University of Nottingham
UK | CHINA | MALAYSIA

COMP3055 Machine Learning

Data Clustering Exercise Solution

Ying Weng

2024 Autumn

K-Means Algorithm

Question

Six 2-dimensional data points

(1, 1), (6, 6), (1, 3), (3, 2), (4, 3), (5, 4)

Cluster them into two clusters and find the cluster centres

K-Means Algorithm

Solution

(1) Euclidean distance to $m_1^0(1,1)$ | Euclidean distance to $m_2^0(3,2)$

$\sqrt{(1-1)^2+(1-1)^2}=0$	$\sqrt{(1-3)^2+(1-2)^2}=2.24$	$\because$ closer to m_1^0, $\therefore$ Assign to C_1
$\sqrt{(6-1)^2+(6-1)^2}=7.07$	$\sqrt{(6-3)^2+(6-2)^2}=5$	$\because$ closer to m_2^0, $\therefore$ Assign to C_2
$\sqrt{(1-1)^2+(3-1)^2}=2$	$\sqrt{(1-3)^2+(3-2)^2}=2.24$	$\because$ closer to m_1^0, $\therefore$ Assign to C_1
$\sqrt{(3-1)^2+(2-1)^2}=2.24$	$\sqrt{(3-3)^2+(2-2)^2}=0$	$\because$ closer to m_2^0, $\therefore$ Assign to C_2
$\sqrt{(4-1)^2+(3-1)^2}=3.61$	$\sqrt{(4-3)^2+(3-2)^2}=1.41$	$\because$ closer to m_2^0, $\therefore$ Assign to C_2
$\sqrt{(5-1)^2+(4-1)^2}=5$	$\sqrt{(5-3)^2+(4-2)^2}=2.83$	$\because$ closer to m_2^0, $\therefore$ Assign to C_2

$\therefore \{(1,1),(1,3)\} \in C_1 \Rightarrow m_1^{(0+1)}$:

$$\frac{1}{2}\sum_{i=1}^{2} x_i = \frac{1+1}{2} = 1$$

$$\frac{1}{2}\sum_{j=1}^{2} x_j = \frac{1+3}{2} = 2$$

$\{(6,6),(3,2),(4,3),(5,4)\} \in C_2 \Rightarrow m_2^{(0+1)}$:

$$\frac{1}{4}\sum_{i=1}^{4} x_i = \frac{6+3+4+5}{4} = 4.5$$

$$\frac{1}{4}\sum_{j=1}^{4} x_j = \frac{6+2+3+4}{4} = 3.75$$

(2) Euclidean distance to $m_1^1(1,2)$ | Euclidean distance to $m_2^1(4.5,3.75)$

$\sqrt{(1-1)^2+(1-2)^2}=1$	$\sqrt{(1-4.5)^2+(1-3.75)^2}=4.45$	$\because$ closer to m_1^1, $\therefore$ Assign to C_1
$\sqrt{(6-1)^2+(6-2)^2}=6.40$	$\sqrt{(6-4.5)^2+(6-3.75)^2}=2.70$	$\because$ closer to m_2^1, $\therefore$ Assign to C_2
$\sqrt{(1-1)^2+(3-2)^2}=1$	$\sqrt{(1-4.5)^2+(3-3.75)^2}=3.58$	$\because$ closer to m_1^1, $\therefore$ Assign to C_1
$\sqrt{(3-1)^2+(2-2)^2}=2$	$\sqrt{(3-4.5)^2+(2-3.75)^2}=2.30$	$\because$ closer to m_1^1, $\therefore$ Assign to C_1
$\sqrt{(4-1)^2+(3-2)^2}=3.16$	$\sqrt{(4-4.5)^2+(3-3.75)^2}=0.90$	$\because$ closer to m_2^1, $\therefore$ Assign to C_2
$\sqrt{(5-1)^2+(4-2)^2}=4.47$	$\sqrt{(5-4.5)^2+(4-3.75)^2}=0.56$	$\because$ closer to m_2^1, $\therefore$ Assign to C_2

$\therefore \{(1,1),(1,3),(3,2)\} \in C_1 \Rightarrow m_1^{(1+1)}$:

$$\frac{1}{3}\sum_{i=1}^{3} x_i = \frac{1+1+3}{3} = 1.67$$

$$\frac{1}{3}\sum_{j=1}^{3} x_j = \frac{1+3+2}{3} = 2$$

$\{(6,6),(4,3),(5,4)\} \in C_2 \Rightarrow m_2^{(1+1)}$:

$$\frac{1}{3}\sum_{i=1}^{3} x_i = \frac{6+4+5}{3} = 5$$

$$\frac{1}{3}\sum_{j=1}^{3} x_j = \frac{6+3+4}{3} = 4.33$$

K-Means Algorithm

Solution

(3) Euclidean distance to m_1^2 (1.67, 2) | Euclidean distance to m_2^2 (5, 4.33)

$\sqrt{(1-1.67)^2+(1-2)^2} = 1.20$	$\sqrt{(1-5)^2+(1-4.33)^2} = 5.20$	$\because$ closer to m_1^2, $\therefore$ Assign to C_1
$\sqrt{(6-1.67)^2+(6-2)^2} = 5.89$	$\sqrt{(6-5)^2+(6-4.33)^2} = 1.95$	$\because$ closer to m_2^2, $\therefore$ Assign to C_2
$\sqrt{(1-1.67)^2+(3-2)^2} = 1.20$	$\sqrt{(1-5)^2+(3-4.33)^2} = 4.22$	$\because$ closer to m_1^2, $\therefore$ Assign to C_1
$\sqrt{(3-1.67)^2+(2-2)^2} = 1.33$	$\sqrt{(3-5)^2+(2-4.33)^2} = 3.07$	$\because$ closer to m_1^2, $\therefore$ Assign to C_1
$\sqrt{(4-1.67)^2+(3-2)^2} = 2.54$	$\sqrt{(4-5)^2+(3-4.33)^2} = 1.66$	$\because$ closer to m_2^2, $\therefore$ Assign to C_2
$\sqrt{(5-1.67)^2+(4-2)^2} = 3.88$	$\sqrt{(5-5)^2+(4-4.33)^2} = 0.33$	$\because$ closer to m_2^2, $\therefore$ Assign to C_2

$\therefore \{(1,1), (1,3), (3,2)\} \in C_1$

$\{(6,6), (4,3), (5,4)\} \in C_2$