

# Taller 6 PR: Clasificación No Lineal

Facultad de Ingeniería  
Departamento de Electrónica

**Nota:** fecha máxima de entrega del informe **viernes 7 de octubre de 2022 a las 3:55 p.m.**  
Por cada minuto de retraso en la entrega se descontará una (1) décima.

## Objetivos:

- Visualizar e interpretar datos de acuerdo con la estadística descriptiva de los mismos.
- Comparar el desempeño de diferentes estrategias de clasificación no lineal en problemas de clasificación multiclase y binaria.
- Utilizar el método de validación simple y cruzada para garantizar que el rendimiento de los clasificadores es independiente de la partición entre los datos de entrenamiento y prueba.

1. (2.5 pt.) El conjunto de datos “*Classify Gestures by Reading Muscle Activity*” (ver data.rar, carpeta “Punto 1”) del repositorio de bases de datos *Kaggle* (<https://www.kaggle.com/kyr7plus/emg-4>) contiene datos pertenecientes a 4 clases  $\{rock-0, scissors-1, paper-2, ok-3\}$ . Cada observación del conjunto de datos está descrita por 8 lecturas consecutivas de 8 sensores de EMG, para un total de 64 atributos de EMG. Los datos se registraron a 200 Hz, lo que significa que cada observación tiene  $8 \cdot (1/200) = 40$  ms de tiempo de grabación.

Con base en lo anterior,

- a.) (0.1 pt.) Divida los datos aleatoriamente en conjuntos de entrenamiento (60 %), validación (20 %) y prueba (20 %), con el fin de realizar *hold-out cross-validation*\*.

\*Sugerencia: procure que las observaciones sean homogéneas para cada conjunto.

- b.) (1.0 pt.) Utilizando los conjuntos de entrenamiento y validación, y un SVM con *kernel* polinomial con transformación binaria OVA (*one vs all*), diseñe una estrategia para determinar el orden del *kernel* que mejor clasifica los datos. Describa su estrategia y presente los resultados en una tabla.
- c.) (0.7 pt.) Utilizando la misma estrategia del ítem previo, cambie el tipo de transformación binaria a OVO (*one vs one*) y presente los resultados en una tabla.
- d.) (0.5 pt.) Utilizando los datos de prueba compute la matriz de confusión de los dos (2) modelos obtenidos en los ítems previos. Obtenga la(s) métrica(s) de rendimiento que considere aporta(n) información relevante al problema de clasificación multiclase.

- e.) (0.2 pt.) Concluya sobre los resultados obtenidos de los clasificadores y sobre la estrategia implementada.

2. (2.5 pt.) El conjunto de datos “*Letter-Recognition*” (ver `data.rar`, carpeta “Punto 2”) del repositorio de bases de datos *Kaggle* (<https://www.kaggle.com/nishan192/letterrecognition-using-svm>) contiene datos pertenecientes a 26 clases. Cada observación del conjunto de datos está descrita por 16 atributos obtenidos a partir de una imagen en blanco y negro de cada una de las 26 letras del alfabeto inglés. Para convertir el problema a clasificación binaria, divida las 26 letras en dos clases de la siguiente forma:

- Clase 1: letras presentes en sus nombres y apellidos.
- Clase 2: el resto de las letras.

Con base en lo anterior,

- a.) (0.1 pt.) Divida los datos aleatoriamente en conjuntos de entrenamiento (80 %) y prueba (20 %).
- b.) (1.0 pt.) Entrene un MLP para resolver este problema de clasificación binaria. Para tal fin, tenga en cuenta los siguientes hiperparámetros y las arquitecturas descritas en la Tabla 1:
- Optimizador: Adam (*Adaptive Moment Estimation*).
  - Función de costo: Error Cuadrático Medio (MSE).
  - Número de épocas: 250.
  - Función de activación: Tangente hiperbólico,  $\forall h_i$ .

Tabla 1: Arquitecturas del MLP.

MLP No.	Arquitectura
1	16 - 8 - 1
2	16 - 16 - 1
3	16 - 32 - 1
4	16 - 8 - 4 - 1
5	16 - 16 - 8 - 1
6	16 - 24 - 16 - 1

- c.) (0.7 pt.) Utilizando el método de validación *k-fold cross-validation* obtenga la(s) métrica(s) de rendimiento que considere aporta(n) información relevante al problema de clasificación binaria, para cada una de las seis (6) arquitecturas.
- d.) (0.2 pt.) Seleccione el mejor modelo con base en el ítem previo. Justifique su respuesta.
- e.) (0.3 pt.) Utilizando los datos de prueba compute la matriz de confusión del modelo seleccionado. Obtenga la(s) métrica(s) de rendimiento que considere aporta(n) información relevante al problema de clasificación binaria.
- f.) (0.2 pt.) Concluya sobre los resultados obtenidos.

**Observaciones:**

Elabore un informe con el desarrollo del taller y envíelo al correo [warenas@javeriana.edu.co](mailto:warenas@javeriana.edu.co) con el asunto “Taller 6 PR\_1986” antes de la fecha establecida.