



Propuesta de Trabajo Fin de Grado en Matemáticas (curso 2022-2023)

Responsable de tutorización: Rafael Molina Soriano
Departamento: Ciencias de la Computación e Inteligencia Artificial
Correo electrónico: rms@decsai.ugr.es

Responsable de cotutorización:
Departamento:
Correo electrónico:

(Rellenar sólo en caso de que la propuesta esté realizada a través de un estudiante)
Estudiante que propone el trabajo: Marian Rojas Arques

Título del trabajo: Clasificación binaria con Procesos Gaussianos usando diferentes modelos de observación e inferencia variacional. Aplicación a imágenes médicas.

Tipología del trabajo (marcar una o varias de las siguientes casillas):

- Complementario de profundización
- Divulgación de las Matemáticas
- Docencia e innovación
- Herramientas informáticas
- Iniciación a la investigación

Materias del grado relacionadas con el trabajo: Informática I y II, Introducción a la Probabilidad, Inferencia estadística, Procesos estocásticos, Estadística Computacional, Análisis funcional.

Descripción y resumen de contenidos:

El objetivo del presente trabajo fin de grado (tfg) es doble. Por un lado, se profundizará en los conocimientos adquiridos en el grado y, por otro, la estudiante se iniciará en un tema de investigación.

Los Procesos Gaussianos son modelos probabilísticos no paramétricos muy populares en tareas de clasificación. Gracias a su formulación probabilística pueden cuantificar la incertidumbre durante el modelado y la predicción. Cuando un Proceso Gaussiano a priori se combina con una verosimilitud Gaussiana (como es el caso de regresión), la inferencia puede hacerse con una expresión de forma cerrada. Sin embargo, en clasificación binaria, la usada distribución de Bernoulli hace que el cálculo de la distribución a posteriori de la función no paramétrica subyacente no sea tratable al no ser el modelo de observación conjugado de la distribución Gaussiana. La inferencia variacional es una forma de inferencia aproximada muy extendida para abordar la estimación de la distribución a posteriori. En ella se propone una factorización para dicha distribución y mediante la optimización de un funcional se estiman los parámetros variacionales óptimos.

La distribución de Pòlya-Gamma permite realizar inferencia bayesiana eficiente para problemas de clasificación binaria con regresión logística, y también cuando se usan Procesos Gaussianos como distribuciones a priori. El uso de esta distribución lleva a la introducción de variables latentes

adicionales que restauran la conjugación. De esta forma, se puede encontrar expresiones en forma cerrada para la distribución a posteriori del Proceso Gaussiano.

En este tfg se estudiarán

- Procesos Gaussianos como distribuciones a priori en problemas de clasificación,
- Modelos de observación en dichos problemas, en concreto, el uso de la distribución de Bernoulli y su obtención a partir de la distribución Polya-Gamma,
- Inferencia variacional para los dos modelos de observación.

y se aplicarán los conocimientos obtenidos a la clasificación de imágenes histológicas de cáncer.

Actividades a desarrollar:

Para la realización del presente tfg se requiere:

- *Estudio de los Procesos Gaussianos para clasificación binaria.*
- *Repaso de la inferencia bayesiana y revisión bibliográfica.*
- *Estudio de la inferencia variacional y su aplicación a los Procesos Gaussianos.*
- *Estudio y uso de la distribución de Pòlya-Gamma y su combinación con Procesos Gaussianos en problemas de clasificación.*
- *Aplicación de los modelos expuestos a la clasificación binaria de imágenes médicas.*

Objetivos matemáticos planteados

Desarrollo de la teoría de Procesos Gaussianos para clasificación binaria.

Desarrollo de la teoría de inferencia usando la distribución de Pòlya-Gamma.

Evaluación de los modelos teóricos en un problema de clasificación automática de imágenes médicas.

Bibliografía para el desarrollo matemático de la propuesta:

1. Bishop, C. M. (2006). Pattern Recognition and Machine Learning.
2. A. E. Esteban, M. López-Pérez, A. Colomer, M. A. Sales, R. Molina, and V. Naranjo, “A New Optical Density Granulometry-Based Descriptor for the Classification of Prostate Histological Images Using Shallow and Deep Gaussian Processes”, Computer Methods and Programs in Biomedicine, vol.178, 303-317, September 2019.
3. Hensman, J., Fusi, N., & Lawrence, N. D. (2013). Gaussian Processes for Big Data. In Uncertainty in Artificial Intelligence (p. 282).
4. Murphy, K. P. (2022). Probabilistic machine learning: an introduction. MIT press.
5. Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. Journal of the American Statistical Association, 108(504), 1339-1349.
6. Wenzel, F., Galy-Fajou, T., Donner, C., Kloft, M., & Opper, M. (2019, July). Efficient Gaussian process classification using Pòlya-Gamma data augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 01, pp. 5417-5424).

Otras referencias (si procede):

A handwritten signature in black ink, appearing to read 'María', with a large, sweeping flourish extending upwards and to the right.

Firma del estudiante
(solo para trabajos propuestos por estudiantes)

Firma del responsable de tutorización
(solo para trabajos propuestos por estudiantes)

Firma del responsable de cotutorización
(solo para trabajos propuestos por estudiantes)

En, Granada, a 25 de abril de 2022