

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

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

## 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 14 de octubre 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 se describe en las preguntas.
- 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.
- El parcial puede trabajarse en grupos de hasta tres personas.
- Incluir en el asunto del correo de envío del parcial: Parcial 1 TAM 2025-2: Nombres de los integrantes.

## 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. Presente un cuadro comparativo de los siguientes regresores, en donde discuta: modelo matemático, función de costo, estrategia de optimización, relación con los esquemas básicos de regresión discutidos en el punto 1, y escalabilidad:
  - LinearRegressor
  - Lasso
  - ElasticNet
  - KernelRidge
  - SGDRegressor
  - BayesianRidge
  - Gaussian Process Regressor
  - Support Vector Machines Regressor
  - RandomForestRegressor
  - GradientBoostingRegressor y XGBoost
3. Consulte en qué consiste y el funcionamiento principal de la librería RAPIDS. Posteriormente, elabore una tabla que compare los métodos e hiperparámetros más relevantes de los regresores mencionados en el punto 2, indicando su implementación o el algoritmo equivalente disponible en RAPIDS.

4. Presente un análisis detallado del conjunto de datos disponible en <https://www.kaggle.com/competitions/nfl-big-data-bowl-2026-prediction/overview>. Incluya la descripción de las variables de entrada y salida, el objetivo principal del concurso, un análisis exploratorio básico, los procedimientos de codificación de valores faltantes y variables categóricas, así como las estrategias de ingeniería de características —si aplican—. Finalmente, proporcione todos los detalles necesarios para comprender, interpretar y analizar adecuadamente el conjunto de datos.

Luego, utilizando un esquema de validación hold-out con una partición del 60 % para entrenamiento, 20 % para validación y 20 % para evaluación del rendimiento, compare los regresores mencionados en el punto 2 —considerando sus posibles implementaciones en RAPIDS, según lo indicado en el punto 3—, empleando el conjunto de datos de entrenamiento disponible en el dataset. Realice las consideraciones necesarias para la implementación y sintonización de cada modelo mediante optimización bayesiana, con el fin de determinar los hiperparámetros más relevantes de cada algoritmo. Justifique la selección de los hiperparámetros a optimizar, así como la rejilla o los rangos de valores definidos para cada modelo, de acuerdo con las características estudiadas y el score a minimizar. Finalmente, presente los rendimientos promedio obtenidos sobre el conjunto de evaluación, junto con su respectiva desviación estándar, empleando las siguientes métricas de desempeño: MAE, MSE, R2 (consultar) y MAPE (consultar).