

# Parcial 2: Teoría de Aprendizaje de Máquina - 2025 - 1

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 17 de julio 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. El Dashboard debe poder ejecutarse de forma intuitiva y agregar el link del vídeo en youtube según se describe en el punto e.
- 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 resolverse en grupos de hasta 3 personas.
- Incluir en el asunto del correo de envío del parcial: Parcial 2 TAM 2025-1: Nombres completos.

## 2. Preguntas

- a) Consulte y presente el modelo y problema de optimización de los siguientes modelos de aprendizaje de máquina:
- PCA
  - UMAP
  - Naive`Bayes.GaussianNB
  - SGDClassifier
  - LogisticRegression
  - LinearDiscriminantAnalysis
  - KNeighborsClassifier
  - SVC
  - RandomForestClassifier
  - GuassianProcessClassifier
  - Clasificadores basados en Deep Learning
- b) Utilizando la base de datos USPS, proyecte el conjunto de datos en un espacio de menor dimensión empleando los métodos PCA y UMAP. En cada caso, represente gráficamente el espacio proyectado, utilizando colores para indicar

la etiqueta de cada muestra. Adicionalmente, incluya algunas imágenes representativas superpuestas sobre el espacio proyectado. Analice las diferencias entre los espacios obtenidos y explore cómo varía la representación al modificar el número de vecinos en UMAP, discutiendo el impacto de este parámetro en la estructura del espacio latente.

- c) A partir de los resultados de proyección obtenidos, seleccione tres métodos de clasificación descritos en el punto (a), para clasificar los datos en la base de datos USPS. Incluya al menos un modelo basado en aprendizaje profundo. Justifique la elección de hiperparámetros, las estrategias de entrenamiento empleadas y cualquier otro aspecto relevante del proceso de modelado. Presente los reportes de desempeño de los clasificadores, así como las curvas ROC correspondientes.
- d) Construya un dashboard interactivo que ilustre los principales conceptos y resultados obtenidos en los puntos a al c.
- e) Presente un video de entre 3 y 5 minutos en el que discuta el dashboard implementado y sintetice los conceptos y resultados más relevantes abordados en los puntos (a) al (c). Al final del video, incluya una descripción e ilustración de las capas de atención y los modelos tipo Transformer, explicando cómo podrían aplicarse para resolver el problema de clasificación estudiado.