

PRÁCTICO 4

ESCALADO MULTIDIMENSIONAL

1. A partir de la matriz de distancias entre ciudades del Uruguay, aplique un escalado multidimensional.
2. Se considera la tabla de individuos siguiente:

ind \ var	x_1	x_2	x_3	x_4	x_5	x_6
\mathbf{x}_1	1	1	0	0	1	1
\mathbf{x}_2	1	1	1	0	0	1
\mathbf{x}_3	1	0	0	1	0	1

Calcule las matrices de similaridad de Sokal-Michener y de Jaccard. Aplicar el escalado multidimensional en ambos casos.

3. Probar que si X es una matriz de datos centrada, se calcula su matriz de disimilitud Q y se aplica un MDS clásico, entonces las columnas de la matriz Y son las componentes principales de X . Comprobarlo a partir de un ejemplo de juguete.
4. Describa con precisión la técnica ISOMAP y cuales son los ventajas respecto del Escalado Multidimensional clásico. Buscar información sobre el conjunto de datos “Swiss roll data set” y aplicar ISOMAP a este conjunto. Comentar.
5. a) Sea E un conjunto de n individuos cuya matriz euclídea de distancias es D y cuya representación en coordenadas principales es X . Se desean obtener las coordenadas de un nuevo individuo, al que llamaremos individuo $n+1$, del cual se conocen los cuadrados de sus distancias a los n individuos del conjunto E . Si $\mathbf{d} = (d_{n+1,1}^2, \dots, d_{n+1,n}^2)'$ es el vector columna que contiene las distancias al cuadrado del individuo $n+1$ a los restantes, demostrar que la representación en coordenadas principales del individuo $n+1$ viene dada por

$$\mathbf{x}_{n+1} = \frac{1}{2} \Lambda^{-1} X'(\mathbf{q} - \mathbf{d})$$

siendo $\mathbf{q} = \text{diag}(Q)$ donde $Q = XX' = U\Lambda U'$.

- b) Se considera el siguiente conjunto de seis individuos formado por cinco animales, león, jirafa, vaca, oveja, gato doméstico, junto con el hombre. Se miden seis variables binarias sobre estos individuos: x_1 =tiene cola, x_2 =es salvaje, x_3 =tiene el cuello largo, x_4 =es animal de granja, x_5 =es carnívoro, x_6 =camina sobre cuatro patas.
 - 1) Obtener la matriz de datos, calcular los coeficientes de similaridad de Sokal-Michener y de Jaccard para cada par de individuos y obtenga las matrices de distancias asociadas.
 - 2) Obtener una representación en coordenadas principales utilizando la matriz de distancias calculada a partir del coeficiente de similaridad de Sokal y Michener.
 - 3) Sin volver a recalcular las coordenadas principales, añadir el elefante al conjunto de animales y obtener sus coordenadas