



Tarea 4
CA-0411
I CICLO 2025

1. Instrucciones

A continuación se muestran las instrucciones de la tarea:

- La solución a cada tarea se debe subir en el aula virtual, no pueden ser enviadas por correo u otro medio.
- Las tareas se pueden hacer en parejas, pero cada persona deberá entregar la solución.
- Todas las tareas tienen el mismo valor en la nota final del curso.
- Las tareas se pueden entregar tarde, pero cada día de atraso tendrá un rebajo de 10 puntos.

2. Preguntas

1. (10 puntos) Complete las demostraciones de las reglas de asignación de la presentación de la semana 7.
2. (20 puntos) Diseñe un algoritmo en pseudocódigo para el Método del Análisis Discriminante Lineal, cuadrático y Naive Bayes según la teoría vista en clase. Programe esto en una clase que se llame `analisis_predictivo`. También incluya métodos para el gráfico del plano principal y del círculo de correlaciones. Compare los resultados con respecto a usar `lda`, `qda` y `bayes` de Python. Para esto use el archivo de datos `EjemploAD.csv`.

3. (20 puntos) Generalización matricial de las inercias:

Se consideran p variables continuas x_1, \dots, x_p observadas en una muestra Ω de n individuos. Cada individuo $i \in \Omega$ se representa con el vector fila $x_i^t = (x_{i1}, \dots, x_{ip})$ y cada variable x_j con el vector columna $x_j = (x_{1j}, \dots, x_{nj})^t$.

Sea y una variable cualitativa que determina una partición $P = \{C_1, \dots, C_r\}$ del conjunto Ω en r grupos. Definiciones:

- X : matriz $n \times p$ centrada por columnas.

- $D = \text{diag}(p_i)$: matriz de pesos de los individuos.
- A cada clase C_s se le asigna un peso $q_s = \sum_{i \in C_s} p_i$ y centro de gravedad $g_s = \frac{1}{q_s} \sum_{i \in C_s} p_i x_i$.
- $D_q = \text{diag}(q_j)$: matriz diagonal de pesos de las clases.
- C_g : matriz con las filas g_s^t como centros de gravedad.

Como X está centrada, el centro de gravedad global es $g = 0$ y la matriz de covarianza total es:

$$V = X^t D X = \sum_{i=1}^n p_i x_i x_i^t = \sum_{s=1}^r \sum_{i \in C_s} p_i x_i x_i^t$$

La matriz de covarianza dentro de clase V_s es:

$$V_s = \frac{1}{q_s} \sum_{i \in C_s} p_i (x_i - g_s)(x_i - g_s)^t$$

Y la matriz de covarianza intra-clase es:

$$V_W = \sum_{s=1}^r q_s V_s = \sum_{s=1}^r \sum_{i \in C_s} p_i (x_i - g_s)(x_i - g_s)^t$$

La matriz de covarianza inter-clase es:

$$V_B = \sum_{s=1}^r q_s g_s g_s^t = C_g^t D_q C_g$$

Demuestre lo siguiente:

- a) $V = V_B + V_W$
- b) $\sum_{s=1}^r q_s g_s = 0$, por tanto $\text{rango}(C_g) \leq r - 1$
- c) $\text{rango}(C_g) = \text{rango}(V_B)$

Además, con el archivo `EjemploAD.csv` calcule: g_A , g_B , g_C , V , V_B , V_W y verifique que $V = V_B + V_W$.

4. (20 puntos) Con la tabla `novatosNBA.csv`, que contiene métricas de desempeño de novatos de la NBA en su primera temporada, realice lo siguiente:

- a) Use LDA, QDA y Bayes en Python para generar un modelo predictivo utilizando el 80 % para aprendizaje y 20 % para prueba. Calcule los índices de precisión e interprete los resultados.
 - b) Construya un **DataFrame** comparando los modelos actuales con los de tareas anteriores, usando las métricas: Precisión Global, Error Global, Precisión Positiva (PP) y Precisión Negativa (PN). Determine cuál modelo es mejor.
5. (20 puntos) Con el conjunto de datos `diabetes.csv`, realice:
 - a) Cargue los datos en Python.
 - b) Genere modelos LDA, QDA y Bayes (75 % entrenamiento, 25 % prueba). Calcule matriz de confusión, precisión global y por clase. ¿Son buenos los resultados? Justifique.
 - c) Compare los modelos con los desarrollados en tareas anteriores en un **DataFrame** con: Precisión Global, Error Global, PP y PN.
 - d) Repita el inciso anterior seleccionando solo 6 variables predictoras. ¿Mejora la predicción?
6. (20 puntos) Para la siguiente tabla (vista en clase), se tiene una nueva fila: $t = (\text{Isabel}, F, 4, ?)$. Prediga manualmente si Isabel pertenece a la clase pequeño, mediano o alto.

Nombre	Género	Altura	Clase
Kristina	F	1	P
Jim	M	5	A
Maggi	F	4	M
Martha	F	4	M
Stephanie	F	2	P
Bob	M	4	M
Kathy	F	1	P
Dave	M	2	P
Worth	M	6	A
Steven	M	6	A
Debbie	F	3	M
Todd	M	5	M
Kim	F	5	M
Amy	F	3	M
Wynette	F	3	M

7. (20 puntos) Para la siguiente tabla, se tiene una nueva fila: $12 = (1, 3, 2, 4, ?)$. Prediga manualmente si el individuo es buen pagador o mal pagador.

Id	Monto Crédito	Ingreso Neto	Monto Cuota	Grado Académico	Buen Pagador
1	2	4	1	4	Sí
2	2	3	1	4	Sí
3	4	1	4	2	No
4	1	4	1	4	Sí
5	3	3	3	2	No
6	3	4	1	4	Sí
7	4	2	3	2	No
8	4	1	3	2	No
9	3	4	1	3	Sí
10	1	3	2	4	Sí
11	1	4	2	4	Sí

Entregables

Debe subir en el Aula Virtual el script de pruebas y un documento autoreproducible con la solución de la tarea.