

Taller 3 AI: Clasificación Lineal

Facultad de Ingeniería
Departamento de Electrónica

Nota: fecha máxima de entrega del informe **martes 15 de marzo de 2022 a las 4:09 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 lineal en un problema de clasificación multiclase.
- Utilizar el método de validación cruzada para garantizar que el rendimiento de los clasificadores es independiente de la partición entre los datos de entrenamiento y prueba.

1. El conjunto de datos “data_2D” (ver archivo *data_2D.npy*) contiene datos pertenecientes a tres clases $\{a, b, c\}$.

- Divida los datos aleatoriamente en conjunto de entrenamiento (85 %) y prueba (15 %). Ambos conjuntos de datos almacénalos en una matriz. **Tenga en cuenta que los datos de prueba serán usados únicamente en el ítem e.**
- Visualice el conjunto de entrenamiento con un color diferente para cada clase.
- Utilizando el método de validación *k-fold cross-validation**, entrene c hiperplanos utilizando la estrategia *one-vs-all* con las siguientes técnicas de clasificación lineal:
 - Algoritmo de Mínimos Cuadrados.
 - Discriminante Logístico.
 - Discriminante de Fisher.
 - Perceptrón.

Para cada iteración obtenga la matriz de confusión de cada hiperplano y, con base en ella, calcule: 1) *Accuracy*, 2) *Error rate*, 3) *Sensitivity* y 4) *Specificity*. Genere una matriz por cada hiperplano de la forma:

Tabla 1: Matriz de validación propuesta para un (1) hiperplano.

i	w_0	w_1	w_2	A_{cc}	E_{rr}	S_e	S_p
1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
k

Con base en la matriz previa, seleccione el mejor hiperplano según su criterio. Como resultado, tendrá los tres (3) mejores hiperplanos para cada técnica de clasificación lineal.

* **Sugerencia:** asuma $k = 10$.

- d.) Implemente un clasificador por medio de los tres (3) hiperplanos seleccionados en el ítem anterior, en donde la clase es asignada de acuerdo al $\max\{y_a, y_b, y_c\}$. En la Figura 1 se ilustra el modelo para obtener el *score* que un hiperplano tiene de haber generado el patrón X_i .

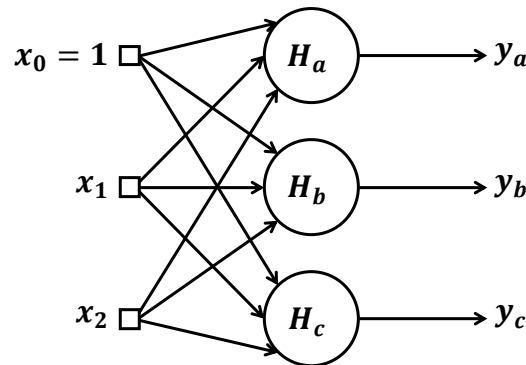


Figura 1: Modelo propuesto para el cálculo del *score*.

- e.) Compute la matriz de confusión para cada técnica de clasificación lineal implementada y calcule la(s) métrica(s) de rendimiento que considere pertinente. Tenga en cuenta que en el proceso de prueba no existe fase de validación.
- f.) Concluya sobre los resultados obtenidos de los clasificadores.

Observaciones:

Elabore un informe con el desarrollo del taller y envíelo al correo warenas@javeriana.edu.co con el asunto “Taller 3 AI_32700” antes de la fecha establecida.