

Parcial 2: Teoría de Aprendizaje de Máquina - 2024 - 2

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 3 de Febrero 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 Consulte y presente el modelo y problema de optimización de los siguientes clasificadores (Ver ver cuadernos guía de clasificadores):
 - Naive`Bayes.GaussianNB
 - SGDClassifier
 - LogisticRegression
 - LinearDiscriminantAnalysis
 - KNeighborsClassifier
 - SVC
 - RandomForestClassifier
 - GuassianProcessClassifier
- 2.2 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.
- 2.3 Presente una comparación visual en 2D, utilizando PCA y UMAP (ver Cuaderno guía reducción de dimensión), en donde se muestre el espacio proyectado, colocando en color el valor de la etiqueta original para cada muestra. Para el caso de PCA, grafique los primeros 2 componentes (vectores propios) en formato imagen de entrada.