

Algoritmo K Medias

Algoritmo K Medias

Tags: [Reconocimiento de Patrones](#), [Tecnicas de agrupamiento](#)

Suponemos que a priori nos dan el número de medias en nuestra distribución

- Colocar aleatoriamente las M medias en nuestro espacio de búsqueda $\mu_1, \mu_2, \dots, \mu_M$
- Calcular una medida de pertenencia de una observación x_i con una clase c_k . Usar lo siguiente:

$$b_{ik} = \begin{cases} 1, & \text{Si } x_i \text{ es el valor más cercano a la media } \mu_k \\ 0, & \text{Si no.} \end{cases}$$

- Calcular las medias usando

$$\mu_k = \frac{\sum_{i=1}^N x_i b_{ik}}{N_k} \quad k = 1, 2, \dots, M \quad N_k = \sum_{i=1}^N b_{ik}$$

- Repetimos los pasos 2 y 3 hasta que las medias no cambien.
-

Ejemplo:

Considere

$$\begin{aligned} x &= \{3, 4, 2, 3, 4, 4\} \\ y &= \{7, 8, 4, 7, 5, 8\} \end{aligned}$$

Tenemos 2 clases, o sea, 2 medias, que, de acuerdo al paso 1, se inicializan aleatoriamente

$$\begin{aligned} \mu_1 &= 5 \\ \mu_2 &= 7 \end{aligned}$$

- Paso 2

Calcular la medida de pertenencia con b_{ik} por medio de distancias. En este caso, distancia Manhattan $D(x, y) = |x_i - \mu_k|$

- Paso 3

Calcular las nuevas medias utilizando

$$\mu_k = \frac{\sum_{i=1}^N x_i b_{ik}}{N_k} \quad k = 1, 2, \dots, M \quad N_k = \sum_{i=1}^N b_{ik}$$

x_i	$D(x_i, \mu_1)$	$D(x_i, \mu_2)$	b_{i_1}	b_{i_2}	$x_i b_{i_1}$	$x_i b_{i_2}$
3	2	4	1	0	3	0
4	1	3	1	0	4	0
2	3	5	1	0	2	0
3	2	4	1	0	3	0
4	1	3	1	0	4	0
4	1	3	1	0	4	0
7	2	0	0	1	0	7
8	3	1	0	1	0	8
4	1	3	1	0	4	0
7	2	0	0	1	0	7
5	0	2	1	0	5	0
8	3	1	0	1	0	8
SUMA			8	4	29	30

$$\mu_1 = \frac{39}{N_1} = \frac{29}{8} = 3.62 \quad \mu_2 = \frac{30}{N_2} = \frac{30}{4} = 7.5$$

References

[Aglomerativas](#)

[K medias difusas](#)