

CLUSTER ANALYSIS

Universitas Putra Indonesia YPTK Padang
Fakultas Ilmu Komputer
Program Studi Teknik Informatika

ALGORITMA K-MEANS

ALGORITMA K-MEANS

Algoritma K-Means

K-means merupakan metode clustering secara partitioning yang memisahkan data ke dalam kelompok yang berbeda

Dengan partitioning secara iteratif, K-Means mampu meminimalkan rata-rata jarak setiap data ke cluster-nya.

ALGORITMA K-MEANS

Dasar algoritma K-means adalah sebagai berikut :

- 1. Tentukan nilai k sebagai jumlah klaster yang ingin dibentuk.*
- 2. Bangkitkan k centroid (titik pusat klaster) awal secara random.*
- 3. Hitung jarak setiap data ke masing-masing centroid menggunakan rumus korelasi antar dua objek yaitu Euclidean Distance*
- 4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya.*

ALGORITMA K-MEANS

5. *Tentukan posisi centroid baru (C_k) dengan cara menghitung nilai rata-rata dari data-data yang ada pada centroid yang sama.*

$$C_k = \left(\frac{1}{n_k} \right) \sum d_i$$

Dimana n_k adalah jumlah dokumen dalam cluster k dan d_i adalah dokumen dalam cluster k .

6. *Kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama.*

ALGORITMA K-MEANS

Contoh :

Dengan menggunakan Algoritma K-Means, temukan grup terbaik dari dua cluster berikut :

$$M_1 = (2 , 5.0),$$

$$M_2 = (2 , 5.5),$$

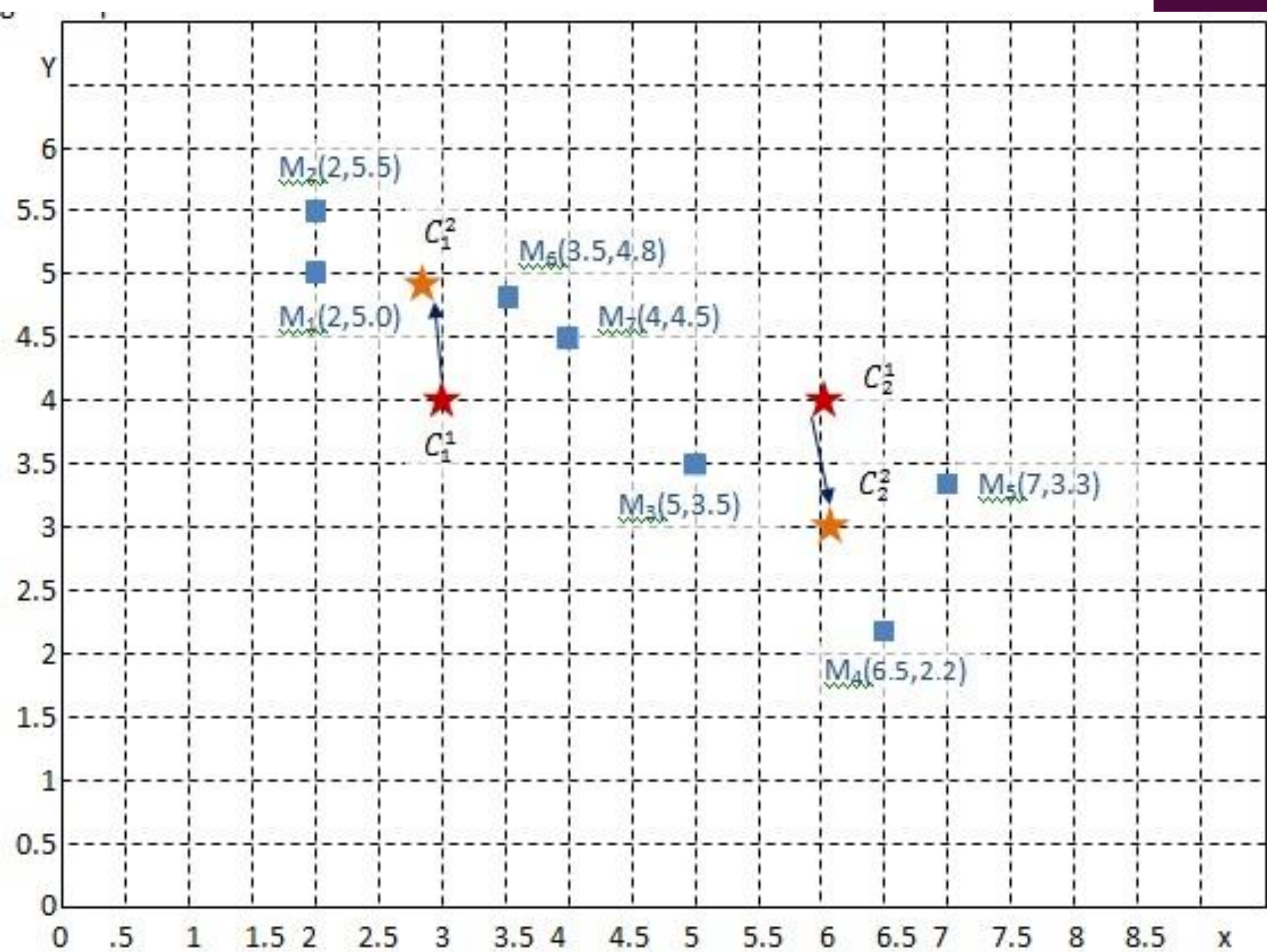
$$M_3 = (5 , 3.5),$$

$$M_4 = (6.5 , 2.2),$$

$$M_5 = (7 , 3.3),$$

$$M_6 = (3.5 , 4.8),$$

$$M_7 = (4 , 4.5)$$



ALGORITMA K-MEANS

Asumsi:

- ☐ *Semua data akan dikelompokkan ke dalam dua kelas*
- ☐ *Center points dari kedua cluster adalah $C_1(3, 4)$, $C_2(6, 4)$*

ALGORITMA K-MEANS

Iterasi 1

a. Menghitung Euclidean Distance dari semua data ke tiap titik pusat pertama

$$D_{11} = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(2 - 3)^2 + (5 - 4)^2} = \sqrt{2} = 1.41$$

$$D_{12} = \sqrt{(M_{2x} - C_{1x})^2 + (M_{2y} - C_{1y})^2} = \sqrt{(2 - 3)^2 + (5.5 - 4)^2} = \sqrt{3.25} = 1.80$$

$$D_{13} = \sqrt{(M_{3x} - C_{1x})^2 + (M_{3y} - C_{1y})^2} = \sqrt{(5 - 3)^2 + (3.5 - 4)^2} = \sqrt{4.25} = 2.06$$

$$D_{14} = \sqrt{(M_{4x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(6.5 - 3)^2 + (2.2 - 4)^2} \neq 15.49 = 3.94$$

$$D_{15} = \sqrt{(M_{5x} - C_{1x})^2 + (M_{5y} - C_{1y})^2} = \sqrt{(7 - 3)^2 + (3.3 - 4)^2} \neq 16.48 = 4.06$$

$$D_{16} = \sqrt{(M_{6x} - C_{1x})^2 + (M_{6y} - C_{1y})^2} = \sqrt{(3.5 - 3)^2 + (4.8 - 4)^2} \neq 0.88 = 0.94$$

$$D_{17} = \sqrt{(M_{7x} - C_{1x})^2 + (M_{7y} - C_{1y})^2} = \sqrt{(4 - 3)^2 + (4.5 - 4)^2} \neq 1.26 = 1.12$$

ALGORITMA K-MEANS

Dengan cara yang sama hitung jarak tiap titik ke titik pusat ke-2 dan kita akan mendapatkan

$$D_{21} = 4.12,$$

$$D_{22} = 4.27,$$

$$D_{23} = 1.12,$$

$$D_{24} = 1.87,$$

$$D_{25} = 1.22,$$

$$D_{26} = 2.62,$$

$$D_{27} = 2.06$$

ALGORITMA K-MEANS

b. Dari penghitungan Euclidean distance, kita dapat membandingkan:

	M_1	M_2	M_3	M_4	M_5	M_6	M_7
<i>jarak ke C_1</i>	1.41	1.80	2.06	3.94	4.06	0.94	1.12
C_2	4.12	4.27	1.12	1.87	1.22	2.62	2.06

$\{M_1, M_2, M_6, M_7\}$ anggota C_1 dan

$\{M_3, M_4, M_5\}$ anggota C_2

ALGORITMA K-MEANS

c. Hitung titik pusat baru

$$c_1 = \left(\frac{2 + 2 + 3.5 + 4}{4}, \frac{5 + 5.5 + 4.8 + 4.5}{4} \right) = (2.88, 4.9)$$
$$c_2 = \left(\frac{5 + 6.5 + 7}{3}, \frac{3.5 + 2.2 + 3.3}{3} \right) = (6.17, 3)$$

ALGORITMA K-MEANS

Iterasi ke 2

a. Hitung Euclidean distance dari tiap data ke titik pusat yang baru Dengan cara yang sama dengan iterasi pertama kita akan mendapatkan perbandingan sebagai berikut:

	M_1	M_2	M_3	M_4	M_5	M_6	M_7
Jarak ke C_1	0.88	1.04	2.57	4.55	4.55	0.64	1.21
C_2	4.62	4.86	1.27	0.87	0.88	3.22	2.64

ALGORITMA K-MEANS

b. Dari perbandingan tersebut kita tahu bahwa

$\{M_1, M_2, M_6, M_7\}$ anggota C_1 dan

$\{M_3, M_4, M_5\}$ anggota C_2

c. Karena anggota kelompok tidak ada yang berubah maka titik pusat pun tidak akan berubah.

Kesimpulan

$\{M_1, M_2, M_6, M_7\}$ anggota C_1 dan

$\{M_3, M_4, M_5\}$ anggota C_2