

Multidimensional Scaling

Ejercicios propuestos

1. Supongamos que las observaciones pueden ser clasificadas en k categorías excluyentes M_1, \dots, M_k . Supongamos que las probabilidades de estas categorías en la población Π_1 son $\mathbf{p} = (p_1, \dots, p_k)$ mientras que en la población Π_2 son $\mathbf{q} = (q_1, \dots, q_k)$. La distancia de Bhattacharyya entre estas dos poblaciones viene dada como

$$d(\mathbf{p}, \mathbf{q}) = \arccos\left(\sum_{i=1}^k \sqrt{p_i q_i}\right).$$

- (a) Crea una función que calcule la distancia de Bhattacharyya entre varias poblaciones.
- (b) La siguiente tabla contiene las frecuencias relativas de 4 categorías excluyentes que se han observado en 5 poblaciones diferentes. Calcula la matriz de distancias Bhattacharyya entre ellas. ¿Qué poblaciones son las más similares?

Población	M_1	M_2	M_3	M_4
1	0.36	0.21	0.23	0.20
2	0.66	0.18	0.11	0.05
3	0.01	0.24	0.62	0.13
4	0.43	0.38	0.08	0.11
5	0.16	0.07	0.09	0.68
6	0.22	0.37	0.25	0.16

- (c) ¿Es una matriz de distancias euclídea? Obtén una representación gráfica de las poblaciones.

2. Se han encontrado 5 herramientas arqueológicas en un yacimiento. Éstas estaban hechas con *piedra*, *bronce* y *hierro* según la siguiente matriz de incidencias:

ω_i	<i>Piedra</i>	<i>Bronce</i>	<i>Hierro</i>
1	0	1	0
2	1	1	0
3	0	1	1
4	0	0	1
5	1	0	0

- (a) Calcula la matriz de distancias adecuada para utilizando la función `dist()`.
- (b) Obtén una representación euclídea de estas herramientas. Interpreta los resultados.
3. Se han considerado 28 tipos de drogas y se ha preguntado a 68 personas acerca del consumo que hacen y se ha construido una matriz de similaridades con las frecuencias del consumo simultaneo dos a dos de estas sustancias. La matriz de similaridades se ha recogido en el fichero *drogas.dat*.

Estudia como se podrían representar los consumos de estas sustancias en una configuración euclídea.