

Parcial 2: Teoría de Aprendizaje de Máquina

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.
- La componente teórica de cada uno de los puntos deberá entregarse a mano. La componente práctica (programación), debe ser enviada al correo electrónico `amalvarezme@unal.edu.co` antes de las 11:00 pm del 26 de octubre de 2023, vía link de GitHub o adjuntando el archivo `.ipynb`. Se debe enviar un solo cuaderno de Python.
- Los códigos deben estar debidamente comentados en las celdas de código, y discutidos/explicados en celdas de texto (markdown). Códigos no comentados ni discutidos, no serán contabilizados en la nota final.
- El presente parcial puede ser desarrollado en parejas (enviar un solo cuaderno con los nombres de los integrantes).

2. Preguntas

- 2.1 (Valor 2.5 puntos). Consulte y presente el modelo y problema de optimización de los siguientes clasificadores (Ver 4'Clasificación):
- Naive' Bayes.GaussianNB
 - SGDClassifier
 - LogisticRegression
 - LinearDiscriminantAnalysis
 - KNeighborsClassifier
 - SVC
 - RandomForestClassifier
 - GaussianProcessClassifier
- 2.2 (Valor 2.5 puntos). Utilizando un esquema de validación cruzada de 5 folds sobre el conjunto de entrenamiento, compare el rendimiento de los métodos del punto 1 sobre la base de datos LFW (ver 6'CVClasificacion'LFW.ipynb) en términos de validación cruzada con GridsearchCV, fijando el score del gridsearch en términos del acc y el f1. Justifique los hiperparámetros a buscar y la rejilla de valores escogida para cada algoritmo según los modelos estudiados en clase y las respuestas del punto 1.