

Tarea 3

Profesor: Felipe Tobar

Auxiliares: N. Aramayo, L. Araya, M. Campos, A. Cuevas y C. Valenzuela

Consultas: Todos el cuerpo docente.

Fecha entrega: 27/4/2018

Formato entrega: Entregue un informe en formato PDF con una extensión de a lo más 2 páginas presentando y analizando sus resultados, detalle la metodología utilizada y adicionalmente debe entregar un jupyter notebook con los códigos que creó para resolver la tarea.

P1. Regresión logística, clasificación y Metropolis-Hastings

Cargue los datos desde el archivo **datosT3.txt**, el cual consiste en un conjunto de 500 muestras - etiquetas de la forma $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, donde $\mathbf{x}_i \in \mathbb{R}^2, y_i \in \{0, 1\} \forall i = 1, \dots, n$. El objetivo de este inciso es implementar un clasificador lineal utilizando como modelo una regresión logística, y tanto máxima verosimilitud como Markov chain Monte Carlo para estimar sus parámetros.

- a) (2 pts.) Muestre en 2D con un color las muestras de las clase 0 y otro distintos las muestras de la clase 1 en base a las etiquetas disponibles. Luego, implemente un modelo de regresión logística para clasificar los datos disponibles, para esto construya la función de log-verosimilitud y estime los parámetros de la regresión logística utilizando L-BFGS. Sobre las muestras presentadas, grafique la recta que separa ambas clases. ¿Pudo clasificar correctamente todas las muestras?
- c) (2 pts.) Implemente el algoritmo de Metropolis-Hastings (MH) para estimar la densidad *a posteriori* del modelo encontrado en a), donde la cadena de Markov puede inicializarse con los parámetros encontrados en esa parte. Genere del orden de 1000 muestras para los parámetros y grafique algunas de las rectas que corresponden a estos parámetros. Discuta su elección del prior sobre los parámetros y muestre cuán sensible son sus soluciones para distintos priors.
- d) (1 pts.) Con sus muestras generadas, aproxime la distribución posterior de cada parámetro usando histogramas y muestrelos ¿Cómo se compara la moda de cada parámetro con lo obtenido con máximo a posteriori? discuta.

P2. Proyecto curso (2 pts.)

Describa en detalle su proyecto y presente un análisis exploratorio de sus datos. Para esto puede utilizar cualquier herramienta de programación científica que desee.