

# Parcial 1: Teoría de Aprendizaje de Máquina - 2025-I

Profesor: Andrés Marino Álvarez Meza, Ph.D.  
Departamento de Ingeniería Eléctrica, Electrónica, y Computación  
Universidad Nacional de Colombia - sede Manizales

## 1. Instrucciones

- Para recibir crédito total por sus respuestas, estas deben estar claramente justificadas e ilustrar sus procedimientos y razonamientos (paso a paso) de forma concreta, clara y completa.
- El parcial debe ser enviado al correo electrónico `amalvarezme@unal.edu.co` antes de las 23:59 del 20 de mayo de 2025, vía link de GitHub, con componentes teóricas de solución a mano en formato pdf. Los componentes de simulación deben presentarse en cuadernos de Python y en un Dashboard según se describe en las preguntas. Si el correo unal o GitHub presentan inconsistencias, enviar los archivos como adjunto en `.zip`.
- Los códigos deben estar debidamente comentados en las celdas correspondientes y explicados en celdas de texto (markdown). Los códigos que no incluyan comentarios ni discusiones no serán considerados en la evaluación final.
- Incluir en el asunto del correo de envío del parcial: Parcial 1 TAM 2025-1: Nombre completo.

## 2. Preguntas

1. Sea el modelo de regresión  $t_n = \phi(\mathbf{x}_n)\mathbf{w}^\top + \eta_n$ , con  $\{t_n \in \mathbb{R}, \mathbf{x}_n \in \mathbb{R}^P\}_{n=1}^N$ ,  $\mathbf{w} \in \mathbb{R}^Q$ ,  $\phi : \mathbb{R}^P \rightarrow \mathbb{R}^Q$ ,  $Q \geq P$ , y  $\eta_n \sim \mathcal{N}(\eta_n|0, \sigma_\eta^2)$ . Presente el problema de optimización y la solución del mismo para los modelos mínimos cuadrados, mínimos cuadrados regularizados, máxima verosimilitud, máximo a-posteriori, Bayesiano con modelo lineal Gaussiano, regresión rígida kernel y mediante procesos Gaussianos. Asuma datos i.i.d. Discuta las diferencias y similitudes entre los modelos estudiados.
2. Utilizando un esquema de validación cruzada de 5 folds, compare el rendimiento de los siguientes regresores de sklearn sobre el conjunto de datos Ames Housing Dataset: <https://www.kaggle.com/datasets/shashanknecrothapa/ames-housing-dataset>:
  - LinearRegresor
  - Lasso
  - ElasticNet
  - KernelRidge
  - SGDRegressor
  - BayesianRidge
  - Gaussian Process Regressor
  - RandomForestRegressor
  - Support Vector Machines Regressor

Utilice GridSearch, RandomSearch y optimización Bayesiana con muestreo tipo Gaussian Processes para encontrar los hiper-parámetros relevantes de cada modelo. Justifique los hiper-parámetros a buscar, la rejilla o cotas de valores escogida para cada algoritmo según los modelos estudiados y el score a minimizar. Finalmente, presente los rendimientos promedios en los datos de evaluación con su respectiva desviación estándar para las siguientes medidas de desempeño: MAE, MSE, R2 (consultar) y MAPE (consultar). Nota: Recuerde incluir y justificar el preprocesamiento utilizado e incluya un análisis exploratorio de los datos.

3. Construya un Dashboard a partir del paquete Streamlit: <https://streamlit.io/> que permita visualizar de forma intuitiva el dataset y comparar el rendimiento de los tres mejores regresores del punto 2. Ver material de apoyo para construcción de Dashboards:

Dashboards con Dash y Streamlit: <https://github.com/crosorior/Dashboards>

Guía Básica Streamlit con Colab: <https://github.com/amalvarezme/SenalesSistemas/tree/master/Dashboards>