

Parcial 2

1.

[20 puntos] Considere los datos que se encuentran en el archivo "Paises2.xlsx".

- (5 puntos) Determine la matriz de distancias con la norma usual (primero estandarice los datos).
- (5 puntos) Grafique los datos en el plano de las dos primeras coordenadas principales, de acuerdo al gráfico obtenido nota la presencia de algún outlier?
- (10 puntos) De acuerdo con el análisis ¿cuál es el vecino más cercano de Colombia?

2.

(5 puntos) En este ejercicio se pide encontrar una matriz de datos X de tal manera que la matriz de distancias en la norma $\|\cdot\|_2$ coincida con la siguiente matriz:

$$\begin{bmatrix} 0 & 3 \\ 3 & 0 \end{bmatrix}$$

3.

[20 puntos] Considere los datos que se encuentran en el archivo Clientes_D1.csv".

El objetivo es construir un cluster usando el algoritmo K-means.

- (5 puntos) Ejecute el algoritmo k-means, con $k = 2$, y determine el cluster al que pertenece la persona del primer registro.
- (10 puntos) Determine el número óptimo de grupos k de la siguiente manera: construye un gráfico bidimensional donde en el eje x varíe el número de clusters y en el eje y varíe la inercia. Este forma de encontrar k se llama el método de Elbow.
Cuál es el valor del k óptimo?
- (5 puntos) Ejecute el algoritmo k-means, con el k óptimo encontrado en el paso anterior, y determine el cluster al que pertenece la persona del primer registro.

4.

[5 puntos] Redacte la demostración del siguiente hecho:

La esperanza condicional $E[Y|X]$ es la solución del problema de minimización

$$f(x) = \arg \min_c E((Y - c)^2 | X = x)$$