos que este número controla la dimensión del vector de características que se extrae de cada corte, por lo que esperamos que influya en la capacidad de decisión de cada clasificador.

Num. características	Modelo	CQ500 bag Acc.	CQ500 bag F1	CQ500 bag Log Loss	CQ500 bag AUC
8	AttCNN	0.6547 ± 0.0387	0.7004 ± 0.0204	0.6238 ± 0.0442	0.9061 ± 0.0088
	G-VGPMIL	$\underline{\mathbf{0.8645 \pm 0.013}}$	$\underline{\mathbf{0.8216 \pm 0.02}}$	$\underline{\mathbf{0.3457 \pm 0.0223}}$	$\underline{\mathbf{0.9235 \pm 0.0058}}$
	VGPMIL	0.8314 ± 0.0077	0.8096 ± 0.0093	0.3766 ± 0.011	0.922 ± 0.0061
32	AttCNN	0.6329 ± 0.0731	0.6924 ± 0.0536	0.6229 ± 0.0823	0.8717 ± 0.0255
	G-VGPMIL	0.8141 ± 0.0076	0.7409 ± 0.0265	0.4637 ± 0.0307	$\mathbf{0.8869 \pm 0.03}$
	VGPMIL	$\mathbf{0.8252 \pm 0.0327}$	$\mathbf{0.7893 \pm 0.0439}$	$\mathbf{0.4189 \pm 0.05}$	0.8812 ± 0.0311
128	AttCNN	0.6673 ± 0.0407	0.712 ± 0.0348	0.5956 ± 0.0736	0.8988 ± 0.0215
	G-VGPMIL	$\mathbf{0.8542 \pm 0.0171}$	$\mathbf{0.8199 \pm 0.0332}$	0.4148 ± 0.041	$\mathbf{0.9115 \pm 0.0197}$
	VGPMIL	0.8392 ± 0.0199	0.813 ± 0.0282	$\mathbf{0.3976 \pm 0.035}$	0.9096 ± 0.0207

Tabla 5.6.: Resultados sobre el conjunto de CQ500. Para cada valor del número de características, D, se recogen las métricas obtenidas por los mejores modelos. Se señala en negrita el mejor valor de cada métrica para cada valor de D, y subrayado el mejor de cada columna.



(a) RSNA, nivel de bolsa. (b) RSNA, nivel de instancia. (c) CQ500, nivel de bolsa.

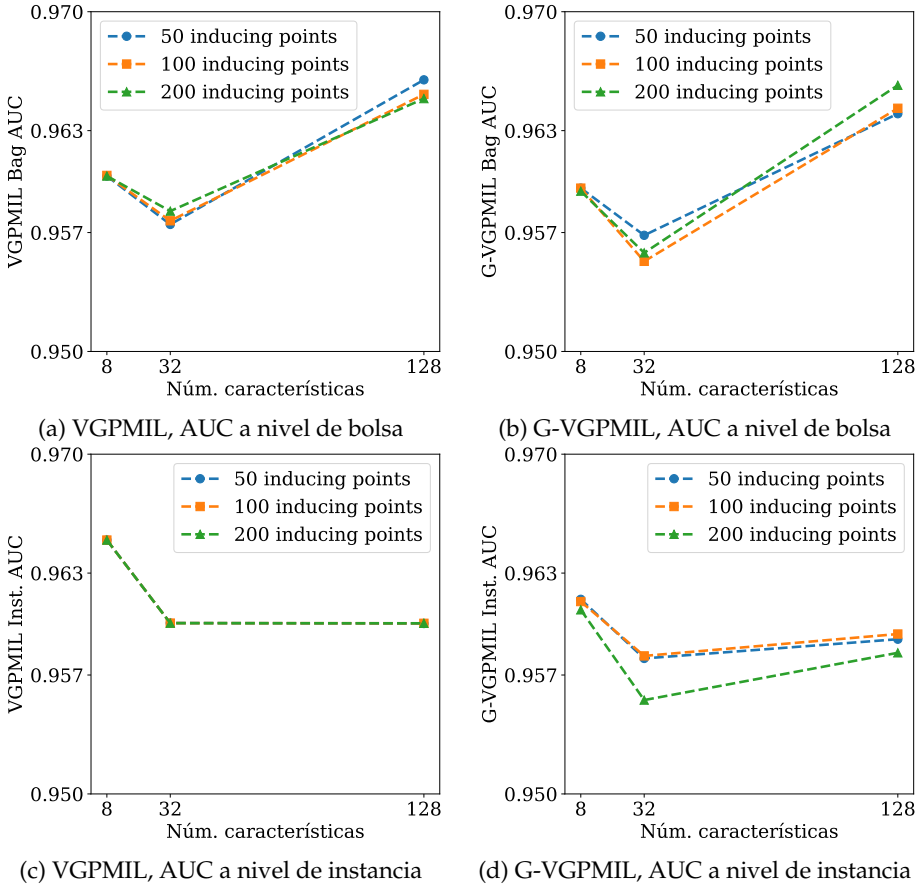
Figura 5.9.: Métrica AUC, tanto a nivel de bolsa como a nivel de instancia, en los conjuntos de RSNA y CQ500.

Para poder analizarlo hemos representado cómo varía la métrica AUC, tanto a nivel de instancia como de bolsa, en la Figura 5.10. El rendimiento empeora cuando fijamos a 32 el número de características, dibujándose una especie de valle en las gráficas. A nivel de bolsa es evidente que con 128 características se obtiene un mejor comportamiento, pero a nivel de instancia el mejor desempeño ocurre cuando este número se fija a 8. Sin embargo, hemos decidido evaluar estos valores de D para seguir el trabajo de [53].

5.4.4.3. El parámetro k de G-VGPMIL

Uno de los aspectos más interesantes de G-VGPMIL es el parámetro k. Con el objetivo de analizar cómo influye este parámetro en el rendimiento del modelo, vamos a estudiar cómo influye en la convergencia de las distribuciones variacionales y su relación con los resultados obtenidos.

Para ello hemos representado en la Figura 5.12 cómo evoluciona la métrica Log Loss con el número de épocas. Una época es una actualización completa de todos los parámetros, y en ella se usa todo el conjunto de entrenamiento. Analizamos los resultados correspondientes a tomar $D = 128$, lengthscale $= \sqrt{D}$ y 200 inducing points. Usando los datos del Apéndice A, el lector



(a) VGPMIL, AUC a nivel de bolsa

(b) G-VGPMIL, AUC a nivel de bolsa

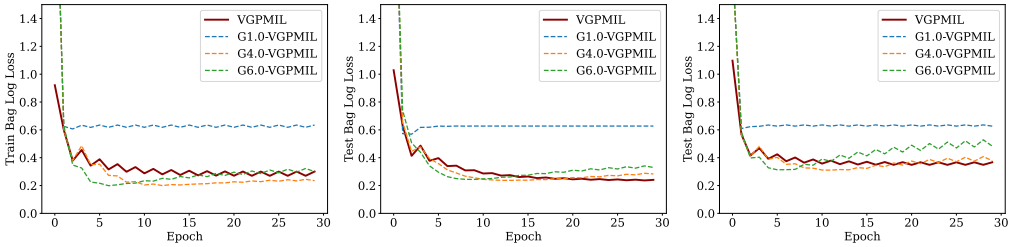
(c) VGPMIL, AUC a nivel de instancia

(d) G-VGPMIL, AUC a nivel de instancia

Figura 5.10.: Influencia del número de características en el rendimiento. Se muestra la evolución de la métrica AUC conforme aumenta el valor de D. Resultados obtenidos a partir del conjunto RSNA.

puede comprobar que para otras configuraciones de estos hiperparámetros el comportamiento es similar. Cuando $k = 1$ la función de pérdida se estanca en un valor alto, pero conforme aumenta k el comportamiento mejora. En la subfigura 5.12a puede comprobarse que cuando $k = 6$ alrededor de la sexta época ya se ha obtenido el mínimo, que se encuentra por debajo del que conseguirá VGPMIL. Sin embargo, las gráficas de test, subfiguras 5.12b y 5.12c, muestran que ocurre cierto sobreajuste y que este es más acusado cuanto mayor es k. Esto convierte a G-VGPMIL en un modelo cuyo proceso de entrenamiento es mucho más eficiente que VGPMIL si introducimos mecanismos para evitar el sobreaprendizaje. Un simple *early stopping* bastaría para solucionar este problema.

Para comprobar la relación de este hecho con la métrica AUC, hemos fijado el número de características en 128 y en la Figura 5.12 hemos recogido las gráficas de las métricas AUC y Log Loss frente al valor de k con el que se han obtenido. Los resultados en ambos conjuntos muestran que cuanto mayor es el parámetro k mejor rendimiento se obtiene. Las subfiguras 5.12a, 5.12b, 5.12d y 5.12e muestran que en el conjunto de RSNA la métrica AUC empeora a nivel de bolsa cuando $k = 4$. Sin embargo, Log Loss mejora considerablemente a ambos niveles. Las subfiguras 5.12c y 5.12f apoyan esta idea ya que la gráfica de Log Loss es estrictamente



(a) Conjunto de entrenamiento de RSNA. (b) Conjunto de test de RSNA. (c) Conjunto CQ500.

Figura 5.11.: Evolución de la métrica Log Loss a lo largo del proceso de entrenamiento. Resultados obtenidos tomando $D = 128$ y 200 inducing points.

decreciente, mientras que la de AUC es estrictamente creciente.

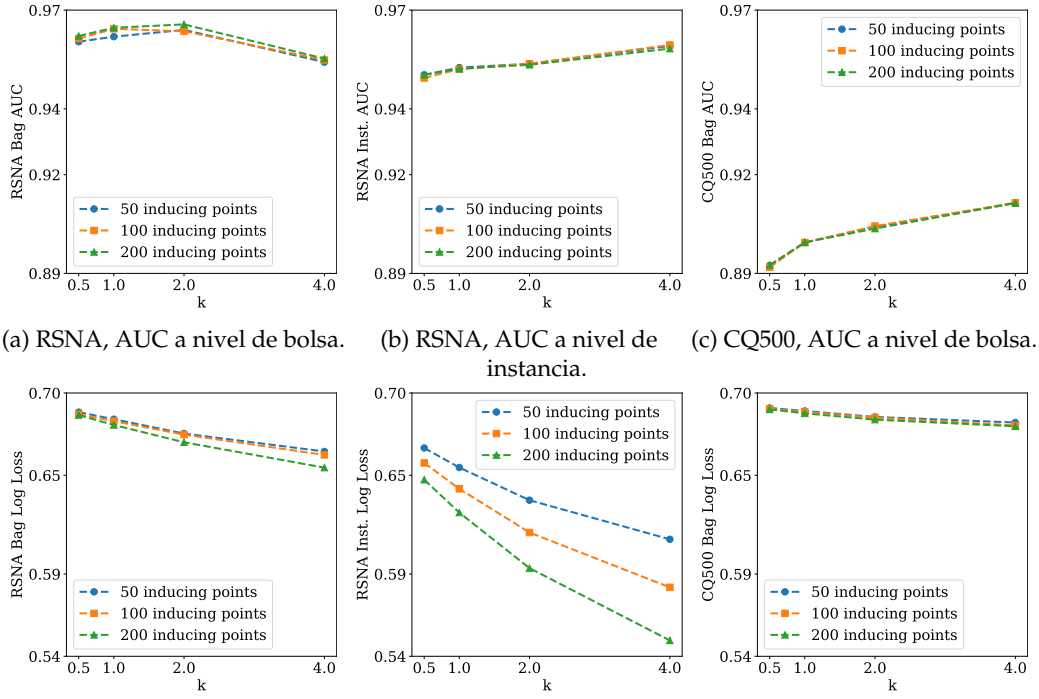
Recordemos que modificando k podíamos alterar la forma de la función θ (ecuaciones (4.12) y (4.13)) a partir de la cual se construye la matriz diagonal $\mathbf{\Theta}$. Cuando $k = 4$, las funciones θ y 2λ presentaban el mismo máximo absoluto. El comportamiento que observamos ya se había dejado ver en las tablas de MNIST y musk, y justifica la sustituición de las variables Pólya-Gamma. Estos resultados convierten a G-VGPMIL en un potente modelo, que destaca por su flexibilidad frente a VGPMIL.

5.4.4.4. Comparación con el estado del arte

La Tabla 5.7 compara los modelos estudiados en este trabajo con otras aproximaciones de la literatura. En todas ellas se entrenó un modelo sobre un conjunto de datos (en algunas también RSNA) para después evaluarlo sobre CQ500. Observamos que todas las propuestas emplean redes convolucionales, pero sólo una se limita a usar etiquetas a nivel de bolsa. En [**?**] emplean etiquetas a nivel de instancia sin imponer ninguna relación entre los objetos de una bolsa. En [29], sin embargo, usan el mecanismo de las capas LSTM para introducir relaciones entre las instancias contiguas. Ambos planteamiento obtienen una métrica AUC superior a nuestros clasificadores, pero no debemos olvidar que requieren un esfuerzo de anotación mucho mayor. Las propuestas basadas en MIL obtienen métricas muy competitivas (obsérvese que AttCNN + G-VGPMIL se encuentra a menos de 4 centesimas de [29]), usando sólo etiquetas a nivel de bolsa. Las aproximaciones presentadas en este trabajo, y en particular G-VGPMIL, son muy superiores al resto de trabajos que usan etiquetas a nivel de bolsa, como es [28].

Modelo	Tipo de etiquetas	CQ500 bag AUC
2D CNNs [8]	Instancia	0.94
2D CNN + LSTM [29]	Instancia	0.96
Segmentación con CNN [28]	Bolsa	0.83
AttCNN	Bolsa	0.9061
AttCNN + VGPMIL	Bolsa	0.922
AttCNN + G-VGPMIL	Bolsa	0.9235

Tabla 5.7.: Comparación con otras aproximaciones de la literatura.



(a) RSNA, AUC a nivel de bolsa.

(b) RSNA, AUC a nivel de instancia.

(c) CQ500, AUC a nivel de bolsa.

(d) RSNA, Log Loss a nivel de bolsa.

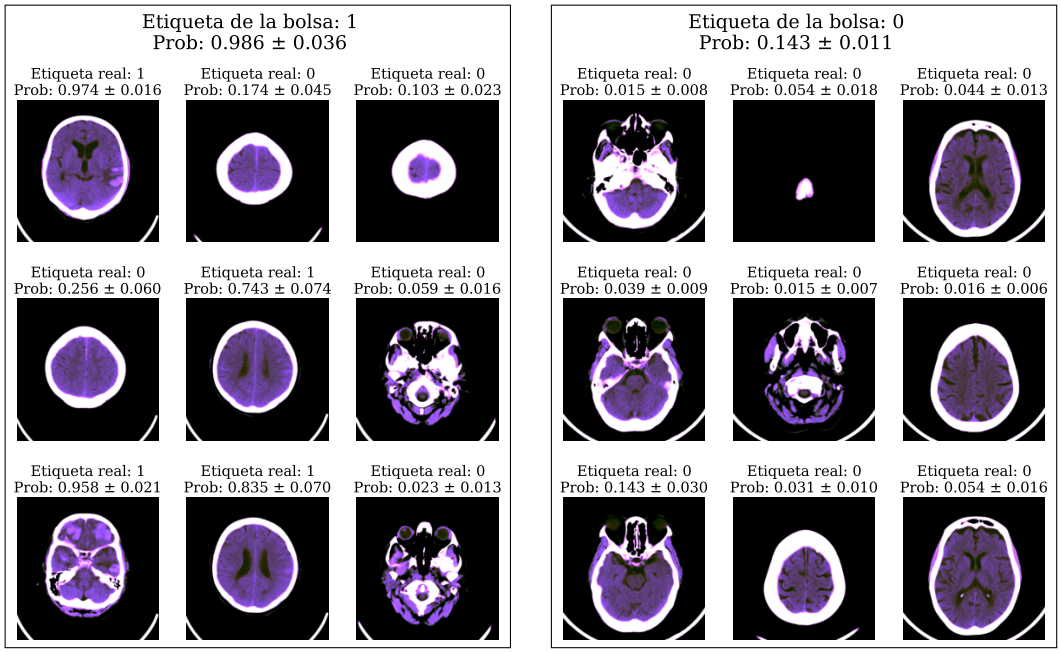
(e) RSNA, Log Loss a nivel de instancia.

(f) CQ500, Log Loss a nivel de bolsa.

Figura 5.12.: Métricas AUC y Log Loss, tanto a nivel de bolsa como a nivel de instancia, en los conjuntos de RSNA y CQ500, frente al parámetro k. Resultados obtenidos tomando $D = 128$.

Para terminar, ofrecemos en la Figura 5.13 un ejemplo de cómo funciona el sistema CAD desarrollado. Dado un escáner completo de un paciente, el modelo AttCNN + G-VGPMIL es capaz de estimar si el paciente sufre o no una hemorragia intracraneal. Esta estimación se da en forma de probabilidad, tanto a nivel de bolsa completa como a nivel de instancia. Además para cada probabilidad se incluye una cantidad que indica el intervalo de confianza de la predicción (obsérvese la ecuación (4.9)). Juntos cuantifican la incertidumbre que se tiene en la estimación ofrecida.

Ante la más que posible pregunta del radiólgo *¿por qué esa bolsa debe tener etiqueta positiva?* nuestro modelo sería capaz de señalar qué instancias le han llevado a tomar esa decisión (es decir, aquellas cuya etiqueta es positiva) y con qué probabilidad. Además, sería capaz de indicar qué seguridad se tiene en tal estimación, lo cual es esencial para advertir al experto sobre qué instancias o ejemplos debe revisar. Ninguno de los modelos basados exclusivamente en arquitecturas profundas, como los de la Tabla 5.7, pueden satisfacer estos requerimientos como lo hacen los modelos probabilísticos.



(a) Predicciones de un escáner con etiqueta positiva.

(b) Predicciones de un escáner con etiqueta negativa.

Figura 5.13.: Predicciones de dos escáneres completos del conjunto de RSNA.

6. Conclusiones y trabajo futuro

El objetivo de este capítulo es resumir las lecciones que hemos aprendido y el trabajo que estas puede motivar en el futuro. Para ello primero ofrecemos un resumen de los problemas que hemos tratado y las conclusiones a las que hemos llegado tras resolverlos. Posteriormente, señalamos dónde podemos centrar nuestros esfuerzos en el futuro.

6.1. Conclusiones

Comenzábamos este trabajo hablando sobre la importancia de una detección temprana y fiable de las hemorragias intracraneales. Justificábamos por qué era necesario elaborar métodos de triaje basados en Sistemas de Diagnóstico Asistido por Ordenador en los que los radiólogos pudieran apoyarse. La construcción de estos sistemas mediante las técnicas habituales de Aprendizaje Automático requiere disponer de bases de datos fuertemente anotadas. Esto es, los radiólogos deben etiquetar una gran cantidad de cortes de TAC, lo cual es una tarea inabordable debido al gran esfuerzo que requiere y al poco tiempo del que disponen.

Lo anterior motivaba el uso de técnicas de Aprendizaje a partir de Múltiples Instancias (*Multiple Instance Learning*, MIL), las cuales sólo requieren una etiqueta para cada escáner completo. Esta información es fácil de conseguir ya que se encuentra disponible en el historial clínico del paciente. Así es cómo surgía el primer objetivo de este trabajo, que era comprender los problemas de tipo MIL y revisar la literatura comprendiendo qué soluciones se han propuesto hasta el momento.

La incertidumbre presente en MIL nos sugería el uso de métodos probabilísticos capaces de cuantificarla. Esto justificó la elección de los Procesos Gaussianos (*Gaussian Processes*, GPs) como herramienta, cuya formulación probabilística permite adaptarlos para modelar la relación entre las instancias y las bolsas en los problemas de MIL. En este trabajo hemos estudiado los fundamentos de estos modelos bayesianos en problemas de aprendizaje supervisado clásicos, para después considerar las aproximaciones basadas en GPs para MIL existentes en la literatura. De todas ellas, nos hemos centrado en un modelo denominado *Variational Gaussian Process Multiple Instance Learning* (VGPMIL), cuyo proceso de inferencia se basa en una desigualdad que puede conducir a soluciones subóptimas.

En este proyecto nos proponíamos explorar técnicas de *data augmentation* para realizar este proceso de inferencia. Concretamente, inspirados por una propuesta previa que usa variables aleatorias Pólya-Gamma, nos preguntábamos qué ocurría al incluir este tipo de variables en el modelo de GPs para MIL. Así es cómo hemos llegado a las dos aportaciones teóricas de este trabajo:

- Hemos construido *Pólya-Gamma Variational Gaussian Process Multiple Instance Learning* (PG-VGPMIL) que propone usar variables Pólya-Gamma para aumentar la probabilidad conjunta y así expresar la función logística a través de ellas. Probamos que PG-VGPMIL es equivalente, en el sentido *Mean Field*, a VGPMIL. Esto significa que existe cierta relación entre la desigualdad de Jaakkola y este tipo de variables aleatorias.

- La definición de las variables Pólya-Gamma nos sugiere sustituirlas por variables Gamma. Así hemos desarrollado *Gamma Variational Gaussian Process Multiple Instance Learning* (G-VGPMIL), que nos conduce a un nuevo modelo de observación. Hemos validado experimentalmente G-VGPMIL usando tres conjuntos diferentes. Los experimentos nos han llevado a concluir que es un modelo muy competitivo que en varias ocasiones se comporta mejor que VGPMIL.

La última parte del trabajo ofrece una solución al problema que lo motiva, la detección de hemorragias intracraneales. Esta solución, que se denomina AttCNN+G-VGPMIL se divide en dos fases. La primera se basa en el uso de redes neuronales y mecanismos de atención para extraer características. La segunda fase emplea G-VGPMIL sobre estas características para obtener predicciones a nivel de bolsa y de instancia. El modelo obtenido sólo necesita etiquetas a nivel de bolsa y ofrece los mejores resultados de todas las metodologías de la literatura de ese tipo, a la vez que se sitúa unas centésimas por debajo de modelos mucho más complejos que emplean etiquetas a nivel de instancia. Esto demuestra que usar metodologías de etiquetado débil es la vía a seguir para construir modelos que no dependan de conjuntos fuertemente anotados. Por otra parte, AttCNN+G-VGPMIL es capaz de informar sobre la incertidumbre de sus decisiones gracias al uso de GPs. Esto es una cualidad esencial dentro de la medicina.

6.2. Trabajo futuro

Los resultados de este Trabajo Fin de Máster se incluirán en un artículo que será enviado a una revista en el Journal Citation Report. Además, cada una de las aportaciones a la literatura que hemos realizado deja abierta una puerta a la investigación. Nos detenemos en cada una de ellas.

1. La equivalencia entre PG-VGPMIL y VGPMIL nos lleva a pensar en la relación existente entre la desigualdad de Jaakkola y las variables Pólya-Gamma. Pensamos que un hecho relacionado ya se apunta en las conclusiones de [51]. Aclarar esta cuestión puede suponer un gran avance a la hora de tratar el modelo de observación logístico.

2. ¿Qué ocurriría al cambiar las variables Gamma por otro tipo de variables aleatorias? La forma general de las distribuciones Pólya-Gamma y Gamma se definen como el producto por un factor exponencial de una versión simple de ellas mismas. Este hecho es el que nos permite calcular la distribución variacional de las nuevas variables. Explorar nuevas variables con esta propiedad se traduce en cambiar la función que define a la matriz diagonal $\mathbf{\Theta}$.

3. El modelo AttCNN+G-VGPMIL obtiene grandes resultados, pero tiene un gran inconveniente: el proceso de entrenamiento no es *end-to-end*, sino que debe realizarse en dos fases. Por tanto, las características extraídas en la primera fase pueden no ser las mejores para G-VGPMIL. Una forma de hacer esto podría ser entrenar G-VGPMIL mediante gradiente descendente y usar alguna herramienta basada en diferenciación automática para guiar el proceso de entrenamiento por toda la red.

A. Tablas

Num. Inducing	Lengthscale	k	Model	Inst. Acc.	Inst. F1	Inst. Log Loss	Inst. AUC	Bag Acc.	Bag F1	Bag Log Loss	Bag AUC
		0.5	G-VGPMIL	**0.9383 ± 0.0089**	**0.5541 ± 0.0089**	0.6874 ± 0.0089	**0.9038 ± 0.0089**	**0.6855 ± 0.0089**	**0.6546 ± 0.0089**	0.6931 ± 0.0089	**0.8025 ± 0.0089**
		1.0	G-VGPMIL	0.9382 ± 0.0089	0.5531 ± 0.0089	0.6851 ± 0.0089	0.9036 ± 0.0089	0.685 ± 0.0089	0.6538 ± 0.0089	0.6931 ± 0.0089	0.802 ± 0.0089
	1.0	2.0	G-VGPMIL	0.938 ± 0.0088	0.5507 ± 0.0088	0.6815 ± 0.0088	0.9032 ± 0.0088	0.6837 ± 0.0088	0.6519 ± 0.0088	0.6931 ± 0.0088	0.8009 ± 0.0088
		4.0	G-VGPMIL	0.9378 ± 0.0089	0.5481 ± 0.0089	**0.6764 ± 0.0089**	0.9026 ± 0.0089	0.6821 ± 0.0089	0.6492 ± 0.0089	**0.6931 ± 0.0089**	0.7992 ± 0.0089
		-	VGPMIL	0.9372 ± 0.0087	0.5422 ± 0.0087	0.6765 ± 0.0087	0.9019 ± 0.0087	0.677 ± 0.0087	0.6423 ± 0.0087	0.6931 ± 0.0087	0.7958 ± 0.0087
		0.5	G-VGPMIL	0.9759 ± 0.0039	0.8623 ± 0.0039	0.6196 ± 0.0039	0.988 ± 0.0039	0.8919 ± 0.0039	0.9032 ± 0.0039	0.6476 ± 0.0039	0.9739 ± 0.0039
		1.0	G-VGPMIL	0.9824 ± 0.0011	0.906 ± 0.0011	0.5724 ± 0.0011	0.9906 ± 0.0011	0.9262 ± 0.0011	0.937 ± 0.0011	0.5994 ± 0.0011	0.9785 ± 0.0011
	3.238	2.0	G-VGPMIL	0.9827 ± 0.002	0.9129 ± 0.002	0.4482 ± 0.002	0.9915 ± 0.002	0.9347 ± 0.002	0.9465 ± 0.002	0.447 ± 0.002	0.9807 ± 0.002
		4.0	G-VGPMIL	0.9842 ± 0.0015	0.9144 ± 0.0015	**0.0657 ± 0.0015**	0.9921 ± 0.0015	0.9331 ± 0.0015	0.9426 ± 0.0015	**0.1991 ± 0.0015**	0.9807 ± 0.0015
		-	VGPMIL	**0.9862 ± 0.0008**	**0.9287 ± 0.0008**	0.1393 ± 0.0008	**0.9933 ± 0.0008**	**0.9466 ± 0.0008**	**0.9557 ± 0.0008**	0.2309 ± 0.0008	**0.9839 ± 0.0008**
		0.5	G-VGPMIL	0.9638 ± 0.0055	0.7835 ± 0.0055	0.6392 ± 0.0055	0.9854 ± 0.0055	0.8388 ± 0.0055	0.8491 ± 0.0055	0.6678 ± 0.0055	0.9667 ± 0.0055
		1.0	G-VGPMIL	0.9764 ± 0.0059	0.8683 ± 0.0059	0.5911 ± 0.0059	0.9873 ± 0.0059	0.9006 ± 0.0059	0.9125 ± 0.0059	0.6313 ± 0.0059	0.9715 ± 0.0059
50	5.477	2.0	G-VGPMIL	0.981 ± 0.004	0.9013 ± 0.004	0.4646 ± 0.004	0.9886 ± 0.004	0.927 ± 0.004	0.9391 ± 0.004	0.4977 ± 0.004	0.9756 ± 0.004
		4.0	G-VGPMIL	0.9824 ± 0.0016	0.905 ± 0.0016	**0.0643 ± 0.0016**	0.9908 ± 0.0016	0.9285 ± 0.0016	0.9387 ± 0.0016	**0.2047 ± 0.0016**	0.9795 ± 0.0016
		-	VGPMIL	**0.9855 ± 0.0013**	**0.9257 ± 0.0013**	0.1414 ± 0.0013	**0.993 ± 0.0013**	**0.9445 ± 0.0013**	**0.954 ± 0.0013**	0.2353 ± 0.0013	**0.9846 ± 0.0013**
		0.5	G-VGPMIL	0.9449 ± 0.0015	0.6124 ± 0.0015	0.6484 ± 0.0015	**0.9923 ± 0.0015**	0.72 ± 0.0015	0.7017 ± 0.0015	0.6788 ± 0.0015	**0.9835 ± 0.0015**
		1.0	G-VGPMIL	0.9684 ± 0.0013	0.8098 ± 0.0013	0.6068 ± 0.0013	0.9914 ± 0.0013	0.8538 ± 0.0013	0.864 ± 0.0013	0.6543 ± 0.0013	0.9816 ± 0.0013
	17.738	2.0	G-VGPMIL	**0.9815 ± 0.0021**	**0.903 ± 0.0021**	0.522 ± 0.0021	0.9891 ± 0.0021	**0.9277 ± 0.0021**	0.9392 ± 0.0021	0.5685 ± 0.0021	0.9768 ± 0.0021
		4.0	G-VGPMIL	0.9811 ± 0.0019	0.9006 ± 0.0019	**0.0781 ± 0.0019**	0.9893 ± 0.0019	0.9259 ± 0.0019	0.9374 ± 0.0019	**0.2177 ± 0.0019**	0.9766 ± 0.0019
		-	VGPMIL	0.9803 ± 0.001	0.8994 ± 0.001	0.2043 ± 0.001	0.9884 ± 0.001	0.9274 ± 0.001	**0.9398 ± 0.001**	0.3175 ± 0.001	0.9756 ± 0.001
		0.5	G-VGPMIL	0.9392 ± 0.0017	0.5546 ± 0.0017	0.649 ± 0.0017	**0.9905 ± 0.0017**	0.6835 ± 0.0017	0.6493 ± 0.0017	0.6805 ± 0.0017	**0.9795 ± 0.0017**
		1.0	G-VGPMIL	0.9649 ± 0.0024	0.7837 ± 0.0024	0.6075 ± 0.0024	0.9892 ± 0.0024	0.8357 ± 0.0024	0.8445 ± 0.0024	0.6575 ± 0.0024	0.9765 ± 0.0024
	30.0	2.0	G-VGPMIL	0.9791 ± 0.002	0.8898 ± 0.002	0.5194 ± 0.002	0.9869 ± 0.002	0.9185 ± 0.002	0.9309 ± 0.002	0.5743 ± 0.002	0.9713 ± 0.002
		4.0	G-VGPMIL	0.979 ± 0.0029	0.8872 ± 0.0029	**0.1129 ± 0.0029**	0.9891 ± 0.0029	0.9171 ± 0.0029	0.9287 ± 0.0029	**0.2697 ± 0.0029**	0.976 ± 0.0029
		-	VGPMIL	**0.9798 ± 0.0006**	**0.895 ± 0.0006**	0.2367 ± 0.0006	0.9876 ± 0.0006	**0.9236 ± 0.0006**	**0.936 ± 0.0006**	0.3577 ± 0.0006	0.9734 ± 0.0006
		0.5	G-VGPMIL	0.9592 ± 0.0045	0.7378 ± 0.0045	0.6869 ± 0.0045	0.9408 ± 0.0045	0.8057 ± 0.0045	0.8094 ± 0.0045	0.6931 ± 0.0045	**0.8802 ± 0.0045**
		1.0	G-VGPMIL	**0.9592 ± 0.0045**	**0.738 ± 0.0045**	0.6845 ± 0.0045	0.9408 ± 0.0045	**0.8059 ± 0.0045**	**0.8095 ± 0.0045**	0.6931 ± 0.0045	0.8801 ± 0.0045
	1.0	2.0	G-VGPMIL	0.9591 ± 0.0046	0.7375 ± 0.0046	0.6806 ± 0.0046	**0.9409 ± 0.0046**	0.8059 ± 0.0046	0.8095 ± 0.0046	0.6931 ± 0.0046	0.8798 ± 0.0046
		4.0	G-VGPMIL	0.959 ± 0.0046	0.7368 ± 0.0046	**0.6751 ± 0.0046**	0.9407 ± 0.0046	0.8051 ± 0.0046	0.8088 ± 0.0046	**0.6931 ± 0.0046**	0.8787 ± 0.0046
		-	VGPMIL	0.959 ± 0.0046	0.7368 ± 0.0046	0.6754 ± 0.0046	0.9407 ± 0.0046	0.8051 ± 0.0046	0.8088 ± 0.0046	0.6931 ± 0.0046	0.8788 ± 0.0046
		0.5	G-VGPMIL	0.9783 ± 0.0004	0.8771 ± 0.0004	0.6215 ± 0.0004	0.9927 ± 0.0004	0.9049 ± 0.0004	0.9155 ± 0.0004	0.6494 ± 0.0004	0.9859 ± 0.0004
		1.0	G-VGPMIL	0.987 ± 0.0009	0.9308 ± 0.0009	0.5739 ± 0.0009	0.9948 ± 0.0009	0.9456 ± 0.0009	0.9537 ± 0.0009	0.5955 ± 0.0009	0.989 ± 0.0009
	3.238	2.0	G-VGPMIL	0.9881 ± 0.0011	0.9397 ± 0.0011	0.4391 ± 0.0011	0.9954 ± 0.0011	0.9581 ± 0.0011	0.9656 ± 0.0011	0.4193 ± 0.0011	0.9895 ± 0.0011
		4.0	G-VGPMIL	0.9875 ± 0.0009	0.9333 ± 0.0009	**0.0509 ± 0.0009**	0.9961 ± 0.0009	0.9469 ± 0.0009	0.9547 ± 0.0009	**0.1598 ± 0.0009**	0.9902 ± 0.0009
		-	VGPMIL	**0.9909 ± 0.0009**	**0.9527 ± 0.0009**	0.1102 ± 0.0009	**0.9968 ± 0.0009**	**0.9645 ± 0.0009**	**0.9705 ± 0.0009**	0.1833 ± 0.0009	**0.9921 ± 0.0009**
		0.5	G-VGPMIL	0.9695 ± 0.0042	0.8181 ± 0.0042	0.638 ± 0.0042	0.9939 ± 0.0042	0.8592 ± 0.0042	0.8695 ± 0.0042	0.6641 ± 0.0042	0.9875 ± 0.0042
		1.0	G-VGPMIL	0.9844 ± 0.0014	0.9156 ± 0.0014	0.5888 ± 0.0014	0.9947 ± 0.0014	0.9351 ± 0.0014	0.9442 ± 0.0014	0.6146 ± 0.0014	0.9887 ± 0.0014
100	5.477	2.0	G-VGPMIL	0.9882 ± 0.0009	0.9398 ± 0.0009	0.4544 ± 0.0009	0.9951 ± 0.0009	0.9558 ± 0.0009	0.9635 ± 0.0009	0.4555 ± 0.0009	0.9898 ± 0.0009
		4.0	G-VGPMIL	0.9867 ± 0.0004	0.9282 ± 0.0004	**0.0476 ± 0.0004**	0.9958 ± 0.0004	0.9438 ± 0.0004	0.952 ± 0.0004	**0.1607 ± 0.0004**	0.9901 ± 0.0004
		-	VGPMIL	**0.9901 ± 0.0008**	**0.9493 ± 0.0008**	0.115 ± 0.0008	**0.9965 ± 0.0008**	**0.9635 ± 0.0008**	**0.9697 ± 0.0008**	0.189 ± 0.0008	**0.9917 ± 0.0008**
		0.5	G-VGPMIL	0.9535 ± 0.0026	0.6916 ± 0.0026	0.6479 ± 0.0026	**0.9954 ± 0.0026**	0.7704 ± 0.0026	0.7677 ± 0.0026	0.6763 ± 0.0026	**0.99 ± 0.0026**
		1.0	G-VGPMIL	0.9735 ± 0.0016	0.8456 ± 0.0016	0.6058 ± 0.0016	0.9945 ± 0.0016	0.8784 ± 0.0016	0.8893 ± 0.0016	0.6498 ± 0.0016	0.9881 ± 0.0016
	17.738	2.0	G-VGPMIL	**0.9834 ± 0.0013**	**0.9136 ± 0.0013**	0.5124 ± 0.0013	0.9913 ± 0.0013	**0.9346 ± 0.0013**	**0.9452 ± 0.0013**	0.5498 ± 0.0013	0.9811 ± 0.0013
		4.0	G-VGPMIL	0.9803 ± 0.0004	0.8935 ± 0.0004	**0.0719 ± 0.0004**	0.9886 ± 0.0004	0.9191 ± 0.0004	0.9304 ± 0.0004	**0.2294 ± 0.0004**	0.9748 ± 0.0004
		-	VGPMIL	0.9817 ± 0.0007	0.9059 ± 0.0007	0.1955 ± 0.0007	0.9895 ± 0.0007	0.9317 ± 0.0007	0.9433 ± 0.0007	0.3073 ± 0.0007	0.9779 ± 0.0007
		0.5	G-VGPMIL	0.947 ± 0.0079	0.629 ± 0.0079	0.649 ± 0.0079	**0.9921 ± 0.0079**	0.7305 ± 0.0079	0.7137 ± 0.0079	0.6788 ± 0.0079	**0.9807 ± 0.0079**
		1.0	G-VGPMIL	0.9677 ± 0.0045	0.8054 ± 0.0045	0.6071 ± 0.0045	0.9906 ± 0.0045	0.8511 ± 0.0045	0.8613 ± 0.0045	0.6551 ± 0.0045	0.9777 ± 0.0045
	30.0	2.0	G-VGPMIL	0.98 ± 0.0015	0.895 ± 0.0015	0.5211 ± 0.0015	0.9888 ± 0.0015	0.9223 ± 0.0015	0.9346 ± 0.0015	0.5767 ± 0.0015	0.9747 ± 0.0015
		4.0	G-VGPMIL	**0.9801 ± 0.002**	0.8943 ± 0.002	**0.112 ± 0.002**	0.9896 ± 0.002	0.9218 ± 0.002	0.9333 ± 0.002	**0.2619 ± 0.002**	0.9767 ± 0.002
		-	VGPMIL	0.9799 ± 0.0004	**0.8957 ± 0.0004**	0.2353 ± 0.0004	0.9877 ± 0.0004	**0.9245 ± 0.0004**	**0.9367 ± 0.0004**	0.3562 ± 0.0004	0.9738 ± 0.0004
		0.5	G-VGPMIL	0.9757 ± 0.0018	0.8604 ± 0.0018	0.6848 ± 0.0018	**0.9664 ± 0.0018**	**0.8916 ± 0.0018**	**0.9027 ± 0.0018**	0.6931 ± 0.0018	**0.9346 ± 0.0018**
		1.0	G-VGPMIL	0.9757 ± 0.0018	0.8604 ± 0.0018	0.6816 ± 0.0018	0.9664 ± 0.0018	0.8914 ± 0.0018	0.9025 ± 0.0018	0.6931 ± 0.0018	0.9339 ± 0.0018
	1.0	2.0	G-VGPMIL	**0.9757 ± 0.0017**	**0.8606 ± 0.0017**	0.6765 ± 0.0017	0.9663 ± 0.0017	0.8911 ± 0.0017	0.9023 ± 0.0017	0.6931 ± 0.0017	0.9333 ± 0.0017
		4.0	G-VGPMIL	0.9756 ± 0.0017	0.86 ± 0.0017	**0.6691 ± 0.0017**	0.9662 ± 0.0017	0.8906 ± 0.0017	0.9019 ± 0.0017	**0.6931 ± 0.0017**	0.9329 ± 0.0017
		-	VGPMIL	0.9756 ± 0.0017	0.8601 ± 0.0017	0.6696 ± 0.0017	0.9662 ± 0.0017	0.8907 ± 0.0017	0.902 ± 0.0017	0.6931 ± 0.0017	0.933 ± 0.0017
		0.5	G-VGPMIL	0.981 ± 0.0017	0.8945 ± 0.0017	0.6216 ± 0.0017	0.9944 ± 0.0017	0.9148 ± 0.0017	0.9253 ± 0.0017	0.6438 ± 0.0017	0.9882 ± 0.0017
		1.0	G-VGPMIL	0.9892 ± 0.0007	0.9428 ± 0.0007	0.5687 ± 0.0007	0.9961 ± 0.0007	0.9537 ± 0.0007	0.961 ± 0.0007	0.5807 ± 0.0007	0.9913 ± 0.0007
	3.238	2.0	G-VGPMIL	0.9897 ± 0.0012	0.9478 ± 0.0012	0.4124 ± 0.0012	0.9963 ± 0.0012	0.9627 ± 0.0012	0.9695 ± 0.0012	0.374 ± 0.0012	0.9913 ± 0.0012
		4.0	G-VGPMIL	0.9897 ± 0.0005	0.9453 ± 0.0005	**0.0429 ± 0.0005**	0.9974 ± 0.0005	0.9576 ± 0.0005	0.964 ± 0.0005	**0.1379 ± 0.0005**	0.9934 ± 0.0005
		-	VGPMIL	**0.9924 ± 0.0006**	**0.9607 ± 0.0006**	0.0957 ± 0.0006	**0.9978 ± 0.0006**	**0.9691 ± 0.0006**	**0.9744 ± 0.0006**	0.1597 ± 0.0006	**0.9944 ± 0.0006**
		0.5	G-VGPMIL	0.9747 ± 0.004	0.853 ± 0.004	0.6394 ± 0.004	0.9957 ± 0.004	0.8838 ± 0.004	0.8944 ± 0.004	0.6633 ± 0.004	0.9901 ± 0.004
		1.0	G-VGPMIL	0.9876 ± 0.0014	0.9341 ± 0.0014	0.5898 ± 0.0014	0.9965 ± 0.0014	0.949 ± 0.0014	0.9565 ± 0.0014	0.6124 ± 0.0014	0.9916 ± 0.0014
200	5.477	2.0	G-VGPMIL	0.9907 ± 0.0009	0.9527 ± 0.0009	0.4481 ± 0.0009	0.9967 ± 0.0009	0.9653 ± 0.0009	0.9714 ± 0.0009	0.4349 ± 0.0009	0.9924 ± 0.0009
		4.0	G-VGPMIL	0.9889 ± 0.0004	0.941 ± 0.0004	**0.0406 ± 0.0004**	0.9972 ± 0.0004	0.9518 ± 0.0004	0.959 ± 0.0004	**0.1378 ± 0.0004**	0.9929 ± 0.0004
		-	VGPMIL	**0.9918 ± 0.0005**	**0.958 ± 0.0005**	0.0999 ± 0.0005	**0.9976 ± 0.0005**	**0.9684 ± 0.0005**	**0.9738 ± 0.0005**	0.1645 ± 0.0005	**0.9939 ± 0.0005**
		0.5	G-VGPMIL	0.9608 ± 0.0029	0.7519 ± 0.0029	0.6473 ± 0.0029	**0.9964 ± 0.0029**	0.8155 ± 0.0029	0.8213 ± 0.0029	0.6734 ± 0.0029	**0.9916 ± 0.0029**
		1.0	G-VGPMIL	0.9777 ± 0.0019	0.8731 ± 0.0019	0.6036 ± 0.0019	0.9953 ± 0.0019	0.8984 ± 0.0019	0.9093 ± 0.0019	0.6427 ± 0.0019	0.9895 ± 0.0019
	17.738	2.0	G-VGPMIL	**0.985 ± 0.0021**	**0.9219 ± 0.0021**	0.5127 ± 0.0021	0.9928 ± 0.0021	**0.9419 ± 0.0021**	**0.9513 ± 0.0021**	0.5474 ± 0.0021	0.9843 ± 0.0021
		4.0	G-VGPMIL	0.9816 ± 0.0018	0.9015 ± 0.0018	**0.0723 ± 0.0018**	0.9893 ± 0.0018	0.9253 ± 0.0018	0.9362 ± 0.0018	**0.2183 ± 0.0018**	0.9768 ± 0.0018
		-	VGPMIL	0.9823 ± 0.0008	0.9093 ± 0.0008	0.1897 ± 0.0008	0.9901 ± 0.0008	0.9347 ± 0.0008	0.9457 ± 0.0008	0.3013 ± 0.0008	0.9792 ± 0.0008
		0.5	G-VGPMIL	0.9444 ± 0.002	0.6074 ± 0.002	0.6487 ± 0.002	**0.9934 ± 0.002**	0.7156 ± 0.002	0.6957 ± 0.002	0.6791 ± 0.002	**0.9859 ± 0.002**
		1.0	G-VGPMIL	0.9656 ± 0.0013	0.789 ± 0.0013	0.6077 ± 0.0013	0.9915 ± 0.0013	0.8402 ± 0.0013	0.8492 ± 0.0013	0.6572 ± 0.0013	0.9817 ± 0.0013
	30.0	2.0	G-VGPMIL	0.9795 ± 0.0014	0.8915 ± 0.0014	0.5185 ± 0.0014	0.9878 ± 0.0014	0.9182 ± 0.0014	0.9308 ± 0.0014	0.5702 ± 0.0014	0.9725 ± 0.0014
		4.0	G-VGPMIL	0.9793 ± 0.0017	0.8866 ± 0.0017	**0.0884 ± 0.0017**	0.9895 ± 0.0017	0.9117 ± 0.0017	0.9232 ± 0.0017	**0.2585 ± 0.0017**	0.9763 ± 0.0017
		-	VGPMIL	**0.9802 ± 0.0005**	**0.8967 ± 0.0005**	0.2346 ± 0.0005	0.9879 ± 0.0005	**0.9253 ± 0.0005**	**0.9373 ± 0.0005**	0.3554 ± 0.0005	0.974 ± 0.0005

Tabla A.1.: Resultados obtenidos en MNIST con PCA. Para cada configuración del número de inducing points, lengthscale y k, se señala en negrita el mejor resultado de cada métrica.

Num. Inducing	Lengthscale	k	Modelo	Inst. Acc.	Inst. F1	Inst. Log Loss	Inst. AUC	Bag Acc.	Bag F1	Bag Log Loss	Bag AUC
50	1.0	0.5	G-VGPMIL	**0.9014 ± 0.0026**	**0.0 ± 0.0**	**0.6931 ± 0.0**	**0.5 ± 0.0**	**0.3905 ± 0.0003**	**0.0 ± 0.0**	**0.6931 ± 0.0**	**0.5 ± 0.0**
		1.0	G-VGPMIL	0.9014 ± 0.0026	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0	0.3905 ± 0.0003	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0
		2.0	G-VGPMIL	0.9014 ± 0.0026	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0	0.3905 ± 0.0003	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0
		4.0	G-VGPMIL	0.9014 ± 0.0026	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0	0.3905 ± 0.0003	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0
	14.5	0.5	G-VGPMIL	0.9173 ± 0.0151	0.2514 ± 0.2087	0.6365 ± 0.0008	0.83 ± 0.0312	0.5184 ± 0.1146	0.3138 ± 0.231	0.6915 ± 0.0025	0.6548 ± 0.082
		1.0	G-VGPMIL	0.9438 ± 0.0116	0.5992 ± 0.1227	0.6069 ± 0.0016	0.8863 ± 0.0225	0.7155 ± 0.0809	0.6899 ± 0.1141	0.682 ± 0.0064	0.7903 ± 0.06
		2.0	G-VGPMIL	**0.9492 ± 0.0046**	**0.7408 ± 0.0237**	0.5513 ± 0.0105	0.9361 ± 0.011	**0.821 ± 0.012**	**0.8518 ± 0.0156**	0.64 ± 0.0195	0.8964 ± 0.0247
		4.0	G-VGPMIL	0.9238 ± 0.0144	0.6948 ± 0.0464	**0.3237 ± 0.076**	**0.9564 ± 0.0054**	0.7895 ± 0.0263	0.8474 ± 0.0159	**0.4712 ± 0.0574**	**0.9253 ± 0.0112**
	28.0	0.5	G-VGPMIL	0.9314 ± 0.0087	0.4651 ± 0.0913	0.6399 ± 0.0005	0.9345 ± 0.0114	0.6291 ± 0.055	0.5589 ± 0.0894	0.6855 ± 0.0022	0.8343 ± 0.0255
		1.0	G-VGPMIL	0.9532 ± 0.0072	0.6949 ± 0.0618	0.6026 ± 0.0025	0.9517 ± 0.0074	0.7744 ± 0.0409	0.7745 ± 0.0511	0.6675 ± 0.0057	0.8879 ± 0.0187
		2.0	G-VGPMIL	**0.9621 ± 0.0019**	**0.8016 ± 0.0104**	0.5295 ± 0.0195	0.9652 ± 0.0032	**0.856 ± 0.0086**	0.8788 ± 0.008	0.6055 ± 0.0135	0.9267 ± 0.0084
		4.0	G-VGPMIL	0.9569 ± 0.0077	0.793 ± 0.0261	**0.1663 ± 0.0384**	**0.9682 ± 0.0021**	0.8497 ± 0.0188	**0.8798 ± 0.0128**	**0.3486 ± 0.0263**	**0.9341 ± 0.0098**
	406.0	0.5	G-VGPMIL	0.9221 ± 0.0022	0.3471 ± 0.0269	0.6504 ± 0.0001	**0.976 ± 0.0017**	0.5624 ± 0.015	0.4397 ± 0.0303	0.6855 ± 0.0005	**0.9394 ± 0.007**
		1.0	G-VGPMIL	0.9368 ± 0.0018	0.5295 ± 0.0216	0.6121 ± 0.0003	0.9753 ± 0.0013	0.6669 ± 0.0139	0.6244 ± 0.0218	0.6732 ± 0.0011	0.9381 ± 0.0066
		2.0	G-VGPMIL	0.9545 ± 0.0022	0.7065 ± 0.0149	0.5479 ± 0.0004	0.9743 ± 0.0012	0.7832 ± 0.0116	0.7857 ± 0.0136	0.6385 ± 0.0024	0.936 ± 0.0055
		4.0	G-VGPMIL	**0.9623 ± 0.0014**	**0.7834 ± 0.0076**	**0.4149 ± 0.0065**	0.9723 ± 0.0025	**0.8411 ± 0.0059**	**0.8569 ± 0.0066**	**0.5484 ± 0.0054**	0.9317 ± 0.0031
	784.0	0.5	G-VGPMIL	0.9091 ± 0.0021	0.145 ± 0.0171	0.6514 ± 0.0002	**0.9743 ± 0.0015**	0.458 ± 0.0089	0.1991 ± 0.0231	0.6884 ± 0.0005	**0.9356 ± 0.0044**
		1.0	G-VGPMIL	0.9099 ± 0.0021	0.1588 ± 0.0287	0.6156 ± 0.0003	0.9729 ± 0.0025	0.4652 ± 0.0156	0.2175 ± 0.0392	0.6847 ± 0.0011	0.9323 ± 0.0028
		2.0	G-VGPMIL	0.9104 ± 0.0027	0.1665 ± 0.0449	0.5591 ± 0.0005	0.9711 ± 0.0037	0.4695 ± 0.0242	0.2276 ± 0.0591	0.6799 ± 0.0026	0.9282 ± 0.0022
		4.0	G-VGPMIL	**0.9134 ± 0.0042**	**0.2136 ± 0.0699**	**0.4921 ± 0.0013**	0.9692 ± 0.0046	**0.4934 ± 0.0373**	**0.2847 ± 0.0833**	**0.6733 ± 0.0056**	0.9236 ± 0.0029
	1.0	-	VGPMIL	0.9014 ± 0.0026	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0	0.3905 ± 0.0003	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0
	14.5	-	VGPMIL	0.929 ± 0.007	0.7033 ± 0.0277	0.3866 ± 0.0307	0.9531 ± 0.0058	0.7989 ± 0.0152	0.8516 ± 0.0101	0.5217 ± 0.0357	0.9216 ± 0.0124
	28.0	-	VGPMIL	0.958 ± 0.0025	**0.7959 ± 0.0132**	**0.2749 ± 0.0101**	0.9679 ± 0.002	**0.8557 ± 0.0103**	**0.8842 ± 0.0084**	**0.426 ± 0.0161**	**0.9338 ± 0.007**
	406.0	-	VGPMIL	**0.9607 ± 0.0018**	0.7682 ± 0.0074	0.4063 ± 0.002	**0.9723 ± 0.0025**	0.829 ± 0.0033	0.8427 ± 0.0035	0.5526 ± 0.0053	0.9316 ± 0.0031
	784.0	-	VGPMIL	**0.9137 ± 0.0041**	**0.2183 ± 0.0687**	**0.4928 ± 0.0008**	**0.9691 ± 0.0046**	**0.4958 ± 0.0369**	**0.2904 ± 0.082**	**0.6729 ± 0.0054**	**0.9235 ± 0.0029**
100	1.0	0.5	G-VGPMIL	0.9014 ± 0.0026	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0	0.3905 ± 0.0003	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0
		1.0	G-VGPMIL	0.9014 ± 0.0026	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0	0.3905 ± 0.0003	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0
		2.0	G-VGPMIL	0.9014 ± 0.0026	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0	0.3905 ± 0.0003	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0
		4.0	G-VGPMIL	0.9014 ± 0.0026	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0	0.3905 ± 0.0003	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0
	14.5	0.5	G-VGPMIL	0.948 ± 0.0053	0.645 ± 0.0523	0.6301 ± 0.001	0.9156 ± 0.0093	0.743 ± 0.0359	0.7332 ± 0.0485	0.6806 ± 0.0033	0.8255 ± 0.0206
		1.0	G-VGPMIL	**0.9618 ± 0.0022**	**0.7783 ± 0.0169**	0.5958 ± 0.0035	0.9415 ± 0.0087	0.8376 ± 0.0145	0.8528 ± 0.0169	0.6559 ± 0.0092	0.8916 ± 0.0104
		2.0	G-VGPMIL	0.9515 ± 0.0035	0.7747 ± 0.0167	**0.5125 ± 0.0169**	**0.9617 ± 0.0082**	**0.8406 ± 0.0121**	**0.8754 ± 0.009**	**0.5702 ± 0.0216**	**0.9327 ± 0.007**
		4.0	G-VGPMIL	0.9591 ± 0.0084	0.7988 ± 0.0261	**0.1784 ± 0.0218**	0.9559 ± 0.0127	0.853 ± 0.0195	0.8807 ± 0.0113	**0.3653 ± 0.0113**	0.9258 ± 0.0126
	28.0	0.5	G-VGPMIL	0.9363 ± 0.0046	0.5236 ± 0.0499	0.6375 ± 0.0004	0.9611 ± 0.0031	0.6644 ± 0.0286	0.6203 ± 0.0451	0.6817 ± 0.0019	0.8881 ± 0.0171
		1.0	G-VGPMIL	0.9579 ± 0.0025	0.7324 ± 0.0184	0.5997 ± 0.0008	0.9704 ± 0.0018	0.799 ± 0.0182	0.8041 ± 0.0208	0.6606 ± 0.0036	0.9218 ± 0.0104
		2.0	G-VGPMIL	**0.9691 ± 0.0019**	**0.8392 ± 0.0126**	0.5266 ± 0.0125	**0.9764 ± 0.0037**	**0.8789 ± 0.0094**	**0.8988 ± 0.0078**	0.5862 ± 0.0158	**0.9456 ± 0.0067**
		4.0	G-VGPMIL	**0.969 ± 0.0041**	**0.8436 ± 0.0165**	**0.1214 ± 0.0129**	0.9786 ± 0.003	**0.8836 ± 0.0108**	**0.9045 ± 0.008**	**0.2943 ± 0.017**	0.9503 ± 0.0053
	406.0	0.5	G-VGPMIL	0.9255 ± 0.0019	0.3932 ± 0.0119	0.6499 ± 0.0002	**0.9863 ± 0.0011**	0.5872 ± 0.0073	0.488 ± 0.0134	0.6838 ± 0.0003	0.9554 ± 0.0041
		1.0	G-VGPMIL	0.9417 ± 0.0015	0.5813 ± 0.0211	0.6112 ± 0.0004	0.9837 ± 0.0015	0.6999 ± 0.0146	0.6739 ± 0.0211	0.6698 ± 0.001	**0.9584 ± 0.0036**
		2.0	G-VGPMIL	0.9588 ± 0.0015	0.7418 ± 0.0093	0.5416 ± 0.0065	0.9783 ± 0.003	0.8082 ± 0.0098	0.815 ± 0.0113	0.6292 ± 0.0061	0.9457 ± 0.0059
		4.0	G-VGPMIL	**0.9583 ± 0.0047**	**0.7469 ± 0.037**	**0.3732 ± 0.0374**	0.9699 ± 0.0037	**0.8142 ± 0.0256**	**0.8246 ± 0.0316**	**0.5445 ± 0.0055**	0.9276 ± 0.006
	784.0	0.5	G-VGPMIL	0.9102 ± 0.0029	0.1633 ± 0.0314	0.6513 ± 0.0002	**0.978 ± 0.0023**	0.4675 ± 0.0155	0.2235 ± 0.04	0.6877 ± 0.0005	**0.9453 ± 0.0039**
		1.0	G-VGPMIL	0.9122 ± 0.0029	0.1973 ± 0.0313	0.6154 ± 0.0004	0.9727 ± 0.0025	0.4842 ± 0.0157	0.2657 ± 0.0385	0.6834 ± 0.0012	0.9333 ± 0.0051
		2.0	G-VGPMIL	0.9145 ± 0.0025	0.235 ± 0.0308	0.5586 ± 0.0006	0.9674 ± 0.0023	0.5035 ± 0.0166	0.3119 ± 0.0383	0.677 ± 0.0022	0.9217 ± 0.0059
		4.0	G-VGPMIL	0.9205 ± 0.0021	0.324 ± 0.0361	0.4909 ± 0.0008	0.9631 ± 0.002	0.5501 ± 0.0202	0.4145 ± 0.0418	0.666 ± 0.0036	0.9122 ± 0.0064
	1.0	-	VGPMIL	0.9014 ± 0.0026	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0	0.3905 ± 0.0003	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0
	14.5	-	VGPMIL	0.9527 ± 0.0026	0.782 ± 0.0129	0.2892 ± 0.0142	0.9641 ± 0.0082	0.8461 ± 0.0075	0.8802 ± 0.0061	0.4241 ± 0.0147	0.9377 ± 0.0067
	28.0	-	VGPMIL	**0.9684 ± 0.0013**	**0.8416 ± 0.0086**	**0.233 ± 0.0071**	**0.9782 ± 0.0023**	**0.8826 ± 0.0044**	**0.9043 ± 0.0032**	**0.3759 ± 0.0132**	**0.9498 ± 0.0042**
	406.0	-	VGPMIL	**0.9607 ± 0.0018**	**0.768 ± 0.0093**	**0.4047 ± 0.0046**	**0.9714 ± 0.003**	**0.8293 ± 0.0055**	**0.843 ± 0.006**	**0.5489 ± 0.0056**	**0.9308 ± 0.0043**
	784.0	-	VGPMIL	0.9206 ± 0.0022	0.3263 ± 0.0358	0.4913 ± 0.0008	0.9631 ± 0.002	0.5514 ± 0.02	0.4171 ± 0.0412	0.6658 ± 0.0035	0.9122 ± 0.0064
200	1.0	0.5	G-VGPMIL	0.9014 ± 0.0026	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0	0.3905 ± 0.0003	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0
		1.0	G-VGPMIL	0.9014 ± 0.0026	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0	0.3905 ± 0.0003	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0
		2.0	G-VGPMIL	0.9014 ± 0.0026	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0	0.3905 ± 0.0003	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0
		4.0	G-VGPMIL	0.9014 ± 0.0026	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0	0.3905 ± 0.0003	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0
	14.5	0.5	G-VGPMIL	0.9541 ± 0.0023	0.7032 ± 0.0142	0.6244 ± 0.0016	0.947 ± 0.006	0.7787 ± 0.0105	0.7812 ± 0.0122	0.6689 ± 0.005	0.8728 ± 0.0098
		1.0	G-VGPMIL	**0.9648 ± 0.0025**	**0.7951 ± 0.0107**	**0.5862 ± 0.0044**	**0.9585 ± 0.0052**	**0.8464 ± 0.0028**	**0.8606 ± 0.0023**	**0.6348 ± 0.0105**	**0.9076 ± 0.0016**
		2.0	G-VGPMIL	0.9641 ± 0.0037	0.8184 ± 0.0146	0.493 ± 0.0151	0.9663 ± 0.006	0.8692 ± 0.0086	0.8926 ± 0.0078	0.5242 ± 0.015	0.9316 ± 0.0074
		4.0	G-VGPMIL	**0.9703 ± 0.0029**	**0.8404 ± 0.0145**	**0.135 ± 0.0061**	0.9669 ± 0.0047	**0.8817 ± 0.0128**	**0.8988 ± 0.0118**	**0.3347 ± 0.0162**	0.9288 ± 0.0106
	28.0	0.5	G-VGPMIL	0.9461 ± 0.0026	0.6287 ± 0.0127	0.6364 ± 0.0006	0.9753 ± 0.0039	0.728 ± 0.0065	0.7149 ± 0.0083	0.6763 ± 0.001	0.9353 ± 0.0079
		1.0	G-VGPMIL	0.9646 ± 0.0023	0.7863 ± 0.0109	0.597 ± 0.0007	**0.98 ± 0.0035**	0.8381 ± 0.0062	0.8487 ± 0.0067	0.6507 ± 0.0018	**0.9502 ± 0.0055**
		2.0	G-VGPMIL	0.9738 ± 0.0024	0.8651 ± 0.0099	0.5185 ± 0.0063	0.9827 ± 0.0035	0.8993 ± 0.0075	**0.9164 ± 0.0066**	0.5613 ± 0.0088	0.9605 ± 0.0048
		4.0	G-VGPMIL	**0.9757 ± 0.0014**	**0.8694 ± 0.0084**	**0.091 ± 0.0046**	0.9831 ± 0.0025	**0.9009 ± 0.0066**	0.9151 ± 0.0069	**0.2617 ± 0.0102**	**0.9687 ± 0.0042**
	406.0	0.5	G-VGPMIL	0.9334 ± 0.0068	0.4892 ± 0.0637	0.6494 ± 0.0004	**0.9893 ± 0.0016**	0.6443 ± 0.0358	0.5856 ± 0.0565	0.6815 ± 0.0012	0.9225 ± 0.002
		1.0	G-VGPMIL	0.9496 ± 0.0059	0.6559 ± 0.0507	0.6096 ± 0.0015	0.9863 ± 0.0019	0.7454 ± 0.0321	0.7352 ± 0.0421	0.6644 ± 0.0037	0.9441 ± 0.0045
		2.0	G-VGPMIL	**0.9615 ± 0.0016**	**0.7641 ± 0.0095**	0.5375 ± 0.0089	0.9807 ± 0.0029	**0.8223 ± 0.0079**	**0.8315 ± 0.0094**	0.62 ± 0.0097	0.9502 ± 0.006
		4.0	G-VGPMIL	0.9571 ± 0.0039	0.7346 ± 0.0281	**0.3525 ± 0.0338**	0.971 ± 0.0026	0.8033 ± 0.0183	0.8119 ± 0.0232	**0.5374 ± 0.0051**	0.9288 ± 0.0046
	784.0	0.5	G-VGPMIL	0.9123 ± 0.002	0.1988 ± 0.0137	0.651 ± 0.0002	**0.9846 ± 0.0016**	0.4855 ± 0.0071	0.2695 ± 0.0174	0.6865 ± 0.0002	**0.96 ± 0.0036**
		1.0	G-VGPMIL	0.9141 ± 0.0022	0.2285 ± 0.0096	0.6149 ± 0.0004	0.9797 ± 0.0015	0.4999 ± 0.0058	0.3042 ± 0.0137	0.6815 ± 0.0005	0.9476 ± 0.0045
		2.0	G-VGPMIL	0.916 ± 0.0018	0.2585 ± 0.0101	0.5579 ± 0.0007	**0.9739 ± 0.0015**	0.5145 ± 0.0075	0.3381 ± 0.017	0.6743 ± 0.0008	0.9344 ± 0.0051
		4.0	G-VGPMIL	0.922 ± 0.0018	0.3464 ± 0.0122	0.4892 ± 0.0019	0.9687 ± 0.0018	0.562 ± 0.0074	0.4392 ± 0.0145	0.6625 ± 0.0012	0.9228 ± 0.0052
	1.0	-	VGPMIL	0.9014 ± 0.0026	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0	0.3905 ± 0.0003	0.0 ± 0.0	0.6931 ± 0.0	0.5 ± 0.0
	14.5	-	VGPMIL	**0.9678 ± 0.0038**	**0.8357 ± 0.0166**	**0.2362 ± 0.0062**	0.9698 ± 0.0059	**0.8806 ± 0.007**	**0.9016 ± 0.0057**	**0.374 ± 0.0104**	**0.9376 ± 0.0071**
	28.0	-	VGPMIL	0.9744 ± 0.0024	0.8691 ± 0.0104	0.1984 ± 0.004	0.983 ± 0.0034	0.9038 ± 0.0079	0.9205 ± 0.0067	0.3275 ± 0.0059	0.961 ± 0.0049
	406.0	-	VGPMIL	0.9609 ± 0.0019	0.7687 ± 0.008	0.3981 ± 0.0009	0.9735 ± 0.0015	0.829 ± 0.0054	0.8424 ± 0.0056	0.541 ± 0.0041	0.9343 ± 0.0045
	784.0	-	VGPMIL	0.9222 ± 0.0017	0.3485 ± 0.0123	0.4905 ± 0.0012	0.9688 ± 0.0018	0.563 ± 0.0072	0.4413 ± 0.0141	0.6623 ± 0.0012	0.9231 ± 0.0051

Tabla A.2.: Resultados obtenidos en MNIST sin usar PCA. Para cada configuración del número de inducing points, lengthscale y *k*, se señala en negrita el mejor resultado de cada métrica.

Num. Inducing	Lengthscale	k	Model	Bag Acc.	Bag F1	Bag Log Loss	Bag AUC
		0.5	G-VGPMIL	**0.6181 ± 0.0831**	**0.4415 ± 0.0831**	**0.6931 ± 0.0831**	**0.6343 ± 0.0831**
		1.0	G-VGPMIL	0.6181 ± 0.0831	0.4415 ± 0.0831	0.6931 ± 0.0831	0.6343 ± 0.0831
	1.0	2.0	G-VGPMIL	0.6181 ± 0.0831	0.4415 ± 0.0831	0.6931 ± 0.0831	0.6343 ± 0.0831
		4.0	G-VGPMIL	0.6181 ± 0.0831	0.4415 ± 0.0831	0.6931 ± 0.0831	0.6343 ± 0.0831
		-	VGPMIL	0.6181 ± 0.0831	0.4415 ± 0.0831	0.6931 ± 0.0831	0.6343 ± 0.0831
		0.5	G-VGPMIL	0.6959 ± 0.1138	0.5537 ± 0.1138	0.6706 ± 0.1138	**0.7519 ± 0.1138**
		1.0	G-VGPMIL	0.6965 ± 0.1397	0.5678 ± 0.1397	0.6625 ± 0.1397	0.7486 ± 0.1397
	6.942	2.0	G-VGPMIL	**0.7076 ± 0.1434**	**0.5845 ± 0.1434**	**0.6554 ± 0.1434**	0.7375 ± 0.1434
		4.0	G-VGPMIL	0.6959 ± 0.1337	0.5633 ± 0.1337	0.6659 ± 0.1337	0.7158 ± 0.1337
		-	VGPMIL	0.6959 ± 0.1337	0.5633 ± 0.1337	0.6585 ± 0.1337	0.7249 ± 0.1337
		0.5	G-VGPMIL	**0.6427 ± 0.0957**	**0.4449 ± 0.0957**	0.6643 ± 0.0957	**0.859 ± 0.0957**
		1.0	G-VGPMIL	0.6427 ± 0.0957	0.4449 ± 0.0957	0.6488 ± 0.0957	0.8415 ± 0.0957
50	12.884	2.0	G-VGPMIL	0.6322 ± 0.1098	0.4418 ± 0.1098	**0.6359 ± 0.1098**	0.8133 ± 0.1098
		4.0	G-VGPMIL	0.6105 ± 0.1038	0.3971 ± 0.1038	0.669 ± 0.1038	0.7894 ± 0.1038
		-	VGPMIL	0.6216 ± 0.1119	0.4162 ± 0.1119	0.6568 ± 0.1119	0.7985 ± 0.1119
		0.5	G-VGPMIL	**0.4895 ± 0.0129**	**0.0 ± 0.0129**	**0.6885 ± 0.0129**	**0.7973 ± 0.0129**
		1.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.695 ± 0.0129	0.7736 ± 0.0129
	83.0	2.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.7266 ± 0.0129	0.7254 ± 0.0129
		4.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.8649 ± 0.0129	0.6842 ± 0.0129
		-	VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.8596 ± 0.0129	0.6889 ± 0.0129
		0.5	G-VGPMIL	**0.4895 ± 0.0129**	**0.0 ± 0.0129**	**0.694 ± 0.0129**	**0.7523 ± 0.0129**
		1.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.7028 ± 0.0129	0.7254 ± 0.0129
	166.0	2.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.7379 ± 0.0129	0.699 ± 0.0129
		4.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.8861 ± 0.0129	0.68 ± 0.0129
		-	VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.8809 ± 0.0129	0.6822 ± 0.0129
		0.5	G-VGPMIL	**0.6953 ± 0.0743**	**0.5956 ± 0.0743**	0.6931 ± 0.0743	**0.7464 ± 0.0743**
		1.0	G-VGPMIL	0.6848 ± 0.0632	0.5874 ± 0.0632	0.6931 ± 0.0632	0.7431 ± 0.0632
	1.0	2.0	G-VGPMIL	0.6848 ± 0.0632	0.5874 ± 0.0632	0.6931 ± 0.0632	0.7431 ± 0.0632
		4.0	G-VGPMIL	0.6848 ± 0.0632	0.5874 ± 0.0632	**0.6931 ± 0.0632**	0.7431 ± 0.0632
		-	VGPMIL	0.6848 ± 0.0632	0.5874 ± 0.0632	0.6931 ± 0.0632	0.7431 ± 0.0632
		0.5	G-VGPMIL	0.6538 ± 0.1069	0.483 ± 0.1069	0.6724 ± 0.1069	0.7716 ± 0.1069
		1.0	G-VGPMIL	0.6538 ± 0.1215	0.4983 ± 0.1215	0.662 ± 0.1215	**0.7788 ± 0.1215**
	6.942	2.0	G-VGPMIL	0.7175 ± 0.1316	0.6143 ± 0.1316	0.6488 ± 0.1316	0.7657 ± 0.1316
		4.0	G-VGPMIL	0.7287 ± 0.1366	0.6191 ± 0.1366	0.6427 ± 0.1366	0.7486 ± 0.1366
		-	VGPMIL	**0.7392 ± 0.1466**	**0.6307 ± 0.1466**	**0.6395 ± 0.1466**	0.7583 ± 0.1466
		0.5	G-VGPMIL	0.6316 ± 0.1077	0.4412 ± 0.1077	0.665 ± 0.1077	**0.8615 ± 0.1077**
		1.0	G-VGPMIL	**0.6427 ± 0.0957**	**0.4449 ± 0.0957**	0.6487 ± 0.0957	0.8519 ± 0.0957
100	12.884	2.0	G-VGPMIL	0.6322 ± 0.1098	0.4418 ± 0.1098	**0.6341 ± 0.1098**	0.8274 ± 0.1098
		4.0	G-VGPMIL	0.6216 ± 0.1119	0.4162 ± 0.1119	0.6623 ± 0.1119	0.7985 ± 0.1119
		-	VGPMIL	0.6216 ± 0.1119	0.4162 ± 0.1119	0.6522 ± 0.1119	0.8081 ± 0.1119
		0.5	G-VGPMIL	**0.4895 ± 0.0129**	**0.0 ± 0.0129**	**0.6885 ± 0.0129**	**0.802 ± 0.0129**
		1.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.695 ± 0.0129	0.7736 ± 0.0129
	83.0	2.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.7266 ± 0.0129	0.7254 ± 0.0129
		4.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.8649 ± 0.0129	0.6842 ± 0.0129
		-	VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.8596 ± 0.0129	0.6889 ± 0.0129
		0.5	G-VGPMIL	**0.4895 ± 0.0129**	**0.0 ± 0.0129**	**0.694 ± 0.0129**	**0.7523 ± 0.0129**
		1.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.7028 ± 0.0129	0.7254 ± 0.0129
	166.0	2.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.7379 ± 0.0129	0.699 ± 0.0129
		4.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.8861 ± 0.0129	0.68 ± 0.0129
		-	VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.8809 ± 0.0129	0.6822 ± 0.0129
		0.5	G-VGPMIL	**0.8053 ± 0.0606**	**0.7896 ± 0.0606**	0.6931 ± 0.0606	**0.8616 ± 0.0606**
		1.0	G-VGPMIL	0.8053 ± 0.0606	0.7896 ± 0.0606	0.6931 ± 0.0606	0.8616 ± 0.0606
	1.0	2.0	G-VGPMIL	0.8053 ± 0.0606	0.7896 ± 0.0606	0.6931 ± 0.0606	0.8591 ± 0.0606
		4.0	G-VGPMIL	0.8053 ± 0.0606	0.7896 ± 0.0606	**0.6931 ± 0.0606**	0.8567 ± 0.0606
		-	VGPMIL	0.8053 ± 0.0606	0.7896 ± 0.0606	0.6931 ± 0.0606	0.8567 ± 0.0606
		0.5	G-VGPMIL	0.7181 ± 0.0899	0.6298 ± 0.0899	0.6685 ± 0.0899	**0.8067 ± 0.0899**
		1.0	G-VGPMIL	0.707 ± 0.1184	0.6027 ± 0.1184	0.6588 ± 0.1184	0.797 ± 0.1184
	6.942	2.0	G-VGPMIL	0.7175 ± 0.1316	0.6143 ± 0.1316	0.6408 ± 0.1316	0.7916 ± 0.1316
		4.0	G-VGPMIL	0.7497 ± 0.1113	0.6741 ± 0.1113	0.6272 ± 0.1113	0.7788 ± 0.1113
		-	VGPMIL	**0.7608 ± 0.1162**	**0.6888 ± 0.1162**	**0.6254 ± 0.1162**	0.7815 ± 0.1162
		0.5	G-VGPMIL	**0.6427 ± 0.0957**	0.4449 ± 0.0957	0.6651 ± 0.0957	**0.8711 ± 0.0957**
		1.0	G-VGPMIL	0.6322 ± 0.1098	0.4418 ± 0.1098	0.648 ± 0.1098	0.8615 ± 0.1098
200	12.884	2.0	G-VGPMIL	0.6427 ± 0.0957	**0.47 ± 0.0957**	**0.6321 ± 0.0957**	0.8346 ± 0.0957
		4.0	G-VGPMIL	0.6322 ± 0.0992	0.4444 ± 0.0992	0.6584 ± 0.0992	0.8005 ± 0.0992
		-	VGPMIL	0.6322 ± 0.0992	0.4444 ± 0.0992	0.648 ± 0.0992	0.8128 ± 0.0992
		0.5	G-VGPMIL	**0.4895 ± 0.0129**	**0.0 ± 0.0129**	**0.6885 ± 0.0129**	**0.802 ± 0.0129**
		1.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.695 ± 0.0129	0.7736 ± 0.0129
	83.0	2.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.7266 ± 0.0129	0.7254 ± 0.0129
		4.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.8649 ± 0.0129	0.6842 ± 0.0129
		-	VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.8596 ± 0.0129	0.6889 ± 0.0129
		0.5	G-VGPMIL	**0.4895 ± 0.0129**	**0.0 ± 0.0129**	**0.694 ± 0.0129**	**0.7523 ± 0.0129**
		1.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.7028 ± 0.0129	0.7254 ± 0.0129
	166.0	2.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.7379 ± 0.0129	0.699 ± 0.0129
		4.0	G-VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.8861 ± 0.0129	0.68 ± 0.0129
		-	VGPMIL	0.4895 ± 0.0129	0.0 ± 0.0129	0.8809 ± 0.0129	0.6822 ± 0.0129

Tabla A.3.: Resultados obtenidos en musk1. Para cada configuración del número de inducing points, lengthscale y *k*, se señala en negrita el mejor resultado de cada métrica.

Num. Inducing	Lengthscale	k	Model	Bag Acc.	Bag F1	Bag Log Loss	Bag AUC
		0.5	G-VGPMIL	$\mathbf{0.6657 \pm 0.0808}$	$\mathbf{0.2133 \pm 0.0808}$	0.6931 ± 0.0808	$\mathbf{0.4679 \pm 0.0808}$
		1.0	G-VGPMIL	0.6657 ± 0.0808	0.2133 ± 0.0808	0.6931 ± 0.0808	0.4679 ± 0.0808
	1.0	2.0	G-VGPMIL	0.6657 ± 0.0808	0.2133 ± 0.0808	0.6931 ± 0.0808	0.4679 ± 0.0808
		4.0	G-VGPMIL	0.6657 ± 0.0808	0.2133 ± 0.0808	$\mathbf{0.6931 \pm 0.0808}$	0.4679 ± 0.0808
		-	VGPMIL	0.6657 ± 0.0808	0.2133 ± 0.0808	0.6931 ± 0.0808	0.4679 ± 0.0808
		0.5	G-VGPMIL	0.7167 ± 0.0793	0.5508 ± 0.0793	0.661 ± 0.0793	0.7575 ± 0.0793
		1.0	G-VGPMIL	$\mathbf{0.7452 \pm 0.1102}$	$\mathbf{0.6239 \pm 0.1102}$	0.638 ± 0.1102	$\mathbf{0.7786 \pm 0.1102}$
	6.942	2.0	G-VGPMIL	0.6976 ± 0.0776	0.5254 ± 0.0776	0.6055 ± 0.0776	0.7238 ± 0.0776
		4.0	G-VGPMIL	0.6943 ± 0.1305	0.4121 ± 0.1305	0.6997 ± 0.1305	0.677 ± 0.1305
		-	VGPMIL	0.6952 ± 0.1126	0.5078 ± 0.1126	$\mathbf{0.603 \pm 0.1126}$	0.732 ± 0.1126
		0.5	G-VGPMIL	0.6767 ± 0.0794	0.3143 ± 0.0794	0.6583 ± 0.0794	0.8068 ± 0.0794
		1.0	G-VGPMIL	0.6767 ± 0.0575	0.3356 ± 0.0575	0.6279 ± 0.0575	0.8166 ± 0.0575
50	12.884	2.0	G-VGPMIL	$\mathbf{0.7557 \pm 0.0488}$	$\mathbf{0.5991 \pm 0.0488}$	$\mathbf{0.5434 \pm 0.0488}$	$\mathbf{0.8177 \pm 0.0488}$
		4.0	G-VGPMIL	0.6571 ± 0.0685	0.2035 ± 0.0685	0.6734 ± 0.0685	0.7775 ± 0.0685
		-	VGPMIL	0.6662 ± 0.0971	0.338 ± 0.0971	0.5475 ± 0.0971	0.7987 ± 0.0971
		0.5	G-VGPMIL	$\mathbf{0.6276 \pm 0.0196}$	$\mathbf{0.0444 \pm 0.0196}$	0.6649 ± 0.0196	$\mathbf{0.8281 \pm 0.0196}$
		1.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	$\mathbf{0.6458 \pm 0.0183}$	0.8171 ± 0.0183
	83.0	2.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	0.6588 ± 0.0183	0.7661 ± 0.0183
		4.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	1.6515 ± 0.0183	0.5398 ± 0.0183
		-	VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	0.9817 ± 0.0183	0.6686 ± 0.0183
		0.5	G-VGPMIL	$\mathbf{0.6176 \pm 0.0183}$	$\mathbf{0.0 \pm 0.0183}$	0.6688 ± 0.0183	$\mathbf{0.7917 \pm 0.0183}$
		1.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	$\mathbf{0.6577 \pm 0.0183}$	0.7537 ± 0.0183
	166.0	2.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	0.6943 ± 0.0183	0.6809 ± 0.0183
		4.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	1.769 ± 0.0183	0.4501 ± 0.0183
		-	VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	1.0429 ± 0.0183	0.5823 ± 0.0183
		0.5	G-VGPMIL	$\mathbf{0.7257 \pm 0.0577}$	$\mathbf{0.4625 \pm 0.0577}$	0.6931 ± 0.0577	$\mathbf{0.5766 \pm 0.0577}$
		1.0	G-VGPMIL	0.7157 ± 0.0473	0.4522 ± 0.0473	0.6931 ± 0.0473	0.5734 ± 0.0473
	1.0	2.0	G-VGPMIL	0.7157 ± 0.0473	0.4522 ± 0.0473	0.6931 ± 0.0473	0.5734 ± 0.0473
		4.0	G-VGPMIL	0.7157 ± 0.0473	0.4522 ± 0.0473	$\mathbf{0.6931 \pm 0.0473}$	0.5734 ± 0.0473
		-	VGPMIL	0.7157 ± 0.0473	0.4522 ± 0.0473	0.6931 ± 0.0473	0.5734 ± 0.0473
		0.5	G-VGPMIL	0.6981 ± 0.098	0.48 ± 0.098	0.6654 ± 0.098	0.7366 ± 0.098
		1.0	G-VGPMIL	0.7362 ± 0.0614	0.5788 ± 0.0614	0.6434 ± 0.0614	0.7579 ± 0.0614
	6.942	2.0	G-VGPMIL	0.7357 ± 0.0348	0.6012 ± 0.0348	0.6169 ± 0.0348	$\mathbf{0.7749 \pm 0.0348}$
		4.0	G-VGPMIL	0.7138 ± 0.147	0.4244 ± 0.147	0.6791 ± 0.147	0.7175 ± 0.147
		-	VGPMIL	$\mathbf{0.7538 \pm 0.1338}$	$\mathbf{0.6061 \pm 0.1338}$	$\mathbf{0.5952 \pm 0.1338}$	0.7553 ± 0.1338
		0.5	G-VGPMIL	0.6767 ± 0.08	0.3453 ± 0.08	0.6566 ± 0.08	0.7877 ± 0.08
		1.0	G-VGPMIL	0.6671 ± 0.0834	0.3408 ± 0.0834	0.6255 ± 0.0834	0.7887 ± 0.0834
100	12.884	2.0	G-VGPMIL	$\mathbf{0.7167 \pm 0.0598}$	$\mathbf{0.5221 \pm 0.0598}$	$\mathbf{0.5472 \pm 0.0598}$	$\mathbf{0.8271 \pm 0.0598}$
		4.0	G-VGPMIL	0.6471 ± 0.0569	0.1944 ± 0.0569	0.6817 ± 0.0569	0.7983 ± 0.0569
		-	VGPMIL	0.6667 ± 0.0724	0.3453 ± 0.0724	0.5484 ± 0.0724	0.8157 ± 0.0724
		0.5	G-VGPMIL	$\mathbf{0.6276 \pm 0.0196}$	$\mathbf{0.0444 \pm 0.0196}$	0.6649 ± 0.0196	$\mathbf{0.8241 \pm 0.0196}$
		1.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	$\mathbf{0.6457 \pm 0.0183}$	0.8213 ± 0.0183
	83.0	2.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	0.6589 ± 0.0183	0.7661 ± 0.0183
		4.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	1.6517 ± 0.0183	0.5356 ± 0.0183
		-	VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	0.9817 ± 0.0183	0.6686 ± 0.0183
		0.5	G-VGPMIL	$\mathbf{0.6176 \pm 0.0183}$	$\mathbf{0.0 \pm 0.0183}$	0.6688 ± 0.0183	$\mathbf{0.7916 \pm 0.0183}$
		1.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	$\mathbf{0.6577 \pm 0.0183}$	0.7495 ± 0.0183
	166.0	2.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	0.6943 ± 0.0183	0.6809 ± 0.0183
		4.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	1.7689 ± 0.0183	0.4501 ± 0.0183
		-	VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	1.0428 ± 0.0183	0.5823 ± 0.0183
		0.5	G-VGPMIL	0.7043 ± 0.0973	0.3941 ± 0.0973	0.6931 ± 0.0973	0.5369 ± 0.0973
		1.0	G-VGPMIL	$\mathbf{0.7138 \pm 0.1129}$	$\mathbf{0.4091 \pm 0.1129}$	0.6931 ± 0.1129	$\mathbf{0.5456 \pm 0.1129}$
	1.0	2.0	G-VGPMIL	0.7138 ± 0.1129	0.4091 ± 0.1129	0.6931 ± 0.1129	0.5456 ± 0.1129
		4.0	G-VGPMIL	0.7138 ± 0.1129	0.4091 ± 0.1129	$\mathbf{0.6931 \pm 0.1129}$	0.5456 ± 0.1129
		-	VGPMIL	0.7138 ± 0.1129	0.4091 ± 0.1129	0.6931 ± 0.1129	0.5456 ± 0.1129
		0.5	G-VGPMIL	0.7167 ± 0.0598	0.5012 ± 0.0598	0.6595 ± 0.0598	0.7801 ± 0.0598
		1.0	G-VGPMIL	0.7467 ± 0.0889	0.5659 ± 0.0889	0.6334 ± 0.0889	$\mathbf{0.795 \pm 0.0889}$
	6.942	2.0	G-VGPMIL	0.7367 ± 0.0671	0.5796 ± 0.0671	0.606 ± 0.0671	0.795 ± 0.0671
		4.0	G-VGPMIL	0.6871 ± 0.0607	0.3933 ± 0.0607	0.6502 ± 0.0607	0.7298 ± 0.0607
		-	VGPMIL	$\mathbf{0.7543 \pm 0.1427}$	$\mathbf{0.6006 \pm 0.1427}$	$\mathbf{0.5723 \pm 0.1427}$	0.7772 ± 0.1427
		0.5	G-VGPMIL	0.7162 ± 0.0694	0.4249 ± 0.0694	0.6532 ± 0.0694	0.8365 ± 0.0694
		1.0	G-VGPMIL	0.7257 ± 0.0577	0.4889 ± 0.0577	0.6204 ± 0.0577	$\mathbf{0.8389 \pm 0.0577}$
200	12.884	2.0	G-VGPMIL	$\mathbf{0.7462 \pm 0.0667}$	$\mathbf{0.5664 \pm 0.0667}$	$\mathbf{0.5415 \pm 0.0667}$	0.8199 ± 0.0667
		4.0	G-VGPMIL	0.6371 ± 0.0747	0.1944 ± 0.0747	0.6632 ± 0.0747	0.7999 ± 0.0747
		-	VGPMIL	0.6767 ± 0.0575	0.3756 ± 0.0575	0.5421 ± 0.0575	0.8181 ± 0.0575
		0.5	G-VGPMIL	$\mathbf{0.6276 \pm 0.0196}$	$\mathbf{0.0444 \pm 0.0196}$	0.6649 ± 0.0196	$\mathbf{0.828 \pm 0.0196}$
		1.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	$\mathbf{0.6457 \pm 0.0183}$	0.8212 ± 0.0183
	83.0	2.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	0.6589 ± 0.0183	0.7661 ± 0.0183
		4.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	1.6514 ± 0.0183	0.5378 ± 0.0183
		-	VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	0.9816 ± 0.0183	0.6686 ± 0.0183
		0.5	G-VGPMIL	$\mathbf{0.6176 \pm 0.0183}$	$\mathbf{0.0 \pm 0.0183}$	0.6688 ± 0.0183	$\mathbf{0.7895 \pm 0.0183}$
		1.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	$\mathbf{0.6577 \pm 0.0183}$	0.7516 ± 0.0183
	166.0	2.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	0.6943 ± 0.0183	0.6809 ± 0.0183
		4.0	G-VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	1.7689 ± 0.0183	0.4543 ± 0.0183
		-	VGPMIL	0.6176 ± 0.0183	0.0 ± 0.0183	1.0428 ± 0.0183	0.5804 ± 0.0183

Tabla A.4.: Resultados obtenidos en musk2. Para cada configuración del número de inducing points, lengthscale y k, se señala en negrita el mejor resultado de cada métrica.

Num. Inducing	Lengthscale	k	Model	RSNA inst. Acc.	RSNA inst. F1	RSNA inst. Log Loss	RSNA inst. AUC	RSNA bag Acc.	RSNA bag F1	RSNA bag Log Loss	RSNA bag AUC	CQ500 bag Acc.	CQ500 bag F1	CQ500 bag Log Loss	CQ500 bag AUC
50	2.82	0.5	G-VGPMIL	0.9389 ± 0.0032	0.7805 ± 0.0032	0.6205 ± 0.0032	0.9488 ± 0.0032	**0.892 ± 0.0032**	**0.8901 ± 0.0032**	0.6578 ± 0.0032	0.9497 ± 0.0032	0.8176 ± 0.0032	0.8015 ± 0.0032	0.6653 ± 0.0032	0.9064 ± 0.0032
		1.0	G-VGPMIL	0.9365 ± 0.0071	0.7634 ± 0.0071	0.5493 ± 0.0071	0.9501 ± 0.0071	0.8907 ± 0.0071	0.8848 ± 0.0071	0.6232 ± 0.0071	0.9489 ± 0.0071	**0.8322 ± 0.0071**	**0.8103 ± 0.0071**	0.6373 ± 0.0071	0.9093 ± 0.0071
		2.0	G-VGPMIL	0.9307 ± 0.0033	0.7414 ± 0.0033	0.3688 ± 0.0033	0.9494 ± 0.0033	0.8787 ± 0.0033	0.8724 ± 0.0033	0.4061 ± 0.0033	0.9527 ± 0.0033	0.8192 ± 0.0033	0.8008 ± 0.0033	0.5088 ± 0.0033	0.9073 ± 0.0033
		4.0	G-VGPMIL	0.9405 ± 0.0013	**0.7878 ± 0.0013**	0.2423 ± 0.0013	0.8968 ± 0.0013	0.8867 ± 0.0013	0.8839 ± 0.0013	0.4012 ± 0.0013	0.9402 ± 0.0013	0.8155 ± 0.0013	0.802 ± 0.0013	0.6668 ± 0.0013	0.9071 ± 0.0013
		-	VGPMIL	**0.9406 ± 0.0018**	0.7862 ± 0.0018	**0.1718 ± 0.0018**	**0.9571 ± 0.0018**	0.888 ± 0.0018	0.8835 ± 0.0018	**0.2832 ± 0.0018**	**0.9547 ± 0.0018**	0.8204 ± 0.0018	0.8034 ± 0.0018	**0.4316 ± 0.0018**	**0.9196 ± 0.0018**
		0.5	G-VGPMIL	0.9351 ± 0.0033	0.7555 ± 0.0033	0.6257 ± 0.0033	0.948 ± 0.0033	**0.904 ± 0.0033**	**0.8956 ± 0.0033**	0.6661 ± 0.0033	0.954 ± 0.0033	0.8412 ± 0.0033	0.812 ± 0.0033	0.6706 ± 0.0033	0.9065 ± 0.0033
		1.0	G-VGPMIL	0.9384 ± 0.0021	0.7799 ± 0.0021	0.5572 ± 0.0021	0.949 ± 0.0021	0.892 ± 0.0021	0.8882 ± 0.0021	0.6287 ± 0.0021	0.956 ± 0.0021	0.8192 ± 0.0021	0.8007 ± 0.0021	0.6424 ± 0.0021	0.9092 ± 0.0021
		2.0	G-VGPMIL	0.9344 ± 0.0029	0.7774 ± 0.0029	0.4082 ± 0.0029	0.9505 ± 0.0029	0.86 ± 0.0029	0.8616 ± 0.0029	0.4843 ± 0.0029	0.9571 ± 0.0029	0.7812 ± 0.0029	0.775 ± 0.0029	0.5426 ± 0.0029	0.9134 ± 0.0029
		4.0	G-VGPMIL	0.9353 ± 0.0022	0.7517 ± 0.0022	0.23 ± 0.0022	0.9205 ± 0.0022	0.8987 ± 0.0022	0.8897 ± 0.0022	0.306 ± 0.0022	0.9585 ± 0.0022	**0.8535 ± 0.0022**	**0.8264 ± 0.0022**	0.4287 ± 0.0022	**0.9221 ± 0.0022**
		-	VGPMIL	**0.9412 ± 0.0016**	**0.7914 ± 0.0016**	**0.167 ± 0.0016**	**0.9598 ± 0.0016**	0.888 ± 0.0016	0.8853 ± 0.0016	**0.27 ± 0.0016**	**0.9594 ± 0.0016**	0.8102 ± 0.0016	0.7968 ± 0.0016	**0.4263 ± 0.0016**	0.9214 ± 0.0016
		0.5	G-VGPMIL	0.9263 ± 0.0046	0.7014 ± 0.0046	0.626 ± 0.0046	0.952 ± 0.0046	0.884 ± 0.0046	0.8685 ± 0.0046	0.6623 ± 0.0046	0.9546 ± 0.0046	0.8535 ± 0.0046	0.811 ± 0.0046	0.6672 ± 0.0046	0.9073 ± 0.0046
		1.0	G-VGPMIL	0.9341 ± 0.0033	0.7497 ± 0.0033	0.5584 ± 0.0033	0.9551 ± 0.0033	0.8933 ± 0.0033	0.8831 ± 0.0033	0.6204 ± 0.0033	0.9567 ± 0.0033	0.8465 ± 0.0033	0.817 ± 0.0033	0.6355 ± 0.0033	0.9122 ± 0.0033
		2.0	G-VGPMIL	0.9384 ± 0.0035	**0.7881 ± 0.0035**	0.423 ± 0.0035	0.9581 ± 0.0035	0.8853 ± 0.0035	0.8841 ± 0.0035	0.503 ± 0.0035	0.9592 ± 0.0035	0.7922 ± 0.0035	0.7816 ± 0.0035	0.5566 ± 0.0035	0.916 ± 0.0035
		4.0	G-VGPMIL	0.9217 ± 0.0026	0.6699 ± 0.0026	0.219 ± 0.0026	0.9615 ± 0.0026	0.8827 ± 0.0026	0.8645 ± 0.0026	0.2795 ± 0.0026	0.9596 ± 0.0026	**0.8657 ± 0.0026**	**0.8251 ± 0.0026**	**0.34 ± 0.0026**	**0.9234 ± 0.0026**
		-	VGPMIL	**0.9397 ± 0.0016**	0.7778 ± 0.0016	**0.1689 ± 0.0016**	**0.9649 ± 0.0016**	**0.8973 ± 0.0016**	**0.8914 ± 0.0016**	**0.2583 ± 0.0016**	**0.9603 ± 0.0016**	0.8314 ± 0.0016	0.8096 ± 0.0016	0.3766 ± 0.0016	0.922 ± 0.0016
100	2.82	0.5	G-VGPMIL	0.9332 ± 0.0069	0.744 ± 0.0069	0.6225 ± 0.0069	0.9471 ± 0.0069	0.8933 ± 0.0069	0.8839 ± 0.0069	0.6614 ± 0.0069	0.9558 ± 0.0069	**0.8412 ± 0.0069**	0.8122 ± 0.0069	0.6683 ± 0.0069	0.9064 ± 0.0069
		1.0	G-VGPMIL	0.9373 ± 0.0038	0.7706 ± 0.0038	0.5536 ± 0.0038	0.9483 ± 0.0038	0.8893 ± 0.0038	0.8833 ± 0.0038	0.6241 ± 0.0038	**0.9567 ± 0.0038**	0.82 ± 0.0038	0.7991 ± 0.0038	0.6403 ± 0.0038	0.9095 ± 0.0038
		2.0	G-VGPMIL	0.9343 ± 0.0038	0.7701 ± 0.0038	0.3945 ± 0.0038	0.9485 ± 0.0038	0.872 ± 0.0038	0.8709 ± 0.0038	0.4629 ± 0.0038	0.956 ± 0.0038	0.7971 ± 0.0038	0.787 ± 0.0038	0.5328 ± 0.0038	0.9103 ± 0.0038
		4.0	G-VGPMIL	0.9396 ± 0.0012	0.7783 ± 0.0012	0.2512 ± 0.0012	0.8891 ± 0.0012	**0.8947 ± 0.0012**	**0.8893 ± 0.0012**	0.3869 ± 0.0012	0.933 ± 0.0012	0.8322 ± 0.0012	**0.8122 ± 0.0012**	0.6053 ± 0.0012	0.9064 ± 0.0012
		-	VGPMIL	**0.9402 ± 0.0019**	**0.7863 ± 0.0019**	**0.1728 ± 0.0019**	**0.9562 ± 0.0019**	0.884 ± 0.0019	0.8795 ± 0.0019	**0.2823 ± 0.0019**	0.9567 ± 0.0019	0.8151 ± 0.0019	0.7996 ± 0.0019	**0.4343 ± 0.0019**	**0.9178 ± 0.0019**
		0.5	G-VGPMIL	0.937 ± 0.0017	0.7657 ± 0.0017	0.6259 ± 0.0017	0.9477 ± 0.0017	0.8987 ± 0.0017	0.8907 ± 0.0017	0.6666 ± 0.0017	0.9534 ± 0.0017	0.8347 ± 0.0017	0.808 ± 0.0017	0.6712 ± 0.0017	0.9081 ± 0.0017
		1.0	G-VGPMIL	0.939 ± 0.002	0.787 ± 0.002	0.5583 ± 0.002	0.9486 ± 0.002	0.8827 ± 0.002	0.88 ± 0.002	0.6301 ± 0.002	0.9542 ± 0.002	0.8029 ± 0.002	0.7889 ± 0.002	0.6442 ± 0.002	0.9099 ± 0.002
		2.0	G-VGPMIL	0.9345 ± 0.0024	0.7892 ± 0.0024	0.4236 ± 0.0024	0.9498 ± 0.0024	0.8387 ± 0.0024	0.8471 ± 0.0024	0.528 ± 0.0024	0.9559 ± 0.0024	0.7555 ± 0.0024	0.7592 ± 0.0024	0.5788 ± 0.0024	0.9146 ± 0.0024
		4.0	G-VGPMIL	0.9364 ± 0.0023	0.7575 ± 0.0023	0.2259 ± 0.0023	0.9229 ± 0.0023	**0.9 ± 0.0023**	**0.8912 ± 0.0023**	0.305 ± 0.0023	0.9583 ± 0.0023	**0.8478 ± 0.0023**	**0.8215 ± 0.0023**	0.4339 ± 0.0023	**0.9224 ± 0.0023**
		-	VGPMIL	**0.9413 ± 0.0016**	**0.7917 ± 0.0016**	**0.1671 ± 0.0016**	**0.9597 ± 0.0016**	0.888 ± 0.0016	0.8853 ± 0.0016	**0.2701 ± 0.0016**	**0.9593 ± 0.0016**	0.8094 ± 0.0016	0.7959 ± 0.0016	**0.4263 ± 0.0016**	0.9214 ± 0.0016
		0.5	G-VGPMIL	0.9262 ± 0.0047	0.7012 ± 0.0047	0.626 ± 0.0047	0.9519 ± 0.0047	0.8853 ± 0.0047	0.8703 ± 0.0047	0.6624 ± 0.0047	0.9547 ± 0.0047	0.8531 ± 0.0047	0.8109 ± 0.0047	0.6673 ± 0.0047	0.9076 ± 0.0047
		1.0	G-VGPMIL	0.9339 ± 0.0034	0.7487 ± 0.0034	0.5584 ± 0.0034	0.9551 ± 0.0034	0.892 ± 0.0034	0.8814 ± 0.0034	0.6205 ± 0.0034	0.9569 ± 0.0034	0.8473 ± 0.0034	0.8175 ± 0.0034	0.6356 ± 0.0034	0.9121 ± 0.0034
		2.0	G-VGPMIL	0.9351 ± 0.0035	0.7646 ± 0.0035	0.4183 ± 0.0035	0.9586 ± 0.0035	0.8853 ± 0.0035	0.8794 ± 0.0035	0.5124 ± 0.0035	0.9589 ± 0.0035	0.82 ± 0.0035	0.8002 ± 0.0035	0.5583 ± 0.0035	0.916 ± 0.0035
		4.0	G-VGPMIL	0.9211 ± 0.0024	0.6662 ± 0.0024	0.221 ± 0.0024	0.9613 ± 0.0024	0.884 ± 0.0024	0.8654 ± 0.0024	0.2817 ± 0.0024	0.9596 ± 0.0024	**0.8633 ± 0.0024**	**0.821 ± 0.0024**	**0.3417 ± 0.0024**	**0.9235 ± 0.0024**
		-	VGPMIL	**0.9397 ± 0.0016**	**0.7778 ± 0.0016**	**0.1689 ± 0.0016**	**0.9649 ± 0.0016**	**0.8973 ± 0.0016**	**0.8914 ± 0.0016**	**0.2583 ± 0.0016**	**0.9604 ± 0.0016**	0.8314 ± 0.0016	0.8096 ± 0.0016	0.3766 ± 0.0016	0.922 ± 0.0016
200	2.82	0.5	G-VGPMIL	0.9344 ± 0.0042	0.7512 ± 0.0042	0.6237 ± 0.0042	0.9466 ± 0.0042	**0.8987 ± 0.0042**	**0.8892 ± 0.0042**	0.6641 ± 0.0042	0.9546 ± 0.0042	**0.8429 ± 0.0042**	0.8144 ± 0.0042	0.6697 ± 0.0042	0.9123 ± 0.0042
		1.0	G-VGPMIL	0.9372 ± 0.0048	0.7744 ± 0.0048	0.5553 ± 0.0048	0.9475 ± 0.0048	0.8853 ± 0.0048	0.8809 ± 0.0048	0.625 ± 0.0048	0.9559 ± 0.0048	0.8065 ± 0.0048	0.7924 ± 0.0048	0.6404 ± 0.0048	0.9145 ± 0.0048
		2.0	G-VGPMIL	0.9339 ± 0.0027	0.7807 ± 0.0027	0.4037 ± 0.0027	0.9482 ± 0.0027	0.8387 ± 0.0027	0.8439 ± 0.0027	0.4873 ± 0.0027	0.9579 ± 0.0027	0.7624 ± 0.0027	0.766 ± 0.0027	0.5514 ± 0.0027	0.9162 ± 0.0027
		4.0	G-VGPMIL	0.9391 ± 0.0021	0.7752 ± 0.0021	0.2531 ± 0.0021	0.889 ± 0.0021	0.8933 ± 0.0021	0.8874 ± 0.0021	0.3846 ± 0.0021	0.9313 ± 0.0021	0.8355 ± 0.0021	**0.8147 ± 0.0021**	0.5936 ± 0.0021	0.9064 ± 0.0021
		-	VGPMIL	**0.9405 ± 0.0018**	**0.7873 ± 0.0018**	**0.1718 ± 0.0018**	**0.9564 ± 0.0018**	0.8853 ± 0.0018	0.8824 ± 0.0018	**0.2807 ± 0.0018**	**0.9583 ± 0.0018**	0.8159 ± 0.0018	0.7993 ± 0.0018	**0.4336 ± 0.0018**	**0.9193 ± 0.0018**
		0.5	G-VGPMIL	0.9369 ± 0.0019	0.7655 ± 0.0019	0.626 ± 0.0019	0.9476 ± 0.0019	0.896 ± 0.0019	0.888 ± 0.0019	0.6664 ± 0.0019	0.954 ± 0.0019	0.8408 ± 0.0019	0.8146 ± 0.0019	0.6711 ± 0.0019	0.9083 ± 0.0019
		1.0	G-VGPMIL	0.939 ± 0.0017	0.786 ± 0.0017	0.5582 ± 0.0017	0.9485 ± 0.0017	0.8893 ± 0.0017	0.8861 ± 0.0017	0.6293 ± 0.0017	0.9546 ± 0.0017	0.8065 ± 0.0017	0.7921 ± 0.0017	0.6435 ± 0.0017	0.9106 ± 0.0017
		2.0	G-VGPMIL	0.9313 ± 0.0056	0.7649 ± 0.0056	0.4139 ± 0.0056	0.9495 ± 0.0056	0.8493 ± 0.0056	0.8522 ± 0.0056	0.499 ± 0.0056	0.9552 ± 0.0056	0.7808 ± 0.0056	0.7741 ± 0.0056	0.5547 ± 0.0056	0.9143 ± 0.0056
		4.0	G-VGPMIL	0.9358 ± 0.0025	0.7545 ± 0.0025	0.2282 ± 0.0025	0.9213 ± 0.0025	**0.9 ± 0.0025**	**0.8912 ± 0.0025**	0.3066 ± 0.0025	0.9583 ± 0.0025	**0.8506 ± 0.0025**	**0.8239 ± 0.0025**	0.4317 ± 0.0025	**0.9224 ± 0.0025**
		-	VGPMIL	**0.9413 ± 0.0015**	**0.7919 ± 0.0015**	**0.1671 ± 0.0015**	**0.9597 ± 0.0015**	0.888 ± 0.0015	0.8853 ± 0.0015	**0.2703 ± 0.0015**	**0.9592 ± 0.0015**	0.8094 ± 0.0015	0.796 ± 0.0015	**0.4266 ± 0.0015**	0.9214 ± 0.0015
		0.5	G-VGPMIL	0.9261 ± 0.0047	0.7007 ± 0.0047	0.626 ± 0.0047	0.9519 ± 0.0047	0.8853 ± 0.0047	0.8703 ± 0.0047	0.6624 ± 0.0047	0.9549 ± 0.0047	0.8535 ± 0.0047	0.8115 ± 0.0047	0.6673 ± 0.0047	0.9076 ± 0.0047
		1.0	G-VGPMIL	0.9341 ± 0.0033	0.7499 ± 0.0033	0.5584 ± 0.0033	0.955 ± 0.0033	0.8933 ± 0.0033	0.8831 ± 0.0033	0.6202 ± 0.0033	0.9569 ± 0.0033	0.8469 ± 0.0033	0.8176 ± 0.0033	0.6354 ± 0.0033	0.9122 ± 0.0033
		2.0	G-VGPMIL	0.9381 ± 0.0049	0.7777 ± 0.0049	0.4181 ± 0.0049	0.9581 ± 0.0049	0.8933 ± 0.0049	0.8891 ± 0.0049	0.5052 ± 0.0049	0.9594 ± 0.0049	0.8061 ± 0.0049	0.79 ± 0.0049	0.5544 ± 0.0049	0.916 ± 0.0049
		4.0	G-VGPMIL	0.9197 ± 0.0018	0.6578 ± 0.0018	0.2258 ± 0.0018	0.9608 ± 0.0018	0.8813 ± 0.0018	0.8624 ± 0.0018	0.2885 ± 0.0018	0.9589 ± 0.0018	**0.8645 ± 0.0018**	**0.8216 ± 0.0018**	**0.3457 ± 0.0018**	**0.9235 ± 0.0018**
		-	VGPMIL	**0.9397 ± 0.0016**	**0.7778 ± 0.0016**	**0.1689 ± 0.0016**	**0.9649 ± 0.0016**	**0.8973 ± 0.0016**	**0.8914 ± 0.0016**	**0.2583 ± 0.0016**	**0.9603 ± 0.0016**	0.8314 ± 0.0016	0.8096 ± 0.0016	0.3766 ± 0.0016	0.922 ± 0.0016

Tabla A.5.: Resultados obtenidos en los conjuntos RSNA y CQ500, fijando a 8 el número de características extraídas con AttCNN. Para cada configuración del número de inducing points, lengthscale y *k*, se señala en negrita el mejor resultado de cada métrica.

Num. Inducing	Lengthscale	k	Model	RSNA inst. Acc.	RSNA inst. F1	RSNA inst. Log Loss	RSNA inst. AUC	RSNA bag Acc.	RSNA bag F1	RSNA bag Log Loss	RSNA bag AUC	CQ500 bag Acc.	CQ500 bag F1	CQ500 bag Log Loss	CQ500 bag AUC
50	5.65	0.5	G-VGPMIL	0.9285 ± 0.0082	0.7436 ± 0.0082	0.604 ± 0.0082	0.9193 ± 0.0082	**0.888 ± 0.0082**	**0.8853 ± 0.0082**	0.664 ± 0.0082	0.9365 ± 0.0082	0.7906 ± 0.0082	0.7682 ± 0.0082	0.6719 ± 0.0082	0.8604 ± 0.0082
		1.0	G-VGPMIL	0.9252 ± 0.0053	0.7374 ± 0.0053	0.5428 ± 0.0053	**0.9212 ± 0.0053**	0.872 ± 0.0053	0.8735 ± 0.0053	0.6304 ± 0.0053	0.9365 ± 0.0053	0.7763 ± 0.0053	0.765 ± 0.0053	0.6491 ± 0.0053	0.864 ± 0.0053
		2.0	G-VGPMIL	0.9243 ± 0.0065	0.7376 ± 0.0065	0.3843 ± 0.0065	0.92 ± 0.0065	0.8653 ± 0.0065	0.8685 ± 0.0065	0.5197 ± 0.0065	**0.9405 ± 0.0065**	0.7615 ± 0.0065	0.7567 ± 0.0065	0.5693 ± 0.0065	0.8628 ± 0.0065
		4.0	G-VGPMIL	0.9284 ± 0.0077	0.7337 ± 0.0077	0.2976 ± 0.0077	0.8564 ± 0.0077	0.884 ± 0.0077	0.8803 ± 0.0077	0.4277 ± 0.0077	0.9157 ± 0.0077	**0.8139 ± 0.0077**	**0.7836 ± 0.0077**	0.5305 ± 0.0077	0.8559 ± 0.0077
		-	VGPMIL	**0.9297 ± 0.0083**	**0.7493 ± 0.0083**	**0.2402 ± 0.0083**	0.9159 ± 0.0083	0.8733 ± 0.0083	0.8723 ± 0.0083	**0.3968 ± 0.0083**	0.9312 ± 0.0083	0.7919 ± 0.0083	0.7708 ± 0.0083	**0.4874 ± 0.0083**	**0.8672 ± 0.0083**
	3.32	0.5	G-VGPMIL	0.9289 ± 0.008	0.7313 ± 0.008	0.6234 ± 0.008	0.9406 ± 0.008	**0.8947 ± 0.008**	0.8862 ± 0.008	0.661 ± 0.008	0.9528 ± 0.008	0.8033 ± 0.008	0.7618 ± 0.008	0.6677 ± 0.008	0.8508 ± 0.008
		1.0	G-VGPMIL	0.9247 ± 0.0091	0.7199 ± 0.0091	0.5556 ± 0.0091	0.9406 ± 0.0091	0.8827 ± 0.0091	0.8765 ± 0.0091	0.6245 ± 0.0091	0.9543 ± 0.0091	0.7928 ± 0.0091	0.7617 ± 0.0091	0.6407 ± 0.0091	0.8566 ± 0.0091
		2.0	G-VGPMIL	0.9243 ± 0.0088	0.7274 ± 0.0088	0.401 ± 0.0088	0.9414 ± 0.0088	0.868 ± 0.0088	0.8668 ± 0.0088	0.4468 ± 0.0088	0.9568 ± 0.0088	0.7747 ± 0.0088	0.7522 ± 0.0088	0.5302 ± 0.0088	0.8621 ± 0.0088
		4.0	G-VGPMIL	0.9297 ± 0.0075	0.7287 ± 0.0075	0.2745 ± 0.0075	0.8899 ± 0.0075	0.892 ± 0.0075	0.8848 ± 0.0075	0.3538 ± 0.0075	0.9435 ± 0.0075	**0.8338 ± 0.0075**	**0.7948 ± 0.0075**	0.5781 ± 0.0075	0.8755 ± 0.0075
		-	VGPMIL	**0.9334 ± 0.0088**	**0.7614 ± 0.0088**	**0.1871 ± 0.0088**	**0.95 ± 0.0088**	0.8907 ± 0.0088	**0.8872 ± 0.0088**	**0.2783 ± 0.0088**	**0.9575 ± 0.0088**	0.8104 ± 0.0088	0.7823 ± 0.0088	**0.4449 ± 0.0088**	**0.8758 ± 0.0088**
	5.65	0.5	G-VGPMIL	0.926 ± 0.0103	0.7185 ± 0.0103	0.6261 ± 0.0103	0.9394 ± 0.0103	0.888 ± 0.0103	0.8779 ± 0.0103	0.6645 ± 0.0103	0.9536 ± 0.0103	0.8094 ± 0.0103	0.7592 ± 0.0103	0.6701 ± 0.0103	0.8514 ± 0.0103
		1.0	G-VGPMIL	0.929 ± 0.0089	0.7466 ± 0.0089	0.5594 ± 0.0089	0.9411 ± 0.0089	**0.892 ± 0.0089**	0.8869 ± 0.0089	0.6264 ± 0.0089	0.9553 ± 0.0089	0.7972 ± 0.0089	0.762 ± 0.0089	0.6423 ± 0.0089	0.8548 ± 0.0089
		2.0	G-VGPMIL	0.9264 ± 0.0063	0.7555 ± 0.0063	0.4254 ± 0.0063	0.9424 ± 0.0063	0.8653 ± 0.0063	0.8677 ± 0.0063	0.5276 ± 0.0063	0.956 ± 0.0063	0.7673 ± 0.0063	0.7516 ± 0.0063	0.5756 ± 0.0063	0.859 ± 0.0063
		4.0	G-VGPMIL	0.9272 ± 0.0061	0.7116 ± 0.0061	0.2605 ± 0.0061	0.9111 ± 0.0061	0.892 ± 0.0061	0.8826 ± 0.0061	0.3208 ± 0.0061	0.9494 ± 0.0061	**0.8355 ± 0.0061**	**0.7898 ± 0.0061**	0.5235 ± 0.0061	**0.8829 ± 0.0061**
		-	VGPMIL	**0.933 ± 0.0071**	**0.7584 ± 0.0071**	**0.1857 ± 0.0071**	**0.9524 ± 0.0071**	0.8907 ± 0.0071	**0.8869 ± 0.0071**	**0.2767 ± 0.0071**	**0.9569 ± 0.0071**	0.8112 ± 0.0071	0.782 ± 0.0071	**0.4355 ± 0.0071**	0.8777 ± 0.0071
	18.82	0.5	G-VGPMIL	0.9154 ± 0.008	0.6451 ± 0.008	0.6266 ± 0.008	0.9445 ± 0.008	0.8693 ± 0.008	0.8481 ± 0.008	0.6626 ± 0.008	0.9513 ± 0.008	0.7701 ± 0.008	0.666 ± 0.008	0.67 ± 0.008	0.8416 ± 0.008
		1.0	G-VGPMIL	0.9223 ± 0.0118	0.6882 ± 0.0118	0.5593 ± 0.0118	0.9471 ± 0.0118	0.8853 ± 0.0118	0.8697 ± 0.0118	0.6235 ± 0.0118	0.9525 ± 0.0118	0.7959 ± 0.0118	0.7255 ± 0.0118	0.6419 ± 0.0118	0.8491 ± 0.0118
		2.0	G-VGPMIL	0.9283 ± 0.0118	0.7324 ± 0.0118	0.4216 ± 0.0118	0.9498 ± 0.0118	0.884 ± 0.0118	0.8763 ± 0.0118	0.5126 ± 0.0118	**0.9536 ± 0.0118**	0.7903 ± 0.0118	0.747 ± 0.0118	0.5663 ± 0.0118	0.8592 ± 0.0118
		4.0	G-VGPMIL	0.9133 ± 0.0037	0.6188 ± 0.0037	0.2519 ± 0.0037	0.9501 ± 0.0037	0.8707 ± 0.0037	0.8481 ± 0.0037	0.3162 ± 0.0037	0.9468 ± 0.0037	0.8169 ± 0.0037	0.7439 ± 0.0037	0.4823 ± 0.0037	**0.8859 ± 0.0037**
		-	VGPMIL	**0.9315 ± 0.0055**	**0.7428 ± 0.0055**	**0.1903 ± 0.0055**	**0.9592 ± 0.0055**	**0.8973 ± 0.0055**	**0.8917 ± 0.0055**	**0.2887 ± 0.0055**	0.9521 ± 0.0055	**0.8228 ± 0.0055**	**0.7877 ± 0.0055**	**0.409 ± 0.0055**	0.8807 ± 0.0055
	5.65	0.5	G-VGPMIL	0.9097 ± 0.0066	0.6016 ± 0.0066	0.627 ± 0.0066	0.9464 ± 0.0066	0.856 ± 0.0066	0.8259 ± 0.0066	0.6623 ± 0.0066	0.9488 ± 0.0066	0.756 ± 0.0066	0.637 ± 0.0066	0.67 ± 0.0066	0.8427 ± 0.0066
		1.0	G-VGPMIL	0.9181 ± 0.0067	0.66 ± 0.0067	0.5609 ± 0.0067	0.9489 ± 0.0067	0.8667 ± 0.0067	0.8456 ± 0.0067	0.6224 ± 0.0067	0.9484 ± 0.0067	0.7876 ± 0.0067	0.7043 ± 0.0067	0.6408 ± 0.0067	0.8503 ± 0.0067
		2.0	G-VGPMIL	0.928 ± 0.01	0.7263 ± 0.01	0.4279 ± 0.01	0.9517 ± 0.01	0.8813 ± 0.01	0.8727 ± 0.01	0.5135 ± 0.01	0.9478 ± 0.01	0.8029 ± 0.01	0.7535 ± 0.01	0.563 ± 0.01	0.8597 ± 0.01
		4.0	G-VGPMIL	0.9105 ± 0.003	0.6024 ± 0.003	0.2369 ± 0.003	0.958 ± 0.003	0.8653 ± 0.003	0.8414 ± 0.003	**0.3138 ± 0.003**	0.9478 ± 0.003	0.8116 ± 0.003	0.7356 ± 0.003	0.458 ± 0.003	**0.8865 ± 0.003**
		-	VGPMIL	**0.9307 ± 0.0058**	**0.7377 ± 0.0058**	**0.2009 ± 0.0058**	**0.9601 ± 0.0058**	**0.8947 ± 0.0058**	**0.8884 ± 0.0058**	0.3161 ± 0.0058	**0.9493 ± 0.0058**	**0.8252 ± 0.0058**	**0.7893 ± 0.0058**	**0.4189 ± 0.0058**	0.8812 ± 0.0058
100	5.65	0.5	G-VGPMIL	0.928 ± 0.0088	0.7403 ± 0.0088	0.6049 ± 0.0088	0.9246 ± 0.0088	0.888 ± 0.0088	0.8852 ± 0.0088	0.6614 ± 0.0088	0.9397 ± 0.0088	0.7964 ± 0.0088	0.7737 ± 0.0088	0.6687 ± 0.0088	0.8548 ± 0.0088
		1.0	G-VGPMIL	0.9274 ± 0.0091	0.7478 ± 0.0091	0.5414 ± 0.0091	0.9249 ± 0.0091	0.8813 ± 0.0091	0.8815 ± 0.0091	0.6271 ± 0.0091	0.9417 ± 0.0091	0.7839 ± 0.0091	0.7656 ± 0.0091	0.6435 ± 0.0091	0.857 ± 0.0091
		2.0	G-VGPMIL	0.9255 ± 0.0087	0.7481 ± 0.0087	0.3975 ± 0.0087	**0.9288 ± 0.0087**	0.868 ± 0.0087	0.8704 ± 0.0087	0.5356 ± 0.0087	**0.9446 ± 0.0087**	0.7685 ± 0.0087	0.7587 ± 0.0087	0.5817 ± 0.0087	0.8597 ± 0.0087
		4.0	G-VGPMIL	0.9311 ± 0.0078	0.7432 ± 0.0078	0.2841 ± 0.0078	0.8641 ± 0.0078	**0.8947 ± 0.0078**	**0.8899 ± 0.0078**	0.404 ± 0.0078	0.9188 ± 0.0078	**0.8217 ± 0.0078**	**0.788 ± 0.0078**	0.5355 ± 0.0078	0.8521 ± 0.0078
		-	VGPMIL	**0.9313 ± 0.0087**	**0.7536 ± 0.0087**	**0.2199 ± 0.0087**	0.9274 ± 0.0087	0.8907 ± 0.0087	0.8879 ± 0.0087	**0.3584 ± 0.0087**	0.9405 ± 0.0087	0.8029 ± 0.0087	0.7785 ± 0.0087	**0.4697 ± 0.0087**	**0.8677 ± 0.0087**
	3.32	0.5	G-VGPMIL	0.9252 ± 0.0082	0.7252 ± 0.0082	0.6262 ± 0.0082	0.9351 ± 0.0082	0.8907 ± 0.0082	0.8836 ± 0.0082	0.6647 ± 0.0082	0.9469 ± 0.0082	0.8057 ± 0.0082	0.7678 ± 0.0082	0.6711 ± 0.0082	0.8494 ± 0.0082
		1.0	G-VGPMIL	0.9258 ± 0.0079	0.7431 ± 0.0079	0.5601 ± 0.0079	0.9363 ± 0.0079	0.88 ± 0.0079	0.878 ± 0.0079	0.6279 ± 0.0079	0.9482 ± 0.0079	0.7948 ± 0.0079	0.7682 ± 0.0079	0.6441 ± 0.0079	0.8567 ± 0.0079
		2.0	G-VGPMIL	0.9254 ± 0.0088	0.7531 ± 0.0088	0.4236 ± 0.0088	0.9376 ± 0.0088	0.864 ± 0.0088	0.8667 ± 0.0088	0.5293 ± 0.0088	0.952 ± 0.0088	0.7627 ± 0.0088	0.7485 ± 0.0088	0.5778 ± 0.0088	0.8617 ± 0.0088
		4.0	G-VGPMIL	0.9222 ± 0.0169	0.6804 ± 0.0169	0.3108 ± 0.0169	0.872 ± 0.0169	0.884 ± 0.0169	0.8712 ± 0.0169	0.3995 ± 0.0169	0.9403 ± 0.0169	**0.8281 ± 0.0169**	**0.7817 ± 0.0169**	0.6126 ± 0.0169	0.8757 ± 0.0169
		-	VGPMIL	**0.9325 ± 0.0091**	**0.7588 ± 0.0091**	**0.1891 ± 0.0091**	**0.9499 ± 0.0091**	**0.892 ± 0.0091**	**0.8884 ± 0.0091**	**0.278 ± 0.0091**	**0.9577 ± 0.0091**	0.8088 ± 0.0091	0.7804 ± 0.0091	**0.4455 ± 0.0091**	**0.8761 ± 0.0091**
	5.65	0.5	G-VGPMIL	0.925 ± 0.0102	0.7158 ± 0.0102	0.6266 ± 0.0102	0.9381 ± 0.0102	0.888 ± 0.0102	0.8782 ± 0.0102	0.6647 ± 0.0102	0.9521 ± 0.0102	0.8143 ± 0.0102	0.7631 ± 0.0102	0.6705 ± 0.0102	0.8522 ± 0.0102
		1.0	G-VGPMIL	0.9285 ± 0.0094	0.7459 ± 0.0094	0.56 ± 0.0094	0.9394 ± 0.0094	**0.896 ± 0.0094**	**0.8911 ± 0.0094**	0.6267 ± 0.0094	0.9544 ± 0.0094	0.8062 ± 0.0094	0.7689 ± 0.0094	0.6429 ± 0.0094	0.8557 ± 0.0094
		2.0	G-VGPMIL	0.9269 ± 0.0069	0.7528 ± 0.0069	0.4215 ± 0.0069	0.9416 ± 0.0069	0.8773 ± 0.0069	0.878 ± 0.0069	0.5109 ± 0.0069	0.9553 ± 0.0069	0.7714 ± 0.0069	0.7499 ± 0.0069	0.5685 ± 0.0069	0.8589 ± 0.0069
		4.0	G-VGPMIL	0.9274 ± 0.0062	0.7126 ± 0.0062	0.2614 ± 0.0062	0.9094 ± 0.0062	0.896 ± 0.0062	0.8869 ± 0.0062	0.3209 ± 0.0062	0.95 ± 0.0062	**0.8367 ± 0.0062**	**0.7923 ± 0.0062**	0.5249 ± 0.0062	**0.8831 ± 0.0062**
		-	VGPMIL	**0.9327 ± 0.0076**	**0.758 ± 0.0076**	**0.186 ± 0.0076**	**0.9525 ± 0.0076**	0.8893 ± 0.0076	0.8857 ± 0.0076	**0.2773 ± 0.0076**	**0.9568 ± 0.0076**	0.8088 ± 0.0076	0.7794 ± 0.0076	**0.4355 ± 0.0076**	0.8781 ± 0.0076
	18.82	0.5	G-VGPMIL	0.9155 ± 0.0082	0.6432 ± 0.0082	0.6267 ± 0.0082	0.9445 ± 0.0082	0.8733 ± 0.0082	0.8526 ± 0.0082	0.6632 ± 0.0082	0.9517 ± 0.0082	0.7721 ± 0.0082	0.6705 ± 0.0082	0.6703 ± 0.0082	0.8424 ± 0.0082
		1.0	G-VGPMIL	0.9229 ± 0.0111	0.6937 ± 0.0111	0.5598 ± 0.0111	0.9471 ± 0.0111	0.8893 ± 0.0111	0.8753 ± 0.0111	0.6236 ± 0.0111	0.9524 ± 0.0111	0.7972 ± 0.0111	0.7295 ± 0.0111	0.642 ± 0.0111	0.8498 ± 0.0111
		2.0	G-VGPMIL	0.9271 ± 0.0012	0.73 ± 0.012	0.4237 ± 0.012	0.9495 ± 0.012	0.8813 ± 0.012	0.8735 ± 0.012	0.5178 ± 0.012	**0.9526 ± 0.012**	0.7948 ± 0.012	0.7496 ± 0.012	0.568 ± 0.012	0.8581 ± 0.012
		4.0	G-VGPMIL	0.914 ± 0.0031	0.6237 ± 0.0031	0.248 ± 0.0031	0.951 ± 0.0031	0.872 ± 0.0031	0.8503 ± 0.0031	0.3112 ± 0.0031	0.9469 ± 0.0031	0.8181 ± 0.0031	0.7467 ± 0.0031	0.4751 ± 0.0031	**0.8859 ± 0.0031**
		-	VGPMIL	**0.9314 ± 0.0054**	**0.7426 ± 0.0054**	**0.1903 ± 0.0054**	**0.9592 ± 0.0054**	**0.8973 ± 0.0054**	**0.8917 ± 0.0054**	**0.2887 ± 0.0054**	0.9521 ± 0.0054	**0.8236 ± 0.0054**	**0.7885 ± 0.0054**	**0.4089 ± 0.0054**	0.8807 ± 0.0054
	5.65	0.5	G-VGPMIL	0.9081 ± 0.006	0.5901 ± 0.006	0.6271 ± 0.006	0.947 ± 0.006	0.8493 ± 0.006	0.8164 ± 0.006	0.6634 ± 0.006	0.9487 ± 0.006	0.7556 ± 0.006	0.6292 ± 0.006	0.6703 ± 0.006	0.8432 ± 0.006
		1.0	G-VGPMIL	0.9191 ± 0.0067	0.6669 ± 0.0067	0.5611 ± 0.0067	0.9492 ± 0.0067	0.8707 ± 0.0067	0.8511 ± 0.0067	0.622 ± 0.0067	0.9483 ± 0.0067	0.7892 ± 0.0067	0.7106 ± 0.0067	0.6406 ± 0.0067	0.8517 ± 0.0067
		2.0	G-VGPMIL	0.9301 ± 0.0009	**0.7435 ± 0.009**	0.4294 ± 0.009	0.9518 ± 0.009	0.892 ± 0.009	0.8867 ± 0.009	0.5081 ± 0.009	0.9485 ± 0.009	0.8024 ± 0.009	0.7676 ± 0.009	0.562 ± 0.009	0.8616 ± 0.009
		4.0	G-VGPMIL	0.9126 ± 0.0038	0.6153 ± 0.0038	0.2325 ± 0.0038	0.9581 ± 0.0038	0.872 ± 0.0038	0.8505 ± 0.0038	**0.3068 ± 0.0038**	0.9478 ± 0.0038	0.8153 ± 0.0038	0.7437 ± 0.0038	0.4514 ± 0.0038	**0.8865 ± 0.0038**
		-	VGPMIL	**0.9307 ± 0.0058**	0.7377 ± 0.0058	**0.2009 ± 0.0058**	**0.9601 ± 0.0058**	**0.8947 ± 0.0058**	**0.8884 ± 0.0058**	0.3161 ± 0.0058	**0.9493 ± 0.0058**	**0.8252 ± 0.0058**	**0.7893 ± 0.0058**	**0.419 ± 0.0058**	0.8812 ± 0.0058
200	5.65	0.5	G-VGPMIL	0.9276 ± 0.0076	0.7353 ± 0.0076	0.6072 ± 0.0076	0.9297 ± 0.0076	0.8893 ± 0.0076	0.8857 ± 0.0076	0.6591 ± 0.0076	**0.9498 ± 0.0076**	0.7806 ± 0.0076	0.7599 ± 0.0076	0.6691 ± 0.0076	0.8604 ± 0.0076
		1.0	G-VGPMIL	0.9279 ± 0.0065	0.7465 ± 0.0065	0.5424 ± 0.0065	0.9313 ± 0.0065	**0.8907 ± 0.0065**	**0.8899 ± 0.0065**	0.6214 ± 0.0065	0.9489 ± 0.0065	0.7786 ± 0.0065	0.7638 ± 0.0065	0.6438 ± 0.0065	0.8611 ± 0.0065
		2.0	G-VGPMIL	0.9252 ± 0.0069	0.7392 ± 0.0069	0.3961 ± 0.0069	**0.9339 ± 0.0069**	0.868 ± 0.0069	0.8687 ± 0.0069	0.5278 ± 0.0069	0.9457 ± 0.0069	0.7702 ± 0.0069	0.756 ± 0.0069	0.5771 ± 0.0069	0.8619 ± 0.0069
		4.0	G-VGPMIL	**0.9313 ± 0.0078**	0.7476 ± 0.0078	0.282 ± 0.0078	0.8691 ± 0.0078	0.8827 ± 0.0078	0.8791 ± 0.0078	0.4073 ± 0.0078	0.9228 ± 0.0078	**0.8132 ± 0.0078**	**0.7843 ± 0.0078**	0.5703 ± 0.0078	0.8558 ± 0.0078
		-	VGPMIL	0.9311 ± 0.0099	**0.7578 ± 0.0099**	**0.214 ± 0.0099**	0.9323 ± 0.0099	0.888 ± 0.0099	0.8861 ± 0.0099	**0.35 ± 0.0099**	0.9417 ± 0.0099	0.7912 ± 0.0099	0.7725 ± 0.0099	**0.4748 ± 0.0099**	**0.8694 ± 0.0099**
	3.32	0.5	G-VGPMIL	0.9234 ± 0.0069	0.7133 ± 0.0069	0.6261 ± 0.0069	0.9331 ± 0.0069	0.8853 ± 0.0069	0.8774 ± 0.0069	0.6631 ± 0.0069	0.9394 ± 0.0069	0.7961 ± 0.0069	0.7547 ± 0.0069	0.6705 ± 0.0069	0.8303 ± 0.0069
		1.0	G-VGPMIL	0.9263 ± 0.0082	0.7365 ± 0.0082	0.5585 ± 0.0082	0.9351 ± 0.0082	**0.8907 ± 0.0082**	0.8854 ± 0.0082	0.6256 ± 0.0082	0.9404 ± 0.0082	0.7928 ± 0.0082	0.7618 ± 0.0082	0.6435 ± 0.0082	0.8374 ± 0.0082
		2.0	G-VGPMIL	0.9246 ± 0.0062	0.7289 ± 0.0062	0.4053 ± 0.0062	0.9384 ± 0.0062	0.8893 ± 0.0062	0.8852 ± 0.0062	0.5074 ± 0.0062	0.9457 ± 0.0062	0.7915 ± 0.0062	0.7648 ± 0.0062	0.565 ± 0.0062	0.8485 ± 0.0062
		4.0	G-VGPMIL	0.9212 ± 0.0127	0.6884 ± 0.0127	0.3033 ± 0.0127	0.88 ± 0.0127	0.8907 ± 0.0127	0.8841 ± 0.0127	0.3661 ± 0.0127	0.9436 ± 0.0127	**0.8271 ± 0.0127**	**0.7898 ± 0.0127**	0.5916 ± 0.0127	0.8743 ± 0.0127
		-	VGPMIL	**0.9316 ± 0.0097**	**0.7576 ± 0.0097**	**0.1895 ± 0.0097**	**0.9501 ± 0.0097**	0.8893 ± 0.0097	**0.8862 ± 0.0097**	**0.2791 ± 0.0097**	**0.9583 ± 0.0097**	0.8075 ± 0.0097	0.7802 ± 0.0097	**0.4471 ± 0.0097**	**0.8757 ± 0.0097**
	5.65	0.5	G-VGPMIL	0.9246 ± 0.0108	0.7184 ± 0.0108	0.6267 ± 0.0108	0.9374 ± 0.0108	0.8893 ± 0.0108	0.8806 ± 0.0108	0.6646 ± 0.0108	0.9534 ± 0.0108	0.8037 ± 0.0108	0.7567 ± 0.0108	0.671 ± 0.0108	0.8472 ± 0.0108
		1.0	G-VGPMIL	0.9284 ± 0.0097	0.7471 ± 0.0097	0.5602 ± 0.0097	0.9394 ± 0.0097	**0.896 ± 0.0097**	**0.8915 ± 0.0097**	0.6265 ± 0.0097	0.955 ± 0.0097	0.7976 ± 0.0097	0.7632 ± 0.0097	0.6432 ± 0.0097	0.8533 ± 0.0097
		2.0	G-VGPMIL	0.9243 ± 0.0091	0.7414 ± 0.0091	0.4226 ± 0.0091	0.9408 ± 0.0091	0.8733 ± 0.0091	0.8725 ± 0.0091	0.5254 ± 0.0091	0.9558 ± 0.0091	0.7698 ± 0.0091	0.7487 ± 0.0091	0.574 ± 0.0091	0.8593 ± 0.0091
		4.0	G-VGPMIL	0.9271 ± 0.0059	0.7121 ± 0.0059	0.2615 ± 0.0059	0.9087 ± 0.0059	0.8893 ± 0.0059	0.8807 ± 0.0059	0.3219 ± 0.0059	0.9501 ± 0.0059	**0.8367 ± 0.0059**	**0.7922 ± 0.0059**	0.5247 ± 0.0059	**0.8832 ± 0.0059**
		-	VGPMIL	**0.9327 ± 0.0076**	**0.7586 ± 0.0076**	**0.1863 ± 0.0076**	**0.9524 ± 0.0076**	0.8893 ± 0.0076	0.8857 ± 0.0076	**0.2779 ± 0.0076**	**0.9572 ± 0.0076**	0.8071 ± 0.0076	0.7783 ± 0.0076	**0.4365 ± 0.0076**	0.8779 ± 0.0076
	18.82	0.5	G-VGPMIL	0.9163 ± 0.0075	0.6494 ± 0.0075	0.6268 ± 0.0075	0.9445 ± 0.0075	0.8733 ± 0.0075	0.8531 ± 0.0075	0.6635 ± 0.0075	0.9513 ± 0.0075	0.7721 ± 0.0075	0.6715 ± 0.0075	0.6705 ± 0.0075	0.8413 ± 0.0075
		1.0	G-VGPMIL	0.9253 ± 0.0084	0.709 ± 0.0084	0.5602 ± 0.0084	0.9473 ± 0.0084	0.888 ± 0.0084	0.8753 ± 0.0084	0.6235 ± 0.0084	0.9522 ± 0.0084	0.7984 ± 0.0084	0.7365 ± 0.0084	0.6421 ± 0.0084	0.8496 ± 0.0084
		2.0	G-VGPMIL	0.9289 ± 0.0105	0.7366 ± 0.0105	0.4239 ± 0.0105	0.9497 ± 0.0105	0.876 ± 0.0105	0.8675 ± 0.0105	0.5188 ± 0.0105	**0.9527 ± 0.0105**	0.7969 ± 0.0105	0.7547 ± 0.0105	0.5685 ± 0.0105	0.8582 ± 0.0105
		4.0	G-VGPMIL	0.9129 ± 0.0041	0.6179 ± 0.0041	0.25 ± 0.0041	0.9519 ± 0.0041	0.8733 ± 0.0041	0.8523 ± 0.0041	0.3118 ± 0.0041	0.9473 ± 0.0041	0.8156 ± 0.0041	0.742 ± 0.0041	0.477 ± 0.0041	**0.8858 ± 0.0041**
		-	VGPMIL	**0.9314 ± 0.0054**	**0.7425 ± 0.0054**	**0.1902 ± 0.0054**	**0.9592 ± 0.0054**	**0.896 ± 0.0054**	**0.8904 ± 0.0054**	**0.2886 ± 0.0054**	0.9521 ± 0.0054	**0.8236 ± 0.0054**	**0.7885 ± 0.0054**	**0.4089 ± 0.0054**	0.8807 ± 0.0054
	5.65	0.5	G-VGPMIL	0.9078 ± 0.0061	0.5882 ± 0.0061	0.627 ± 0.0061	0.9471 ± 0.0061	0.8493 ± 0.0061	0.8157 ± 0.0061	0.6632 ± 0.0061	0.9489 ± 0.0061	0.7548 ± 0.0061	0.6268 ± 0.0061	0.6702 ± 0.0061	0.8436 ± 0.0061
		1.0	G-VGPMIL	0.9156 ± 0.0078	0.6415 ± 0.0078	0.5601 ± 0.0078	0.9493 ± 0.0078	0.864 ± 0.0078	0.8394 ± 0.0078	0.6228 ± 0.0078	0.9486 ± 0.0078	0.7823 ± 0.0078	0.6918 ± 0.0078	0.6405 ± 0.0078	0.8522 ± 0.0078
		2.0	G-VGPMIL	0.9218 ± 0.0112	0.6863 ± 0.0112	0.42 ± 0.0112	0.9525 ± 0.0112	0.884 ± 0.0112	0.869 ± 0.0112	0.5165 ± 0.0112	0.949 ± 0.0112	0.8032 ± 0.0112	0.7468 ± 0.0112	0.5649 ± 0.0112	0.8604 ± 0.0112
		4.0	G-VGPMIL	0.9108 ± 0.0064	0.6039 ± 0.0064	0.2419 ± 0.0064	0.9555 ± 0.0064	0.868 ± 0.0064	0.8447 ± 0.0064	0.3218 ± 0.0064	0.947 ± 0.0064	0.8141 ± 0.0064	0.7409 ± 0.0064	0.4637 ± 0.0064	**0.8869 ± 0.0064**
		-	VGPMIL	**0.9307 ± 0.0058**	**0.7377 ± 0.0058**	**0.2009 ± 0.0058**	**0.9601 ± 0.0058**	**0.8947 ± 0.0058**	**0.8884 ± 0.0058**	**0.3161 ± 0.0058**	**0.9493 ± 0.0058**	**0.8256 ± 0.0058**	**0.7899 ± 0.0058**	**0.419 ± 0.0058**	0.8812 ± 0.0058

Tabla A.6.: Resultados obtenidos en los conjuntos RSNA y CQ500, fijando a 32 el número de características extraídas con AttCNN. Para cada configuración del número de inducing points, lengthscale y k, se señala en negrita el mejor resultado de cada métrica.

Num. Inducing	Lengthscale	k	Model	RSNA inst. Acc.	RSNA inst. F1	RSNA inst. Log Loss	RSNA inst. AUC	RSNA bag Acc.	RSNA bag F1	RSNA bag Log Loss	RSNA bag AUC	CQ500 bag Acc.	CQ500 bag F1	CQ500 bag Log Loss	CQ500 bag AUC
		0.5	G-VGPMIL	**0.9222 ± 0.0129**	0.693 ± 0.0129	0.6666 ± 0.0129	0.867 ± 0.0129	**0.9067 ± 0.0129**	0.899 ± 0.0129	0.6884 ± 0.0129	**0.9113 ± 0.0129**	**0.8254 ± 0.0129**	**0.799 ± 0.0129**	0.6909 ± 0.0129	**0.853 ± 0.0129**
		1.0	G-VGPMIL	0.9222 ± 0.0132	0.695 ± 0.0132	0.6547 ± 0.0132	0.8691 ± 0.0132	0.9053 ± 0.0132	**0.8991 ± 0.0132**	0.6841 ± 0.0132	0.9095 ± 0.0132	0.8153 ± 0.0132	0.7912 ± 0.0132	0.6891 ± 0.0132	0.8444 ± 0.0132
	1.0	2.0	G-VGPMIL	0.9221 ± 0.0134	0.6984 ± 0.0134	0.6349 ± 0.0134	0.8727 ± 0.0134	0.9013 ± 0.0134	0.8969 ± 0.0134	0.6754 ± 0.0134	0.9075 ± 0.0134	0.8052 ± 0.0134	0.7856 ± 0.0134	0.6855 ± 0.0134	0.8385 ± 0.0134
		4.0	G-VGPMIL	0.922 ± 0.0136	**0.7003 ± 0.0136**	**0.6111 ± 0.0136**	**0.8739 ± 0.0136**	0.8893 ± 0.0136	0.8863 ± 0.0136	**0.6645 ± 0.0136**	0.9 ± 0.0136	0.7915 ± 0.0136	0.7759 ± 0.0136	**0.6821 ± 0.0136**	0.8314 ± 0.0136
		-	VGPMIL	0.9219 ± 0.0135	0.7 ± 0.0135	0.616 ± 0.0135	0.8732 ± 0.0135	0.8907 ± 0.0135	0.8874 ± 0.0135	0.6665 ± 0.0135	0.9002 ± 0.0135	0.7919 ± 0.0135	0.7765 ± 0.0135	0.6825 ± 0.0135	0.8329 ± 0.0135
		0.5	G-VGPMIL	0.9263 ± 0.0067	0.7236 ± 0.0067	0.6195 ± 0.0067	0.9362 ± 0.0067	0.8973 ± 0.0067	0.8893 ± 0.0067	0.6595 ± 0.0067	0.9586 ± 0.0067	0.8155 ± 0.0067	0.7849 ± 0.0067	0.6688 ± 0.0067	0.8668 ± 0.0067
		1.0	G-VGPMIL	0.9302 ± 0.004	0.7515 ± 0.004	0.5525 ± 0.004	0.9389 ± 0.004	**0.908 ± 0.004**	**0.9045 ± 0.004**	0.6207 ± 0.004	0.9619 ± 0.004	0.8095 ± 0.004	0.7886 ± 0.004	0.6396 ± 0.004	0.8844 ± 0.004
	6.15	2.0	G-VGPMIL	0.929 ± 0.0045	0.7575 ± 0.0045	0.403 ± 0.0045	0.9403 ± 0.0045	0.888 ± 0.0045	0.8873 ± 0.0045	0.5113 ± 0.0045	0.964 ± 0.0045	0.7985 ± 0.0045	0.7849 ± 0.0045	0.5623 ± 0.0045	0.8919 ± 0.0045
		4.0	G-VGPMIL	0.9313 ± 0.0067	0.7306 ± 0.0067	0.2548 ± 0.0067	0.9015 ± 0.0067	0.9067 ± 0.0067	0.8988 ± 0.0067	0.2924 ± 0.0067	0.9437 ± 0.0067	**0.8489 ± 0.0067**	**0.8205 ± 0.0067**	0.4345 ± 0.0067	**0.905 ± 0.0067**
		-	VGPMIL	**0.9336 ± 0.0076**	**0.7634 ± 0.0076**	**0.1832 ± 0.0076**	**0.9484 ± 0.0076**	0.9067 ± 0.0076	0.9033 ± 0.0076	**0.2465 ± 0.0076**	**0.9649 ± 0.0076**	0.8237 ± 0.0076	0.805 ± 0.0076	**0.4045 ± 0.0076**	0.9025 ± 0.0076
		0.5	G-VGPMIL	0.9235 ± 0.0038	0.7025 ± 0.0038	0.6257 ± 0.0038	0.9379 ± 0.0038	0.8867 ± 0.0038	0.8725 ± 0.0038	0.6623 ± 0.0038	0.9604 ± 0.0038	0.8043 ± 0.0038	0.745 ± 0.0038	0.6703 ± 0.0038	0.8605 ± 0.0038
		1.0	G-VGPMIL	0.9263 ± 0.0048	0.7356 ± 0.0048	0.5593 ± 0.0048	0.9391 ± 0.0048	0.8973 ± 0.0048	0.8903 ± 0.0048	0.6218 ± 0.0048	0.9619 ± 0.0048	0.8103 ± 0.0048	0.7788 ± 0.0048	0.6415 ± 0.0048	0.8677 ± 0.0048
50	11.31	2.0	G-VGPMIL	0.9246 ± 0.0062	0.7556 ± 0.0062	0.4275 ± 0.0062	0.9401 ± 0.0062	0.8747 ± 0.0062	0.8765 ± 0.0062	0.5118 ± 0.0062	0.9633 ± 0.0062	0.7831 ± 0.0062	0.7721 ± 0.0062	0.5664 ± 0.0062	0.8792 ± 0.0062
		4.0	G-VGPMIL	0.9289 ± 0.0056	0.7152 ± 0.0056	0.2464 ± 0.0056	0.9255 ± 0.0056	0.8973 ± 0.0056	0.888 ± 0.0056	0.2857 ± 0.0056	0.9535 ± 0.0056	**0.8522 ± 0.0056**	**0.8193 ± 0.0056**	0.4109 ± 0.0056	**0.9101 ± 0.0056**
		-	VGPMIL	**0.9344 ± 0.0069**	**0.7647 ± 0.0069**	**0.1797 ± 0.0069**	**0.9534 ± 0.0069**	**0.908 ± 0.0069**	**0.9037 ± 0.0069**	**0.246 ± 0.0069**	**0.966 ± 0.0069**	0.8294 ± 0.0069	0.8087 ± 0.0069	**0.3917 ± 0.0069**	0.9038 ± 0.0069
		0.5	G-VGPMIL	0.9108 ± 0.0046	0.604 ± 0.0046	0.6273 ± 0.0046	0.9479 ± 0.0046	0.86 ± 0.0046	0.8297 ± 0.0046	0.6616 ± 0.0046	0.9604 ± 0.0046	0.7963 ± 0.0046	0.7041 ± 0.0046	0.6669 ± 0.0046	0.8824 ± 0.0046
		1.0	G-VGPMIL	0.9196 ± 0.0083	0.6629 ± 0.0083	0.5612 ± 0.0083	0.9501 ± 0.0083	0.8787 ± 0.0083	0.8577 ± 0.0083	0.6194 ± 0.0083	**0.9613 ± 0.0083**	0.8237 ± 0.0083	0.7614 ± 0.0083	0.6339 ± 0.0083	0.8901 ± 0.0083
	69.65	2.0	G-VGPMIL	0.9276 ± 0.0118	0.7184 ± 0.0118	0.4285 ± 0.0118	0.9516 ± 0.0118	0.8907 ± 0.0118	0.8793 ± 0.0118	0.5117 ± 0.0118	0.9608 ± 0.0118	0.8172 ± 0.0118	0.7816 ± 0.0118	0.5523 ± 0.0118	0.8981 ± 0.0118
		4.0	G-VGPMIL	0.911 ± 0.0052	0.6011 ± 0.0052	0.2457 ± 0.0052	0.9576 ± 0.0052	0.8667 ± 0.0052	0.8419 ± 0.0052	0.3187 ± 0.0052	0.9531 ± 0.0052	**0.8429 ± 0.0052**	0.7872 ± 0.0052	**0.3976 ± 0.0052**	**0.9114 ± 0.0052**
		-	VGPMIL	**0.9337 ± 0.0073**	**0.7482 ± 0.0073**	**0.2054 ± 0.0073**	**0.96 ± 0.0073**	**0.8933 ± 0.0073**	**0.8879 ± 0.0073**	**0.3093 ± 0.0073**	0.9582 ± 0.0073	0.8392 ± 0.0073	**0.813 ± 0.0073**	0.3976 ± 0.0073	0.9096 ± 0.0073
		0.5	G-VGPMIL	0.9043 ± 0.0046	0.5559 ± 0.0046	0.6277 ± 0.0046	0.9503 ± 0.0046	0.8547 ± 0.0046	0.821 ± 0.0046	0.661 ± 0.0046	**0.9604 ± 0.0046**	0.7919 ± 0.0046	0.6931 ± 0.0046	0.6661 ± 0.0046	0.8924 ± 0.0046
		1.0	G-VGPMIL	0.9167 ± 0.0064	0.6399 ± 0.0064	0.5628 ± 0.0064	0.9525 ± 0.0064	0.8747 ± 0.0064	0.8522 ± 0.0064	0.6187 ± 0.0064	0.96 ± 0.0064	0.8287 ± 0.0064	0.7683 ± 0.0064	0.6321 ± 0.0064	0.8994 ± 0.0064
	128.0	2.0	G-VGPMIL	**0.9273 ± 0.0097**	**0.71 ± 0.0097**	0.4363 ± 0.0097	0.9537 ± 0.0097	**0.892 ± 0.0097**	**0.8806 ± 0.0097**	0.5199 ± 0.0097	0.9564 ± 0.0097	0.8342 ± 0.0097	0.7961 ± 0.0097	0.5547 ± 0.0097	0.9042 ± 0.0097
		4.0	G-VGPMIL	0.9033 ± 0.0052	0.5474 ± 0.0052	0.2422 ± 0.0052	**0.9591 ± 0.0052**	0.86 ± 0.0052	0.8292 ± 0.0052	0.3517 ± 0.0052	0.9542 ± 0.0052	0.8242 ± 0.0052	0.7471 ± 0.0052	0.4271 ± 0.0052	**0.9081 ± 0.0052**
		-	VGPMIL	0.9266 ± 0.0092	0.7043 ± 0.0092	**0.2327 ± 0.0092**	0.9582 ± 0.0092	0.888 ± 0.0092	0.8784 ± 0.0092	**0.3473 ± 0.0092**	0.955 ± 0.0092	**0.8419 ± 0.0092**	**0.8067 ± 0.0092**	**0.4184 ± 0.0092**	0.9071 ± 0.0092
		0.5	G-VGPMIL	0.9263 ± 0.0066	0.7174 ± 0.0066	0.6575 ± 0.0066	0.8781 ± 0.0066	0.8987 ± 0.0066	0.8918 ± 0.0066	0.6871 ± 0.0066	**0.91 ± 0.0066**	**0.8218 ± 0.0066**	**0.7989 ± 0.0066**	0.6904 ± 0.0066	**0.8518 ± 0.0066**
		1.0	G-VGPMIL	**0.9268 ± 0.0066**	0.7242 ± 0.0066	0.6418 ± 0.0066	0.8816 ± 0.0066	**0.9 ± 0.0066**	**0.8944 ± 0.0066**	0.6829 ± 0.0066	0.9056 ± 0.0066	0.8066 ± 0.0066	0.79 ± 0.0066	0.6885 ± 0.0066	0.8433 ± 0.0066
	1.0	2.0	G-VGPMIL	0.9264 ± 0.0056	0.7282 ± 0.0056	0.6152 ± 0.0056	0.8841 ± 0.0056	0.892 ± 0.0056	0.8876 ± 0.0056	0.6747 ± 0.0056	0.9056 ± 0.0056	0.7906 ± 0.0056	0.7802 ± 0.0056	0.6851 ± 0.0056	0.8353 ± 0.0056
		4.0	G-VGPMIL	0.9265 ± 0.0055	0.7309 ± 0.0055	**0.5819 ± 0.0055**	0.8845 ± 0.0055	0.8907 ± 0.0055	0.8875 ± 0.0055	**0.6624 ± 0.0055**	0.9022 ± 0.0055	0.7846 ± 0.0055	0.7776 ± 0.0055	**0.6803 ± 0.0055**	0.8327 ± 0.0055
		-	VGPMIL	0.9267 ± 0.0054	**0.7316 ± 0.0054**	0.589 ± 0.0054	**0.8846 ± 0.0054**	0.8907 ± 0.0054	0.8875 ± 0.0054	0.6645 ± 0.0054	0.903 ± 0.0054	0.7834 ± 0.0054	0.7765 ± 0.0054	0.6809 ± 0.0054	0.8327 ± 0.0054
		0.5	G-VGPMIL	0.9229 ± 0.0063	0.7168 ± 0.0063	0.6225 ± 0.0063	0.9334 ± 0.0063	0.896 ± 0.0063	0.8893 ± 0.0063	0.6606 ± 0.0063	0.9569 ± 0.0063	0.8039 ± 0.0063	0.774 ± 0.0063	0.6686 ± 0.0063	0.871 ± 0.0063
		1.0	G-VGPMIL	0.9269 ± 0.0071	0.7431 ± 0.0071	0.5553 ± 0.0071	0.9356 ± 0.0071	0.8947 ± 0.0071	0.8898 ± 0.0071	0.6226 ± 0.0071	0.9606 ± 0.0071	0.8072 ± 0.0071	0.783 ± 0.0071	0.6412 ± 0.0071	0.8754 ± 0.0071
	6.15	2.0	G-VGPMIL	0.9281 ± 0.0067	0.7497 ± 0.0067	0.4063 ± 0.0067	0.9382 ± 0.0067	0.884 ± 0.0067	0.8825 ± 0.0067	0.4896 ± 0.0067	0.9636 ± 0.0067	0.808 ± 0.0067	0.7882 ± 0.0067	0.5444 ± 0.0067	0.8832 ± 0.0067
		4.0	G-VGPMIL	0.9315 ± 0.0067	0.736 ± 0.0067	0.252 ± 0.0067	0.9099 ± 0.0067	0.9093 ± 0.0067	0.902 ± 0.0067	0.2821 ± 0.0067	0.9535 ± 0.0067	**0.8379 ± 0.0067**	**0.8114 ± 0.0067**	0.4839 ± 0.0067	**0.9001 ± 0.0067**
		-	VGPMIL	**0.9332 ± 0.0079**	**0.7635 ± 0.0079**	**0.185 ± 0.0079**	**0.9476 ± 0.0079**	**0.9107 ± 0.0079**	**0.9071 ± 0.0079**	**0.2471 ± 0.0079**	**0.9651 ± 0.0079**	0.8238 ± 0.0079	0.8042 ± 0.0079	**0.4083 ± 0.0079**	0.8997 ± 0.0079
		0.5	G-VGPMIL	0.9228 ± 0.0056	0.7094 ± 0.0056	0.6265 ± 0.0056	0.9371 ± 0.0056	0.892 ± 0.0056	0.8809 ± 0.0056	0.6625 ± 0.0056	0.9606 ± 0.0056	0.8038 ± 0.0056	0.7603 ± 0.0056	0.6706 ± 0.0056	0.861 ± 0.0056
		1.0	G-VGPMIL	0.926 ± 0.0053	0.7362 ± 0.0053	0.56 ± 0.0053	0.9394 ± 0.0053	0.9067 ± 0.0053	0.9009 ± 0.0053	0.6239 ± 0.0053	0.9643 ± 0.0053	0.8108 ± 0.0053	0.779 ± 0.0053	0.6425 ± 0.0053	0.8735 ± 0.0053
100	11.31	2.0	G-VGPMIL	0.9292 ± 0.0067	0.7435 ± 0.0067	0.4084 ± 0.0067	0.9419 ± 0.0067	0.9 ± 0.0067	0.8924 ± 0.0067	0.5084 ± 0.0067	0.9626 ± 0.0067	0.8291 ± 0.0067	0.7977 ± 0.0067	0.5529 ± 0.0067	0.8864 ± 0.0067
		4.0	G-VGPMIL	0.9291 ± 0.0056	0.7156 ± 0.0056	0.2465 ± 0.0056	0.9258 ± 0.0056	0.904 ± 0.0056	0.894 ± 0.0056	0.283 ± 0.0056	0.9548 ± 0.0056	**0.8542 ± 0.0056**	**0.8199 ± 0.0056**	0.4148 ± 0.0056	**0.9115 ± 0.0056**
		-	VGPMIL	**0.9346 ± 0.0076**	**0.7659 ± 0.0076**	**0.1797 ± 0.0076**	**0.9537 ± 0.0076**	**0.912 ± 0.0076**	**0.9081 ± 0.0076**	**0.2461 ± 0.0076**	**0.9648 ± 0.0076**	0.8323 ± 0.0076	0.8107 ± 0.0076	**0.3907 ± 0.0076**	0.904 ± 0.0076
		0.5	G-VGPMIL	0.9112 ± 0.0045	0.6059 ± 0.0045	0.6274 ± 0.0045	0.9483 ± 0.0045	0.8587 ± 0.0045	0.8279 ± 0.0045	0.6619 ± 0.0045	0.9614 ± 0.0045	0.7975 ± 0.0045	0.7066 ± 0.0045	0.6672 ± 0.0045	0.8828 ± 0.0045
		1.0	G-VGPMIL	0.9221 ± 0.0049	0.6788 ± 0.0049	0.5618 ± 0.0049	0.9504 ± 0.0049	0.884 ± 0.0049	0.8655 ± 0.0049	0.6189 ± 0.0049	**0.962 ± 0.0049**	0.8274 ± 0.0049	0.7701 ± 0.0049	0.6339 ± 0.0049	0.8905 ± 0.0049
	69.65	2.0	G-VGPMIL	0.9316 ± 0.0058	0.7435 ± 0.0058	0.4308 ± 0.0058	0.9521 ± 0.0058	**0.8987 ± 0.0058**	**0.8907 ± 0.0058**	0.5065 ± 0.0058	0.9614 ± 0.0058	0.8213 ± 0.0058	0.7922 ± 0.0058	0.5497 ± 0.0058	0.8987 ± 0.0058
		4.0	G-VGPMIL	0.9119 ± 0.0037	0.6073 ± 0.0037	0.2407 ± 0.0037	0.9586 ± 0.0037	0.8707 ± 0.0037	0.8477 ± 0.0037	0.3109 ± 0.0037	0.9532 ± 0.0037	**0.8454 ± 0.0037**	0.7929 ± 0.0037	**0.3895 ± 0.0037**	**0.9115 ± 0.0037**
		-	VGPMIL	**0.9337 ± 0.0072**	**0.7481 ± 0.0072**	**0.2054 ± 0.0072**	**0.96 ± 0.0072**	0.8933 ± 0.0072	0.8879 ± 0.0072	**0.3092 ± 0.0072**	0.9582 ± 0.0072	0.8392 ± 0.0072	**0.813 ± 0.0072**	0.3976 ± 0.0072	0.9096 ± 0.0072
		0.5	G-VGPMIL	0.9078 ± 0.004	0.5811 ± 0.004	0.6277 ± 0.004	0.9494 ± 0.004	0.8573 ± 0.004	0.8258 ± 0.004	0.6597 ± 0.004	**0.9605 ± 0.004**	0.7996 ± 0.004	0.7099 ± 0.004	0.6656 ± 0.004	0.8919 ± 0.004
		1.0	G-VGPMIL	0.9179 ± 0.0047	0.6478 ± 0.0047	0.563 ± 0.0047	0.9521 ± 0.0047	0.8733 ± 0.0047	0.8509 ± 0.0047	0.6181 ± 0.0047	0.9604 ± 0.0047	0.8295 ± 0.0047	0.7702 ± 0.0047	0.6319 ± 0.0047	0.8994 ± 0.0047
	128.0	2.0	G-VGPMIL	**0.928 ± 0.0096**	**0.7149 ± 0.0096**	0.4364 ± 0.0096	0.9537 ± 0.0096	**0.8933 ± 0.0096**	**0.883 ± 0.0096**	0.5187 ± 0.0096	0.9567 ± 0.0096	0.8334 ± 0.0096	0.7967 ± 0.0096	0.5542 ± 0.0096	0.9044 ± 0.0096
		4.0	G-VGPMIL	0.9067 ± 0.0045	0.5715 ± 0.0045	0.233 ± 0.0045	**0.9594 ± 0.0045**	0.8693 ± 0.0045	0.8439 ± 0.0045	**0.3358 ± 0.0045**	0.9542 ± 0.0045	0.8323 ± 0.0045	0.7667 ± 0.0045	**0.4126 ± 0.0045**	**0.9076 ± 0.0045**
		-	VGPMIL	0.9265 ± 0.0092	0.7041 ± 0.0092	**0.2327 ± 0.0092**	0.9582 ± 0.0092	0.8867 ± 0.0092	0.8772 ± 0.0092	0.3473 ± 0.0092	0.955 ± 0.0092	**0.8419 ± 0.0092**	**0.8067 ± 0.0092**	0.4184 ± 0.0092	0.9071 ± 0.0092
		0.5	G-VGPMIL	**0.9257 ± 0.0094**	0.7187 ± 0.0094	0.6473 ± 0.0094	0.8841 ± 0.0094	**0.8947 ± 0.0094**	**0.8882 ± 0.0094**	0.6866 ± 0.0094	**0.9122 ± 0.0094**	**0.8118 ± 0.0094**	**0.7899 ± 0.0094**	0.69 ± 0.0094	**0.8469 ± 0.0094**
		1.0	G-VGPMIL	0.9251 ± 0.0087	0.7237 ± 0.0087	0.6273 ± 0.0087	0.8871 ± 0.0087	0.8827 ± 0.0087	0.8782 ± 0.0087	0.6806 ± 0.0087	0.91 ± 0.0087	0.8037 ± 0.0087	0.785 ± 0.0087	0.6876 ± 0.0087	0.8413 ± 0.0087
	1.0	2.0	G-VGPMIL	0.9253 ± 0.008	0.7312 ± 0.008	0.5936 ± 0.008	**0.89 ± 0.008**	0.8787 ± 0.008	0.8753 ± 0.008	0.6701 ± 0.008	0.9067 ± 0.008	0.7886 ± 0.008	0.7747 ± 0.008	0.6838 ± 0.008	0.8345 ± 0.008
		4.0	G-VGPMIL	0.925 ± 0.0074	**0.7336 ± 0.0074**	**0.5496 ± 0.0074**	0.8898 ± 0.0074	0.8827 ± 0.0074	0.8811 ± 0.0074	**0.6546 ± 0.0074**	0.9102 ± 0.0074	0.7769 ± 0.0074	0.7665 ± 0.0074	**0.6798 ± 0.0074**	0.822 ± 0.0074
		-	VGPMIL	0.9248 ± 0.0071	0.7332 ± 0.0071	0.5591 ± 0.0071	0.8898 ± 0.0071	0.88 ± 0.0071	0.8788 ± 0.0071	0.6576 ± 0.0071	0.9101 ± 0.0071	0.7765 ± 0.0071	0.7663 ± 0.0071	0.6803 ± 0.0071	0.8223 ± 0.0071
		0.5	G-VGPMIL	0.9219 ± 0.006	0.7081 ± 0.006	0.624 ± 0.006	0.9332 ± 0.006	0.9053 ± 0.006	0.8979 ± 0.006	0.6608 ± 0.006	0.9597 ± 0.006	0.8128 ± 0.006	0.7728 ± 0.006	0.6695 ± 0.006	0.8631 ± 0.006
		1.0	G-VGPMIL	0.9238 ± 0.0068	0.7297 ± 0.0068	0.5569 ± 0.0068	0.9353 ± 0.0068	0.904 ± 0.0068	0.899 ± 0.0068	0.6208 ± 0.0068	0.9611 ± 0.0068	0.8039 ± 0.0068	0.778 ± 0.0068	0.641 ± 0.00068	0.8729 ± 0.0068
	6.15	2.0	G-VGPMIL	0.9225 ± 0.0053	0.7325 ± 0.0053	0.4018 ± 0.0053	0.9364 ± 0.0053	0.892 ± 0.0053	0.8876 ± 0.0053	0.486 ± 0.0053	0.9618 ± 0.0053	0.7849 ± 0.0053	0.7654 ± 0.0053	0.5435 ± 0.0053	0.8816 ± 0.0053
		4.0	G-VGPMIL	0.9322 ± 0.0056	0.7386 ± 0.0056	0.2497 ± 0.0056	0.9066 ± 0.0056	**0.9107 ± 0.0056**	**0.9029 ± 0.0056**	0.2858 ± 0.0056	0.9514 ± 0.0056	**0.8481 ± 0.0056**	**0.8202 ± 0.0056**	0.4505 ± 0.0056	**0.9042 ± 0.0056**
		-	VGPMIL	**0.9333 ± 0.0076**	**0.7661 ± 0.0076**	**0.1864 ± 0.0076**	**0.9471 ± 0.0076**	0.9053 ± 0.0076	0.9018 ± 0.0076	**0.2483 ± 0.0076**	**0.9646 ± 0.0076**	0.8198 ± 0.0076	0.8009 ± 0.0076	**0.4091 ± 0.0076**	0.898 ± 0.0076
		0.5	G-VGPMIL	0.9225 ± 0.0046	0.7036 ± 0.0046	0.6265 ± 0.0046	0.9376 ± 0.0046	0.8947 ± 0.0046	0.8824 ± 0.0046	0.6622 ± 0.0046	0.9621 ± 0.0046	0.8149 ± 0.0046	0.7643 ± 0.0046	0.6691 ± 0.0046	0.8717 ± 0.0046
		1.0	G-VGPMIL	0.926 ± 0.0052	0.7266 ± 0.0052	0.5593 ± 0.0052	0.9404 ± 0.0052	0.9027 ± 0.0052	0.8942 ± 0.0052	0.6238 ± 0.0052	0.9646 ± 0.0052	0.8291 ± 0.0052	0.7901 ± 0.0052	0.6404 ± 0.0052	0.8787 ± 0.0052
200	11.31	2.0	G-VGPMIL	0.928 ± 0.0036	0.7328 ± 0.0036	0.4011 ± 0.0036	0.9415 ± 0.0036	0.8973 ± 0.0036	0.8902 ± 0.0036	0.4757 ± 0.0036	**0.9657 ± 0.0036**	0.8217 ± 0.0036	0.7859 ± 0.0036	0.5273 ± 0.0036	0.8879 ± 0.0036
		4.0	G-VGPMIL	0.9298 ± 0.0059	0.7204 ± 0.0059	0.2431 ± 0.0059	0.9249 ± 0.0059	0.9053 ± 0.0059	0.8963 ± 0.0059	0.2801 ± 0.0059	0.9553 ± 0.0059	**0.8526 ± 0.0059**	**0.8198 ± 0.0059**	0.4162 ± 0.0059	**0.911 ± 0.0059**
		-	VGPMIL	**0.9346 ± 0.0078**	**0.7662 ± 0.0078**	**0.18 ± 0.0078**	**0.9537 ± 0.0078**	**0.912 ± 0.0078**	**0.9081 ± 0.0078**	**0.246 ± 0.0078**	0.9649 ± 0.0078	0.829 ± 0.0078	0.8074 ± 0.0078	**0.3903 ± 0.0078**	0.9039 ± 0.0078
		0.5	G-VGPMIL	0.9113 ± 0.0044	0.6066 ± 0.0044	0.6274 ± 0.0044	0.9483 ± 0.0044	0.8587 ± 0.0044	0.8279 ± 0.0044	0.6619 ± 0.0044	0.9611 ± 0.0044	0.7971 ± 0.0044	0.7058 ± 0.0044	0.6672 ± 0.0044	0.8828 ± 0.0044
		1.0	G-VGPMIL	0.9217 ± 0.0056	0.6763 ± 0.0056	0.5618 ± 0.0056	0.9504 ± 0.0056	0.8813 ± 0.0056	0.862 ± 0.0056	0.6191 ± 0.0056	**0.9621 ± 0.0056**	0.8274 ± 0.0056	0.7698 ± 0.0056	0.634 ± 0.0056	0.8905 ± 0.0056
	69.65	2.0	G-VGPMIL	0.9284 ± 0.0096	0.7229 ± 0.0096	0.4304 ± 0.0096	0.9515 ± 0.0096	0.8907 ± 0.0096	0.8777 ± 0.0096	0.5139 ± 0.0096	0.9607 ± 0.0096	0.8205 ± 0.0096	0.7834 ± 0.0096	0.5535 ± 0.0096	0.8972 ± 0.0096
		4.0	G-VGPMIL	0.9113 ± 0.0045	0.6031 ± 0.0045	0.2431 ± 0.0045	0.9583 ± 0.0045	0.868 ± 0.0045	0.8436 ± 0.0045	0.3149 ± 0.0045	0.9535 ± 0.0045	**0.8421 ± 0.0045**	0.7879 ± 0.0045	**0.3935 ± 0.0045**	**0.9114 ± 0.0045**
		-	VGPMIL	**0.9337 ± 0.0072**	**0.7481 ± 0.0072**	**0.2054 ± 0.0072**	**0.96 ± 0.0072**	**0.8933 ± 0.0072**	**0.8879 ± 0.0072**	**0.3092 ± 0.0072**	0.9582 ± 0.0072	0.8392 ± 0.0072	**0.813 ± 0.0072**	0.3976 ± 0.0072	0.9096 ± 0.0072
		0.5	G-VGPMIL	0.9057 ± 0.0028	0.5663 ± 0.0028	0.6277 ± 0.0028	0.9503 ± 0.0028	0.856 ± 0.0028	0.8229 ± 0.0028	0.6602 ± 0.0028	**0.9607 ± 0.0028**	0.7968 ± 0.0028	0.7031 ± 0.0028	0.6657 ± 0.0028	0.8926 ± 0.0028
		1.0	G-VGPMIL	0.9167 ± 0.0059	0.6401 ± 0.0059	0.5628 ± 0.0059	0.9522 ± 0.0059	0.8733 ± 0.0059	0.8505 ± 0.0059	0.6185 ± 0.0059	0.9602 ± 0.0059	0.8287 ± 0.0059	0.7683 ± 0.0059	0.632 ± 0.0059	0.8994 ± 0.0059
	128.0	2.0	G-VGPMIL	0.9218 ± 0.0095	0.6756 ± 0.0095	0.4339 ± 0.0095	0.9534 ± 0.0095	0.8827 ± 0.0095	0.8667 ± 0.0095	0.5269 ± 0.0095	0.957 ± 0.0095	0.827 ± 0.0095	0.7817 ± 0.0095	0.5601 ± 0.0095	0.9036 ± 0.0095
		4.0	G-VGPMIL	0.8955 ± 0.0062	0.4885 ± 0.0062	0.2654 ± 0.0062	0.9579 ± 0.0062	0.8347 ± 0.0062	0.7938 ± 0.0062	0.3922 ± 0.0062	0.9529 ± 0.0062	0.7993 ± 0.0062	0.7067 ± 0.0062	0.4687 ± 0.0062	**0.9075 ± 0.0062**
		-	VGPMIL	**0.9265 ± 0.0092**	**0.7041 ± 0.0092**	**0.2327 ± 0.0092**	**0.9582 ± 0.0092**	**0.8867 ± 0.0092**	**0.8772 ± 0.0092**	**0.3473 ± 0.0092**	0.955 ± 0.0092	**0.8419 ± 0.0092**	**0.8067 ± 0.0092**	**0.4184 ± 0.0092**	0.9071 ± 0.0092

Tabla A.7.: Resultados obtenidos en los conjuntos RSNA y CQ500, fijando a 128 el número de características extraídas con AttCNN. Para cada configuración del número de inducing points, lengthscale y k, se señala en negrita el mejor resultado de cada métrica.

Bibliografía

Las referencias se listan por orden alfabético. Aquellas referencias con más de un autor están ordenadas de acuerdo con el primer autor.

[1] Andrews, Stuart, Tsochantaridis, Ioannis, & Hofmann, Thomas. 2002. Support vector machines for multiple-instance learning. *Advances in neural information processing systems*, **15**. [Citado en pág. 7]

[2] Babenko, Boris, Dollár, Piotr, Tu, Zhuowen, & Belongie, Serge. 2008. Simultaneous learning and alignment: Multi-instance and multi-pose learning. *In: Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition*. [Citado en pág. 8]

[3] Bandyopadhyay, Sanghamitra, Ghosh, Dip, Mitra, Ramkrishna, & Zhao, Zhongming. 2015. MBS-TAR: multiple instance learning for predicting specific functional binding sites in microRNA targets. *Scientific reports*, **5**(1), 1–12. [Citado en pág. 8]

[4] Bishop, Christopher M, & Nasrabadi, Nasser M. 2006. *Pattern recognition and machine learning*. Vol. 4. Springer. [Citado en págs. 11 and 19]

[5] Blei, David M, Kucukelbir, Alp, & McAuliffe, Jon D. 2017. Variational inference: A review for statisticians. *Journal of the American statistical Association*, **112**(518), 859–877. [Citado en pág. 19]

[6] Boyd, Stephen, Boyd, Stephen P, & Vandenberghe, Lieven. 2004. *Convex optimization*. Cambridge university press. [Citado en pág. 21]

[7] Carbonneau, Marc-André, Cheplygina, Veronika, Granger, Eric, & Gagnon, Ghyslain. 2018. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition*, **77**, 329–353. [Citado en págs. 8 and 9]

[8] Chilamkurthy, Sasank, Ghosh, Rohit, Tanamala, Swetha, Biviji, Mustafa, Campeau, Norbert G, Venugopal, Vasantha Kumar, Mahajan, Vidur, Rao, Pooja, & Warier, Prashant. 2018. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *The Lancet*, **392**(10162), 2388–2396. [Citado en págs. 1, 47, and 54]

[9] Damianou, Andreas. 2015. *Deep Gaussian processes and variational propagation of uncertainty*. Ph.D. thesis, University of Sheffield. [Citado en págs. 11, 16, and 19]

[10] Dietterich, Thomas G, Lathrop, Richard H, & Lozano-Pérez, Tomás. 1997. Solving the multiple instance problem with axis-parallel rectangles. *Artificial intelligence*, **89**(1-2), 31–71. [Citado en págs. 5 and 7]

[11] Doi, Kunio. 2007. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, **31**(4-5), 198–211. [Citado en pág. 8]

[12] Durrett, Rick. 2019. *Probability: theory and examples*. Vol. 49. Cambridge University Press. [Citado en pág. 19]

[13] Eksi, Ridvan, Li, Hong-Dong, Menon, Rajasree, Wen, Yuchen, Omenn, Gilbert S., Kretzler, Matthias, & Guan, Yuanfang. 2013. Systematically Differentiating Functions for Alternatively Spliced Isoforms through Integrating RNA-seq Data. *PLOS Computational Biology*, **9**(11), 1–16. [Citado en pág. 8]

[14] Flanders, Adam E, Prevedello, Luciano M, Shih, George, Halabi, Safwan S, Kalpathy-Cramer, Jayashree, Ball, Robyn, Mongan, John T, Stein, Anouk, Kitamura, Felipe C, Lungren, Matthew P, *et al.* . 2020. Construction of a machine learning dataset through collaboration: the RSNA 2019 brain CT hemorrhage challenge. *Radiology: Artificial Intelligence*, **2**(3), e190211. [Citado en pág. 46]

[15] Garcia-Garcia, Alberto, Orts-Escolano, Sergio, Oprea, Sergiu, Villena-Martinez, Victor, & Garcia-Rodriguez, Jose. 2017. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*. [Citado en pág. 8]

[16] Gebel, James M, & Broderick, Joseph P. 2000. Intracerebral hemorrhage. *Neurologic clinics*, **18**(2), 419–438. [Citado en pág. 1]

[17] Gibbs, Mark N, & MacKay, David JC. 2000. Variational Gaussian process classifiers. *IEEE Transactions on Neural Networks*, **11**(6), 1458–1464. [Citado en pág. 15]

[18] Goodfellow, Ian, Bengio, Yoshua, & Courville, Aaron. 2016. *Deep learning*. MIT press. [Citado en pág. 1]

[19] Haußmann, Manuel, Hamprecht, Fred A, & Kandemir, Melih. 2017. Variational bayesian multiple instance learning with gaussian processes. *Pages 6570–6579 of: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [Citado en págs. 2, 3, 8, 23, 24, and 28]

[20] Ilse, Maximilian, Tomczak, Jakub, & Welling, Max. 2018. Attention-based deep multiple instance learning. *Pages 2127–2136 of: International conference on machine learning*. PMLR. [Citado en pág. 7]

[21] Jaakkola, Tommi S, & Haussler, David. 1999. Probabilistic kernel regression models. *In: Seventh International Workshop on Artificial Intelligence and Statistics*. PMLR. [Citado en pág. 15]

[22] Jaakkola, Tommi S, & Jordan, Michael I. 2000. Bayesian parameter estimation via variational methods. *Statistics and Computing*, **10**(1), 25–37. [Citado en págs. 2, 3, and 27]

[23] Kim, Minyoung, & De la Torre, Fernando. 2010. Gaussian processes multiple instance learning. *In: ICML*. [Citado en págs. 7, 8, and 25]

[24] Liu, Ying, Zhang, Dengsheng, Lu, Guojun, & Ma, Wei-Ying. 2007. A survey of content-based image retrieval with high-level semantics. *Pattern recognition*, **40**(1), 262–282. [Citado en pág. 8]

[25] Mangasarian, Olvi L, & Wild, Edward W. 2008. Multiple instance classification via successive linear programming. *Journal of optimization theory and applications*, **137**(3), 555–568. [Citado en pág. 7]

[26] Maron, Oded, & Lozano-Pérez, Tomás. 1997. A framework for multiple-instance learning. *Advances in neural information processing systems*, **10**. [Citado en pág. 7]

[27] Minka, Thomas Peter. 2001. *A family of algorithms for approximate Bayesian inference*. Ph.D. thesis, Massachusetts Institute of Technology. [Citado en pág. 15]

[28] Monteiro, Miguel, Newcombe, Virginia FJ, Mathieu, Francois, Adatia, Krishma, Kamnitsas, Konstantinos, Ferrante, Enzo, Das, Tilak, Whitehouse, Daniel, Rueckert, Daniel, Menon, David K, *et al.* . 2020. Multiclass semantic segmentation and quantification of traumatic brain injury lesions on head CT using deep learning: an algorithm development and multicentre validation study. *The Lancet Digital Health*, **2**(6), e314–e322. [Citado en pág. 54]

[29] Nguyen, Nhan T, Tran, Dat Q, Nguyen, Nghia T, & Nguyen, Ha Q. 2020. A CNN-LSTM architecture for detection of intracranial hemorrhage on CT scans. *medRxiv*. [Citado en pág. 54]

[30] Parisi, Giorgio, & Shankar, Ramamurti. 1988. Statistical field theory. [Citado en pág. 20]

[31] Parizel, P, Makkat, S, Van Miert, E, Van Goethem, J, Van den Hauwe, L, & De Schepper, A. 2001. Intracranial hemorrhage: principles of CT and MRI interpretation. *European radiology*, **11**(9), 1770–1783. [Citado en pág. 1]

[32] Polson, Nicholas G, Scott, James G, & Windle, Jesse. 2013. Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American statistical Association*, **108**(504), 1339–1349. [Citado en pág. 31]

[33] Quinonero-Candela, Joaquin, & Rasmussen, Carl Edward. 2005. A unifying view of sparse approximate Gaussian process regression. *The Journal of Machine Learning Research*, **6**, 1939–1959. [Citado en pág. 17]

[34] Rasmussen, Carl Edward. 2003. Gaussian processes in machine learning. *Pages 63–71 of: Summer school on machine learning*. Springer. [Citado en págs. 2, 3, 7, and 11]

[35] Rodríguez-Yáñez, M, Castellanos, M, Freijo, MM, Fernández, JC López, Martí-Fàbregas, J, Nombela, F, Simal, P, Castillo, J, Díez-Tejedor, E, Fuentes, B, *et al.* . 2013. Guías de actuación clínica en la hemorragia intracerebral. *Neurología*, **28**(4), 236–249. [Citado en pág. 1]

[36] Schmidt, Arne, Silva-Rodríguez, Julio, Molina, Rafael, & Naranjo, Valery. 2022. Efficient Cancer Classification by Coupling Semi Supervised and Multiple Instance Learning. *IEEE Access*, **10**, 9763–9773. [Citado en págs. 9 and 15]

[37] Seeger, Matthias. 1999. Bayesian model selection for support vector machines, Gaussian processes and other kernel classifiers. *Advances in neural information processing systems*, **12**. [Citado en pág. 15]

[38] Seeger, Matthias W, Williams, Christopher KI, & Lawrence, Neil D. 2003. Fast forward selection to speed up sparse Gaussian process regression. *Pages 254–261 of: International Workshop on Artificial Intelligence and Statistics*. PMLR. [Citado en pág. 17]

[39] Snelson, Edward, & Ghahramani, Zoubin. 2005. Sparse Gaussian processes using pseudo-inputs. *Advances in neural information processing systems*, **18**. [Citado en págs. 17 and 24]

[40] Sobrino, Justin, & Shafi, Shahid. 2013. Timing and causes of death after injuries. *Pages 120–123 of: Baylor University Medical Center Proceedings*, vol. 26. Taylor & Francis. [Citado en pág. 1]

[41] Strub, WM, Leach, JL, Tomsick, T, & Vagal, A. 2007. Overnight preliminary head CT interpretations provided by residents: locations of misidentified intracranial hemorrhage. *American journal of neuroradiology*, **28**(9), 1679–1682. [Citado en pág. 1]

[42] Titsias, Michalis. 2009. Variational learning of inducing variables in sparse Gaussian processes. *Pages 567–574 of: Artificial intelligence and statistics*. PMLR. [Citado en págs. 17, 18, and 19]

[43] Tompson, Jonathan, Goroshin, Ross, Jain, Arjun, LeCun, Yann, & Bregler, Christoph. 2015. Efficient object localization using convolutional networks. *Pages 648–656 of: Proceedings of the IEEE conference on computer vision and pattern recognition*. [Citado en pág. 8]

[44] Tong, Tong, Wolz, Robin, Gao, Qinquan, Guerrero, Ricardo, Hajnal, Joseph V, Rueckert, Daniel, Initiative, Alzheimer's Disease Neuroimaging, *et al.* . 2014. Multiple instance learning for classification of dementia in brain MRI. *Medical image analysis*, **18**(5), 808–818. [Citado en pág. 9]

[45] Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N, Kaiser, Łukasz, & Polosukhin, Illia. 2017. Attention is all you need. *Advances in neural information processing systems*, **30**. [Citado en pág. 7]

[46] Vezhnevets, Alexander, & Buhmann, Joachim M. 2010. Towards weakly supervised semantic segmentation by means of multiple instance and multitask learning. *Pages 3249–3256 of: 2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE. [Citado en pág. 8]

[47] Wang, Fulton, & Pinar, Ali. 2021. The Multiple Instance Learning Gaussian Process Probit Model. *Pages 3034–3042 of: International Conference on Artificial Intelligence and Statistics*. PMLR. [Citado en pág. 8]

[48] Wang, Xinggang, Yan, Yongluan, Tang, Peng, Bai, Xiang, & Liu, Wenyu. 2018. Revisiting multiple instance neural networks. *Pattern Recognition*, **74**, 15–24. [Citado en pág. 7]

[49] Wei, Xiu-Shen, & Zhou, Zhi-Hua. 2016. An empirical study on image bag generators for multi-instance learning. *Machine learning*, **105**(2), 155–198. [Citado en pág. 8]

[50] Weidmann, Nils, Frank, Eibe, & Pfahringer, Bernhard. 2003. A two-level learning method for generalized multi-instance problems. *Pages 468–479 of: European Conference on Machine Learning*. Springer. [Citado en págs. 1, 2, 5, and 6]

[51] Wenzel, Florian, Galy-Fajou, Théo, Donner, Christan, Kloft, Marius, & Opper, Manfred. 2019. Efficient Gaussian process classification using Pòlya-Gamma data augmentation. *Pages 5417–5424 of: Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33. [Citado en págs. 2, 3, 23, 30, and 58]

[52] Williams, Christopher KI, & Barber, David. 1998. Bayesian classification with Gaussian processes. *IEEE Transactions on pattern analysis and machine intelligence*, **20**(12), 1342–1351. [Citado en pág. 15]

[53] Wu, Yunan, Schmidt, Arne, Hernández-Sánchez, Enrique, Molina, Rafael, & Katsaggelos, Aggelos K. 2021. Combining attention-based multiple instance learning and gaussian processes for CT hemorrhage detection. *Pages 582–591 of: International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. [Citado en págs. 2, 3, 49, 50, and 52]

[54] Yan, Yongluan, Wang, Xinggang, Guo, Xiaojie, Fang, Jiemin, Liu, Wenyu, & Huang, Junzhou. 2018. Deep multi-instance learning with dynamic pooling. *Pages 662–677 of: Asian Conference on Machine Learning*. PMLR. [Citado en pág. 7]

[55] Yang, Jian, Zhang, David, Frangi, Alejandro F, & Yang, Jing-yu. 2004. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE transactions on pattern analysis and machine intelligence*, **26**(1), 131–137. [Citado en pág. 42]

[56] Yasser, EL-Manzalawy, Dobbs, Drena, & Honavar, Vasant. 2010. Predicting MHC-II binding affinity using multiple instance regression. *IEEE/ACM Transactions on computational biology and bioinformatics*, **8**(4), 1067–1079. [Citado en pág. 8]

[57] Zhang, Dingwen, Han, Junwei, Cheng, Gong, & Yang, Ming-Hsuan. 2021. Weakly supervised object localization and detection: A survey. *IEEE transactions on pattern analysis and machine intelligence*. [Citado en pág. 8]

[58] Zhang, Qi, Goldman, Sally A, Yu, Wei, & Fritts, Jason E. 2002. Content-based image retrieval using multiple-instance learning. *Page 2 of: ICML*, vol. 1. Citeseer. [Citado en págs. 7 and 8]