

Análisis de Datos Parte II

Tarea #2: Clasificación Bayesiana paramétrica

Entregar el **23 de marzo**

Profesor: Dr. Wilfrido Gómez Flores

1. Instrucciones generales

- Para realizar los programas de la práctica se puede utilizar el lenguaje de programación que mejor te convenga.
- Entregar un reporte en un documento PDF de máximo 10 cuartillas utilizando el machote de L^AT_EX que puedes bajar en la página del curso.
- Entregar los códigos fuente de las rutinas utilizadas en las prácticas.
- En la carpeta personalizada en OneDrive, crear una nueva carpeta llamada **Tarea2** y colocar allí el reporte. Los códigos fuente se colocan en una subcarpeta llamada **Codigos**.
- El límite de entrega es hasta las 23:59 horas del día señalado. Por cada día de retraso, la calificación de la tarea se reducirá en un 10%.
- Favor de buscar al profesor para aclarar dudas sobre la tarea. No se resolverán dudas relacionadas al lenguaje de programación.

2. Descripción de la tarea

Fuente de datos

Para realizar esta práctica se proveen conjuntos de datos multiclase del mundo real en archivos `.csv`. Cada estudiante tiene asignado un conjunto de datos particular que puede descargarse de la página del curso. Cada renglón en el archivo es un dato, las primeras d columnas corresponden con las variables, y la última columna contiene las c etiquetas de clase.

Ejercicio #1: Clasificadores Bayesianos

Implementar los siguientes clasificadores vistos en la clase AD-04:

- Clasificador de mínima distancia Euclidiana (Ecuación 20).
- Clasificador de mínima distancia Mahalanobis (Ecuación 25).
- Clasificador Bayesiano general (Ecuación 27).
- Clasificador naive-Bayes (Ecuación 31).

Ejercicio #2: Evaluación de clasificadores

- a) Con los programas realizados en los ejercicios #1 y #2 de la Tarea #1, evaluar el desempeño de los clasificadores del Ejercicio #1.
- b) Utilizar validación cruzada de k-pliegues definir (p. ej. $k = 10$) y medir el desempeño de clasificación multiclase de cada pliegue. El desempeño estimado de cada índice es el promedio de todos los k pliegues.