Hierarchical Clustering

Overview

- Agglomerative Vs Divisive Clustering
- Contoh Aglomerative
- Latihan dan diskusi

Hierarchical Clustering

- Hierarchical Clustering adalah metode analisis kelompok yang berusaha untuk membangun sebuah hirarki kelompok data.
- Strategi pengelompokannya umumnya ada 2 jenis yaitu Agglomerative (Bottom-Up) dan Divisive (Top-Down). (Pada bagian ini akan dibatasi hanya menggunakan konsep Agglomerative).
- **Algoritma Agglomerative Hierarchical Clustering** :
  1. Hitung Matrik Jarak antar data.
  2. Ulangi langkah 3 dan 4 higga hanya satu kelompok yang tersisa.
  3. Gabungkan dua kelompok terdekat berdasarkan parameter kedekatan yang ditentukan.
  4. Perbarui Matrik Jarak antar data untuk merepresentasikan kedekatan diantara kelompok baru dan kelompok yang masih tersisa.
  5. Selesai.

Rumus Umum

- Membentuk Matrik Jarak,
  - misal dengan Manhattan Distance :

$$D_{man}(x,y)=\sum_{j=1}^{d}\left|x_j-y_j\right|$$

  - atau menggunakan Euclidian Distance :

$$D(x_2,x_1)=\sqrt{\sum_{j=1}^{d}\left|x_{2j}-x_{1j}\right|^2}$$

- Pengelompokan data secara hierarki:
  - Single linkage:

$$d_{uv}=\min\{d_{uv}\},d_{uv}\in D$$

  - Complete Linkage:

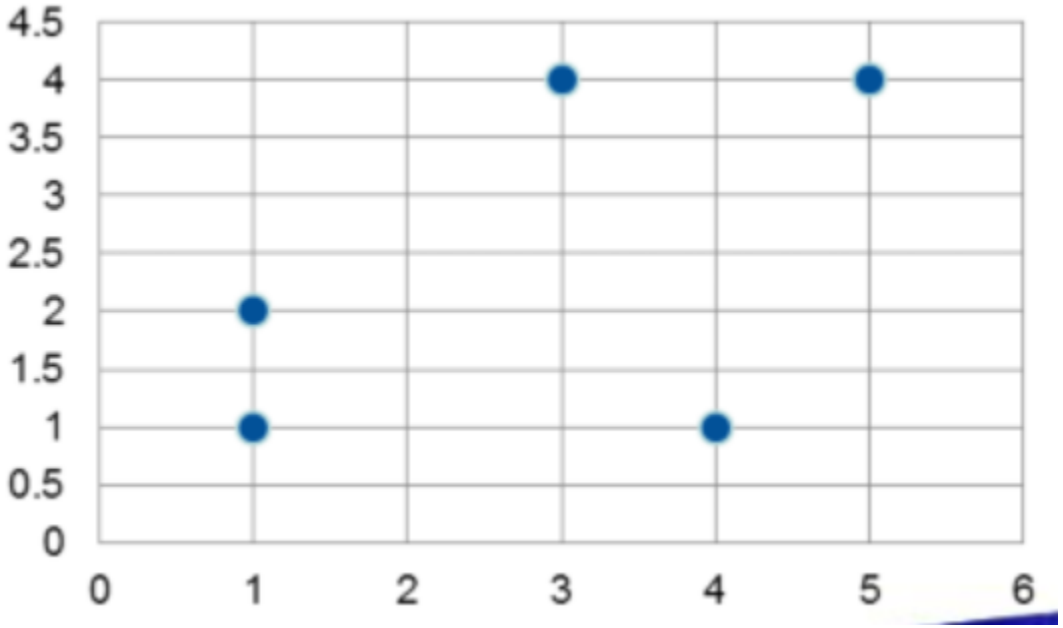
$$d_{uv}=\max\{d_{uv}\},d_{uv}\in D$$

  - Average Linkage:

$$d_{uv}=average\{d_{uv}\},d_{uv}\in D$$

Contoh Studi Kasus

Data	Fitur x	Fitur y
1	1	1
2	4	1
3	1	2
4	3	4
5	5	4



Selesaikan hierarchical clustering (agglomerative) dengan menggunakan min linkage, complete linkage , average linkage!

Single Linkage
Step 1. Hitung Matrik Jarak antar data.

	1	2	3	4	5
1	0				
2		0			
3			0		
4				0	
5					0

Data	Fitur x	Fitur y
1	1	1
2	4	1
3	1	2
4	3	4
5	5	4

- Jarak masing2 data
  - D(1,2): 3
  - D(1,3):1
  - D(1,4):5
  - D(1,5):7
  - D(2,3):4
  - D(2,4):4
  - D(2,5):4
  - D(3,4):4
  - D(3,5):6
  - D(4,5):2

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

Iterasi 1: (langkah 3) penggabungan 2 jarak terdekat

Dman	1	2	3	4	5
1	0	3		5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

- Dengan memperlakukan data sebagai kelompok, selanjutnya kita pilih jarak dua kelompok yang terkecil.
- terpilih kelompok 1 dan 3, sehingga kedua kelompok ini digabungkan.
- Langkah 4: perbarui matriks jarak

Single Linkage
Step 4. Hitung ulang Matrik Jarak antar data.

	1,3	2	4	5
1,3	0	3	4	?
2		0	4	4
4			0	2
5				0

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

- Jarak masing2 data (untuk data cluster dengan individu, pakai single linkage)
  - D((1,3),2): min(D(1,2), D(3,2) :Min(3,4) = 3
  - D((1,3),4): min(D(1,4), D(3,4) : Min(5,4) = 4
  - D((1,3),5):): ?

  - D(2,4):4
  - D(2,5):4

  - D(3,4):4
  - D(3,5):2

  - D(4,5)

Single Linkage
Step 3. gabungkan 2 kelompok data yg paling dekat

	1,3	2	4	5
1,3	0	3	4	6
2		0	4	4
4			0	2
5				0

Single Linkage
Step 4. Hitung ulang Matrik Jarak antar data.

	1,3	2	4,5
1,3	0	3	4
2		0	?
4,5			0

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

- Jarak masing2 data (untuk data cluster dengan individu, pakai single linkage)
  - D((1,3),2): min(D(1,2), D(3,2) :Min(3,4) = 3
  - D((1,3),(4,5)): min(D(1,4), D(1,5),D(3,4),D(3,5))
    - Min(5,7,4,6) =4
  - D(2,(4,5)):

Single Linkage
Step 3. gabungkan 2 kelompok data yg paling dekat

	1,3	2	4,5
~~1,3~~	0	~~3~~	~~4~~
2		0	4
4,5			0

Single Linkage
Step 4. Hitung ulang Matrik Jarak antar data.

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

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

- Jarak masing2 data (untuk data cluster dengan individu, pakai single linkage)
  - D((1,3,2),(4,5)):

- kelompok (132) dan (45) digabung untuk menjadi kelompok tunggal dari lima data, yaitu kelompok (13245) dengan jarak terdekat 4.

Dendogram

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

→

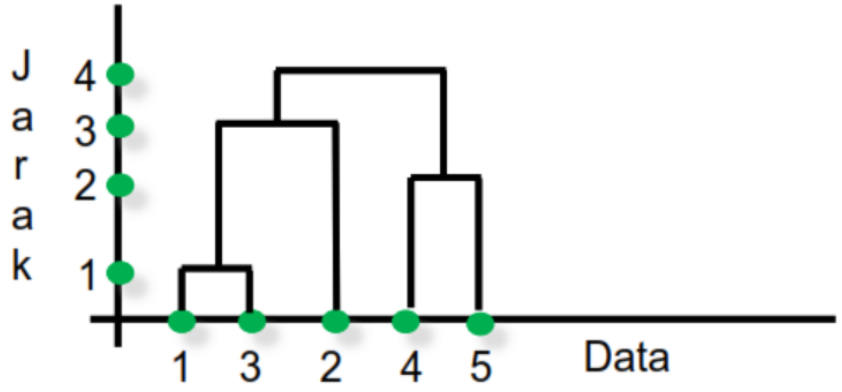
Dman	(13)	2	4	5
(13)	0	3	4	6
2	3	0	4	4
4	4	4	0	2
5	6	4	2	0

→

Dman	(45)	(13)	2
(45)	0	4	4
(13)	4	0	3
2	4	3	0

→

Dman	(132)	(45)
(132)	0	4
(45)	4	0



Complete Linkage
Step 1. Hitung Matrik Jarak antar data.

	1	2	3	4	5
1	0				
2		0			
3			0		
4				0	
5					0

Data	Fitur x	Fitur y
1	1	1
2	4	1
3	1	2
4	3	4
5	5	4

- Jarak masing2 data
  - D(1,2): 3
  - D(1,3):1
  - D(1,4):5
  - D(1,5):7
  - D(2,3):4
  - D(2,4):4
  - D(2,5):4
  - D(3,4):4
  - D(3,5):6
  - D(4,5):2

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

Iterasi 1: (langkah 3) penggabungan 2 jarak terdekat

Dman	1	2	3	4	5
1	0	3		5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

- Dengan memperlakukan data sebagai kelompok, selanjutnya kita pilih jarak dua kelompok yang terkecil.
- terpilih kelompok 1 dan 3, sehingga kedua kelompok ini digabungkan.
- Langkah 4: perbarui matriks jarak

Complete Linkage
Step 4. Hitung ulang Matrik Jarak antar data.

	1,3	2	4	5
1,3	0	4	5	?
2		0	4	4
4			0	2
5				0

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

- Jarak masing2 data (untuk data cluster dengan individu, pakai single linkage)
  - D((1,3),2): max(D(1,2), D(3,2) :Max(3,4) = 4
  - D((1,3),4): max(D(1,4), D(3,4) : Max(5,4) = 5
  - D((1,3),5):): ?
  - D(2,4):4
  - D(2,5):4
  - D(3,4):4
  - D(3,5):2
  - D(4,5)

Complete Linkage
Step 3. gabungkan 2 kelompok data yg paling dekat

	1,3	2	4	5
1,3	0	4	4	6
2		0	4	4
4			0	2
5				0

Complete Linkage
Step 4. Hitung ulang Matrik Jarak antar data.

	1,3	2	4,5
1,3	0	4	7
2		0	?
4,5			0

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

- Jarak masing2 data (untuk data cluster dengan individu, pakai single linkage)
  - D((1,3),2): max(D(1,2), D(3,2) :Min(3,4) = 4
  - D((1,3),(4,5)): max(D(1,4), D(1,5),D(3,4),D(3,5))
    - max(5,7,4,6) =7
  - D(2,(4,5)):

Complete Linkage
Step 3. gabungkan 2 kelompok data yg paling dekat

	1,3	2		4,5
1,3	0	4		7
2		0		4
4,5				0

Complete Linkage
Step 4. Hitung ulang Matrik Jarak antar data.

	1,3,2	4,5
1,3,2	0	7
4,5	7	0

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

- Jarak masing2 data (untuk data cluster dengan individu, pakai single linkage)
  - D((1,3,2),(4,5)):

- kelompok (132) dan (45) digabung untuk menjadi kelompok tunggal dari lima data, yaitu kelompok (13245) dengan jarak terjauh 7.

Dendogram

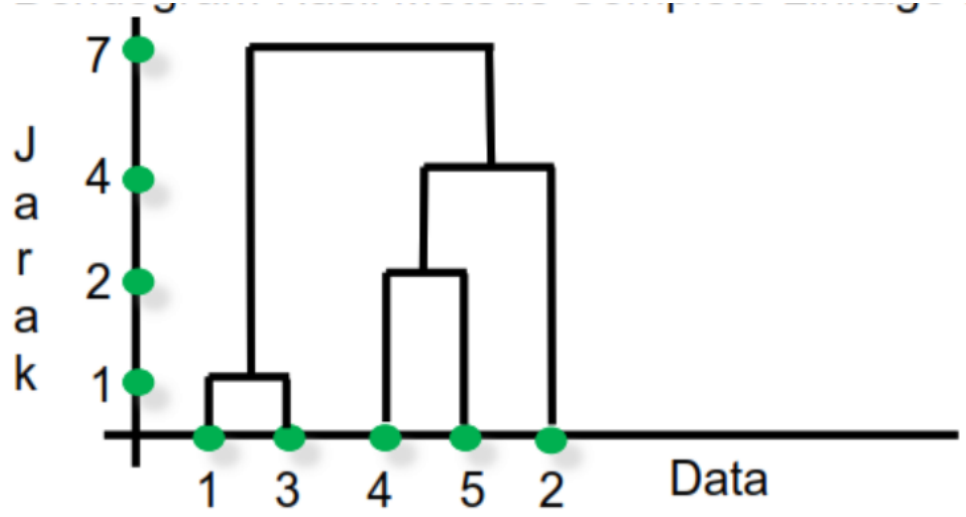
AlternatiV Dendogram

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

Dman	(13)	2	4	5
(13)	0	4	5	7
2	4	0	4	4
4	5	4	0	2
5	7	4	2	0

Dman	(45)	(13)	2
(45)	0	7	4
(13)	7	0	4
2	4	4	0

Dman	(452)	(13)
(452)	0	7
(13)	7	0



Single Vs Complete