

1. a) Hasil cluster dari k-Means clustering ditentukan berdasarkan inisialisasi centroidnya. Jika dilakukan k-Means Clustering 2 kali berturut-turut dengan centroid yang sama, maka hasil clusternya sama. Namun jika centroidnya berbeda, maka hasil clusteringnya berbeda.

b) Dengan melihat plot Silhouette Coefficient maka didapatkan jumlah cluster terbaik adalah 5, karena pada  $k=5$  hasil nilai silhouette paling tinggi yang mendekati 1.



2. Lakukan clustering dan gambarkan dendogramnya menggunakan complete link AHC

Diketahui jarak masing-masing data

	1	2	3	4	5
1	0	9	3	6	11
2	9	0	7	5	10
3	3	7	0	9	2
4	6	5	9	0	8
5	11	10	2	8	0

Mengisi matriks jarak dengan mencari nilai maksimal

•> Iterasi 1

jarak terkecil adalah 2

maka terpilih kelompok 3 dan 5

Perbarui matriks jarak

	1	2	3,5	4
1	0	9	11	6
2	9	0	10	5
3,5	11	10	0	9
4	6	5	9	0

•> Iterasi 2

jarak terkecil adalah 5

maka terpilih kelompok 2 dan 4

perbarui matriks jarak

	1	2,4	3,5
1	0	9	11
2,4	9	0	10
3,5	11	10	0

•> Iterasi 3

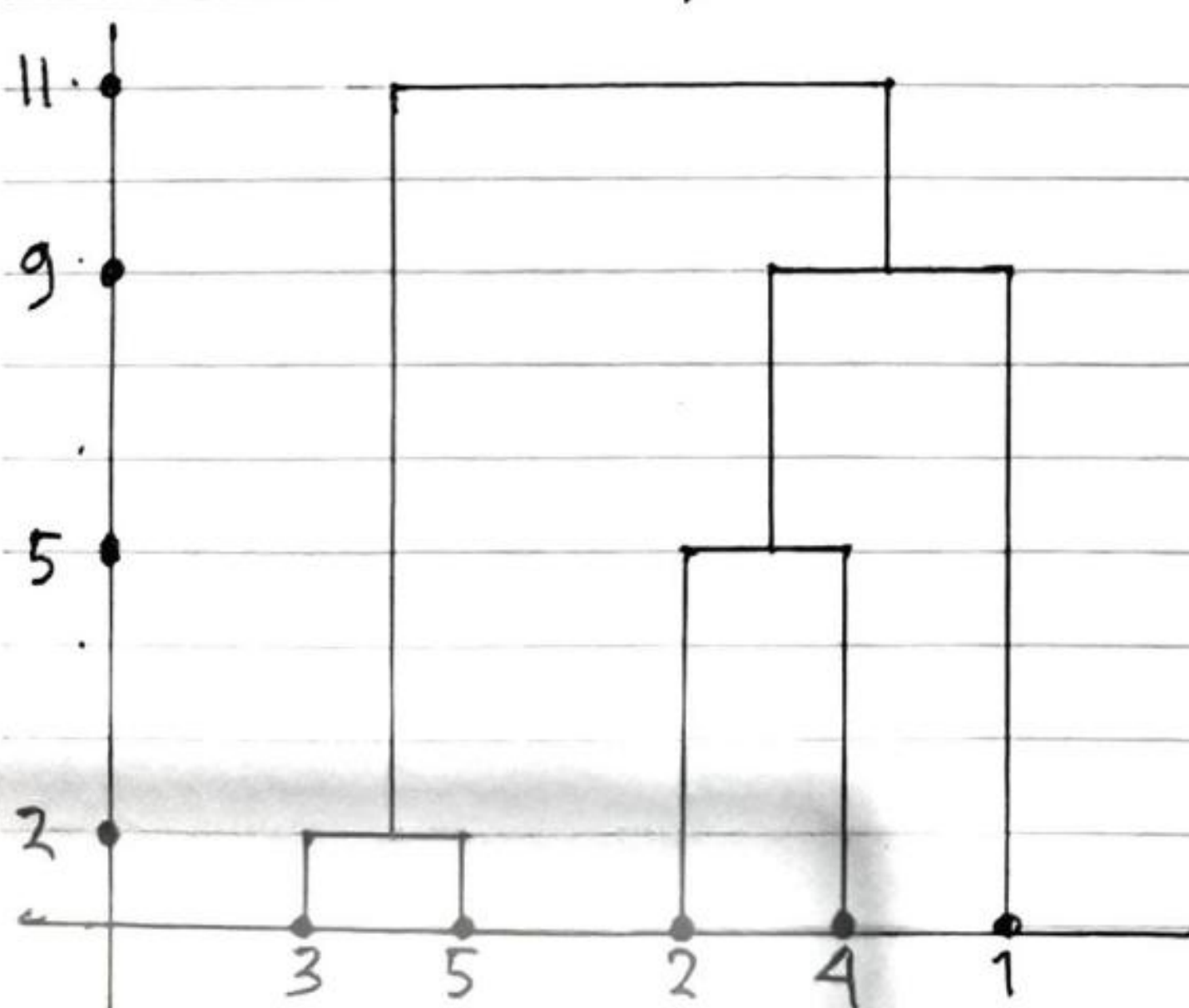
jarak terkecil adalah 9

maka terpilih kelompok 2,4 dan 1

perbarui matriks jarak

	2,4,1	3,5
2,4,1	0	11
3,5	11	0

Gambar Dendogram





3. a) k-means tidak bisa menghasilkan hasil cluster seperti pada gambar, karena k-means menggunakan centroid untuk clustering. Jadi tidak bisa membentuk cluster yang memanjang, dll.

b) •> kohesi : Untuk mengukur seberapa dekat jarak antar masing-masing data pada suatu cluster. Jadi kualitas cluster yang baik, jika memiliki nilai kohesi yang kecil.

•> Separasi : Untuk mengukur seberapa jauh jarak antar cluster. Jadi kualitas cluster yang baik, jika memiliki nilai separasi yang tinggi.



4. a) Saya tidak setuju, karena terdapat curse dimensionality ketika menambah fitur atau dimensi. Maka dari itu, Semakin banyak fitur belum tentu mendapatkan hasil yang maksimal. Dan jika fiturnya terlalu banyak bisa menyebabkan overfitting dan hasil yang didapatkan kurang baik.

b) Diketahui

$$\text{Eigen values} = \begin{pmatrix} 9,279 \\ 4,326 \end{pmatrix} \quad \text{Eigen vector} = \begin{pmatrix} 0,1222 & -0,9925 \\ 0,9925 & 0,1222 \end{pmatrix}$$

$$\text{Data A} = \begin{bmatrix} 8 & 6 \\ 9 & 3 \\ 4 & 7 \\ 5 & 2 \\ 7 & 10 \\ 9 & 8 \end{bmatrix}$$

$$\text{Rata-rata } x = \frac{42}{6} = 7$$

$$\text{Rata-rata } y = \frac{36}{6} = 6$$

$$\text{Jumlah} = 42 \quad 36$$

$$\text{Data Adjust} = \text{Data} - \text{rata-rata}$$

$$\text{Data Adjust} = \begin{bmatrix} 1 & 0 \\ 2 & -3 \\ -3 & 1 \\ -2 & -4 \\ 0 & 4 \\ 2 & 2 \end{bmatrix}$$

Transpose

$$\begin{bmatrix} 1 & 2 & -3 & -2 & 0 & 2 \\ 0 & -3 & 1 & -4 & 4 & 2 \end{bmatrix}$$

$$\text{Feature Vector} = \begin{bmatrix} 0,1222 & -0,9925 \\ 0,9925 & 0,1222 \end{bmatrix}$$

Transpose ↓

$$\text{feature vector}^T = \begin{bmatrix} 0,1222 & 0,9925 \\ -0,9925 & 0,1222 \end{bmatrix}$$

Final Data

$$\begin{bmatrix} 0,1222 \\ -2,7331 \\ 0,6259 \\ -4,2144 \\ 3,97 \\ 2,2294 \end{bmatrix}$$

$$\text{Final Data} = \text{Row Feature Vector} \times \text{Row Data Adjust}$$

Matriks  $6 \times 1$



# 5) Markov Chain

Diketahui probabilitas huruf pertama = e adalah 50%

$$P(e|ta) = P(x_0=e)P(x_1=t|x_0=e)P(x_2=a|x_1=t)$$

$$= 0,5 \times 0,1 \times 0,1$$

$$= 0,005$$

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$