

Parcial 1: Teoría de Aprendizaje de Máquina 2024-II

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

1. Instrucciones

- Para recibir el crédito total, sus respuestas deben estar justificadas de manera clara, detallada y concreta, mostrando los procedimientos y razonamientos paso a paso.
- Está permitido el uso de herramientas de inteligencia artificial (IA). Si las utiliza, por favor, declare explícitamente cómo fueron empleadas en la resolución de cada pregunta. Incluya los prompts (consultas) y las iteraciones realizadas con las IA durante el desarrollo del parcial.
- La entrega del parcial debe realizarse antes de las 23:59 del 14 de enero de 2025 al email `amalvarezme@unal.edu.co` mediante un enlace de GitHub.
 - 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.

2. Preguntas

- 2.1 (Valor 2.5 puntos). 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 (inferencia) 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.
- 2.2 (Valor 2.5 puntos) Utilizando un esquema de validación cruzada de 5 folds, compare el rendimiento de los siguientes regresores sobre el conjunto de datos NBA scored prediction de Kaggle:
- LinearRegresor
 - Lasso
 - ElasticNet
 - KernelRidge
 - SGDRegressor
 - BayesianRidge
 - Gaussian Process Regressor
 - RandomForestRegressor
 - Support Vector Machines Regressor

fijando el score del gridsearch en términos del error absoluto medio y el error cuadrático medio (MAE y MSE). Justifique los hiperparámetros a buscar y la rejilla de valores escogida para cada algoritmo según los modelos estudiados en clase. 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).