Lloyd – um exemplo

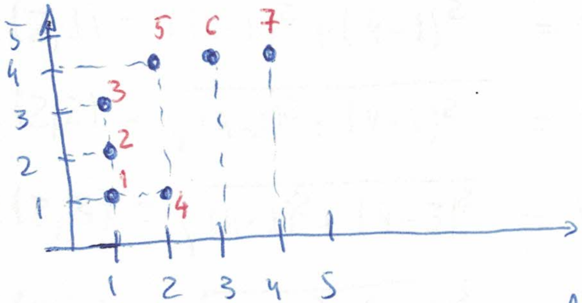
representante – média

$\boxed{\text{Ex } 25}$ $(\mathcal{A}, d, m) \longrightarrow (\mathbb{R}^2, \text{métrica euclideana}, \text{centróide})$

$K = 2$

$$D = \left\{ \overset{x^1}{\binom{1}{1}}, \overset{x^2}{\binom{1}{2}}, \overset{x^3}{\binom{1}{3}}, \overset{x^4}{\binom{2}{1}}, \overset{x^5}{\binom{2}{4}}, \overset{x^6}{\binom{3}{4}}, \overset{x^7}{\binom{4}{4}} \right\} \; ; \; N = |D| = 7. \; ; \; Eps = \text{valor positivo pequeno}$$

↓ a usar na $\boxed{CP}$



Aplicar o algoritmo de Lloyd com $\mathcal{M}(0) = \left\{ \binom{0}{0}, \binom{4}{4} \right\}$.

Vamos calcular a distância entre o representante do cluster m^j e o elemento x_i usando métrica euclideana:

$$\text{distância}(\underset{\text{representante } j}{j}, \overset{\text{elemento } i \text{ de } D}{i}) = \sqrt{(m_1^j - x_1^i)^2 + (m_2^j - x_2^i)^2}$$

Assim, as distâncias dos pontos de D ao 1º representante $\binom{0}{0}$ são:

$d(1,1) = \sqrt{(0-1)^2 + (0-1)^2} = \sqrt{2} \simeq 1.4142\ldots$

$d(1,2) = \sqrt{(0-1)^2 + (0-2)^2} = \sqrt{5} \simeq 2.23\ldots$

$d(1,3) = \sqrt{(0-1)^2 + (0-3)^2} = \sqrt{10} \simeq 3.16\ldots$

$d(1,4) = \sqrt{(0-2)^2 + (0-1)^2} = \sqrt{5} \simeq 2.23\ldots$

$d(1,5) = \sqrt{(0-2)^2 + (0-4)^2} = \sqrt{20} \simeq 4.47\ldots$

$d(1,6) = \sqrt{(0-3)^2 + (0-4)^2} = \sqrt{25} = 5$

$d(1,7) = \sqrt{(0-4)^2 + (0-4)^2} = \sqrt{32} \simeq 5.65\ldots$

As distâncias dos pontos de D ao 2º representante $\begin{pmatrix} 4 \\ 4 \end{pmatrix}$ são:

$d(2,1) = \sqrt{(4-1)^2 + (4-1)^2} = \sqrt{18} \simeq 4.24\ldots$

$d(2,2) = \sqrt{(4-1)^2 + (4-2)^2} = \sqrt{13} \simeq 3.60\ldots$

$d(2,3) = \sqrt{(4-1)^2 + (4-3)^2} = \sqrt{10} \simeq 3.16$

$d(2,4) = \sqrt{(4-2)^2 + (4-1)^2} = \sqrt{13} \simeq 3.60\ldots$

$d(2,5) = \sqrt{(4-2)^2 + (4-4)^2} = \sqrt{4} = 2$

$d(2,6) = \sqrt{(4-3)^2 + (4-4)^2} = 1$

$d(2,7) = \sqrt{(4-4)^2 + (4-4)^2} = 0$

A nova partição vai atribuir a cada ponto de D a nova partição de acordo c/ a menor distância ao seu representante.

$P(1) = \{ 1, 1, 1, 1, 2, 2, 2 \}$

$\min(1.41, 4.24) = 1.41 \rightarrow x^1 \in$ cluster 1

$\min(2.23, 3.6) = 2.23 \rightarrow x^2 \in$ cluster 1

$\min(3.16, 3.16) = 3.16 \rightarrow x^3 \in$ cluster 1 ou 2 (critério de empate – definir 1 ex se dist igual, escolhe cluster

$(\sqrt{5}, \sqrt{13}) = \sqrt{5} \rightarrow x^4 \in$ cluster 1

$\min(\sqrt{20}, \sqrt{13}) = \sqrt{13} \rightarrow x^5 \in$ cluster 2

$\min(5, 1) = 1 \rightarrow x^6 \in$ cluster 2

$\min(\sqrt{32}, 0) = 0 \rightarrow x^7 \in$ cluster 2

$\longrightarrow \mathcal{P}(1) = \{1, 1, 1, 1, 2, 2, 2\}$

Depois de calculada a nova partição, vamos calcular os novos representantes.
Como a métrica escolhida foi euclidiana, o representante é neste caso o centro de massa (média).

$$\mathcal{M}^1(1) = \frac{1}{4}\begin{pmatrix} 1+1+1+2 \\ 1+2+3+1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 5 \\ 7 \end{pmatrix} = \begin{pmatrix} 1.25 \\ 1.75 \end{pmatrix}$$

$$\mathcal{M}^2(1) = \frac{1}{3}\begin{pmatrix} 2+3+4 \\ 4+4+4 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 9 \\ 12 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

$$\therefore \mathcal{M}(1) = \left\{ \begin{pmatrix} 1.25 \\ 1.75 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right\}$$

CP pode ser por exemplo: $CP = \sum_{i=1}^{K} d(\mathcal{M}^i(1), \mathcal{M}^i(0))$

Neste caso $CP = d\left(\begin{pmatrix} 1.25 \\ 1.75 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) + d\left(\begin{pmatrix} 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 \\ 4 \end{pmatrix}\right)$. Se for maior do que ε,

continuamos!

Voltamos a calcular a distância de cada um dos pontos de D aos representantes dos clusters, para escolher novos representantes

- Distâncias a $\begin{pmatrix} 1.25 \\ 1.75 \end{pmatrix}$

$d(1,1) = \sqrt{(1.25-1)^2+(1.75-1)^2} \simeq \boxed{0.79}$

$d(1,2) = \sqrt{(1.25-1)^2+(1.75-2)^2} \simeq \boxed{0.35}$

$d(1,3) = \sqrt{(1.25-1)^2+(1.75-3)^2} \simeq \boxed{1.27}$

$d(1,4) = \sqrt{(1.25-2)^2+(1.75-1)^2} \simeq \boxed{1.06}$

$d(1,5) = \sqrt{(1.25-2)^2+(1.75-4)^2} \simeq 2.37$

$d(1,6) = \sqrt{(1.25-3)^2+(1.75-4)^2} \simeq 2.85$

$d(1,7) = \sqrt{(1.25-4)^2+(1.75-4)^2} \simeq 3.55$

- Distâncias a $\begin{pmatrix} 3 \\ 4 \end{pmatrix}$

$d(2,1) = \sqrt{(3-1)^2+(4-1)^2} \simeq 3.6$

$d(2,2) = \sqrt{(3-1)^2+(4-2)^2} \simeq 2.82$

$d(2,3) = \sqrt{(3-1)^2+(4-3)^2} \simeq 2.23$

$d(2,4) = \sqrt{(3-2)^2+(4-1)^2} \simeq 3.16$

$d(2,5) = \sqrt{(3-2)^2+(4-4)^2} \simeq \boxed{1}$

$d(2,6) = \sqrt{(3-3)^2+(4-4)^2} \simeq \boxed{0}$

$d(2,7) = \sqrt{(3-4)^2+(4-4)^2} \simeq \boxed{0}$

$\mathcal{P}(2) = \{1,1,1,1,2,2,2\}$ a partição não mudou!

$\downarrow$ calcular representantes - centroide - média

$\mathcal{M}(2) = \left\{ \begin{pmatrix} 1.25 \\ 1.75 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right\} \longrightarrow CP = 0$

$\therefore$ O cluster 1 tem os pontos x^1, x^2, x^3, x^4

O cluster 2 tem os pontos x^5, x^6, x^7