

1. (30 pts) Considere  $\mathbf{D} = (d_{ij})$  matriz de distancias y suponga que consideramos  $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ . Sea  $\mathbf{A} = (a_{ij})$  con  $a_{ij} = -\frac{1}{2}d_{ij}^2$  y  $b_{ij} = a_{ij} - \bar{a}_i - \bar{a}_j + \bar{a}$ , donde

$$\bar{a}_i = \frac{1}{n} \sum_{j=1}^n a_{ij}, \quad \bar{a}_j = \frac{1}{n} \sum_{i=1}^n a_{ij}, \quad \bar{a} = \frac{1}{n} \sum_{i=1}^n \bar{a}_i = \frac{1}{n} \sum_{j=1}^n \bar{a}_j.$$

Considere

$$\mathbf{B} = \left( \mathbf{I} - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbf{A} \left( \mathbf{I} - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right).$$

Muestre que  $\mathbf{B}$  es semidefinida positiva.

2. (30 pts) Sea  $U_1$  y  $U_2$  dos variables aleatorias independientes  $U(0, 1)$ . Suponga  $\mathbf{X} = (X_1, X_2, X_3, X_4)^\top$  donde

$$X_1 = U_1, \quad X_2 = U_2, \quad X_3 = U_1 + U_2, \quad X_4 = U_1 - U_2.$$

Calcule la matrix de correlación  $\mathbf{R}$  de  $\mathbf{X}$ . ¿Cuántas componentes (PC) son de interés? Muestre que

$$\mathbf{t}_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{t}_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \\ 1 \end{pmatrix},$$

son vectores propios de  $\mathbf{R}$  asociados a  $\lambda$ 's no triviales. Interprete las dos primeras PC obtenidas.

3. (40 pts) Conjunto de datos desde Mardia, Kent y Bibby sobre 88 estudiantes quienes tomaron exámenes sobre 5 materias. Las primeras 2 materias fueron realizadas con el libro de texto cerrado, mientras que los últimos 3 se llevaron a cabo con el libro abierto. Realice un análisis de componentes principales y escriba un **muy breve** reporte de sus resultados.

El conjunto de datos, llamado `examScor.csv`, se encuentran disponibles en la página:

<http://fosorios.mat.utfsm.cl/teaching.html#MAT269>