Hierarchical clustering Technique

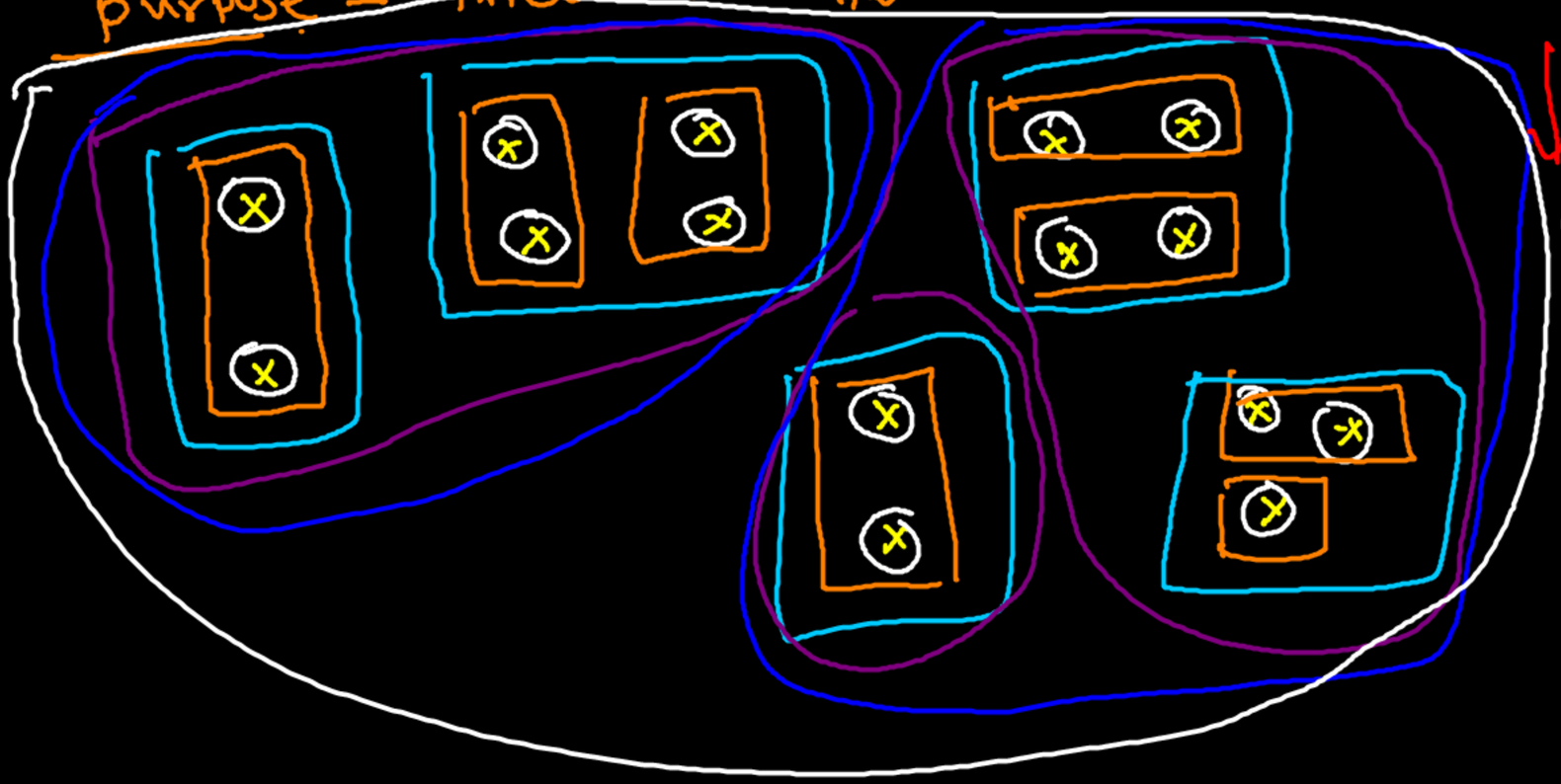
- Agglomerative & Divisive — Top to Bottom
  - Agglomerative: very popular
  - Divisive: Not used frequently
- Dendrograms

Grouping / Segmentation → Two methods

- Agglomerative: more popular
- Divisive: less popular.

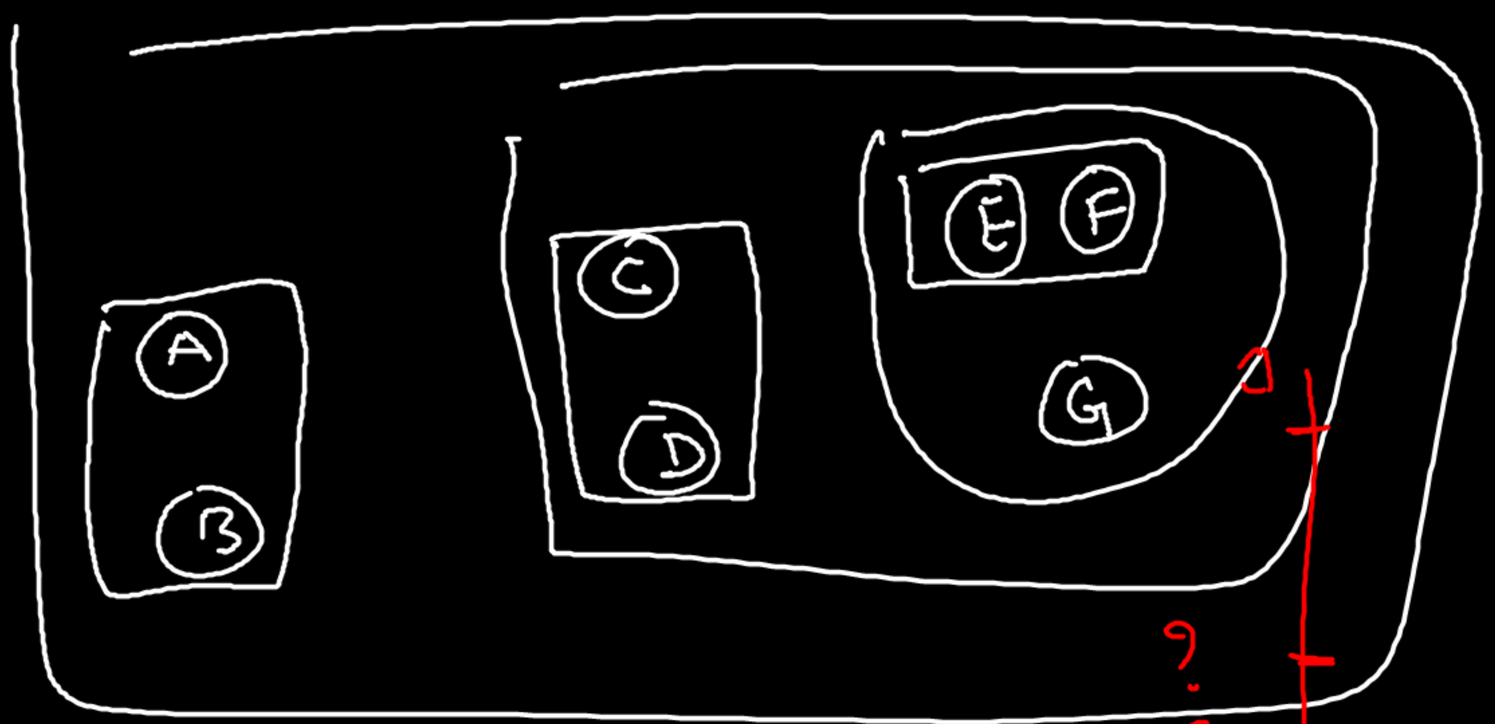
Agglomerative method – bottom to top approach

purpose – individual to one cluster

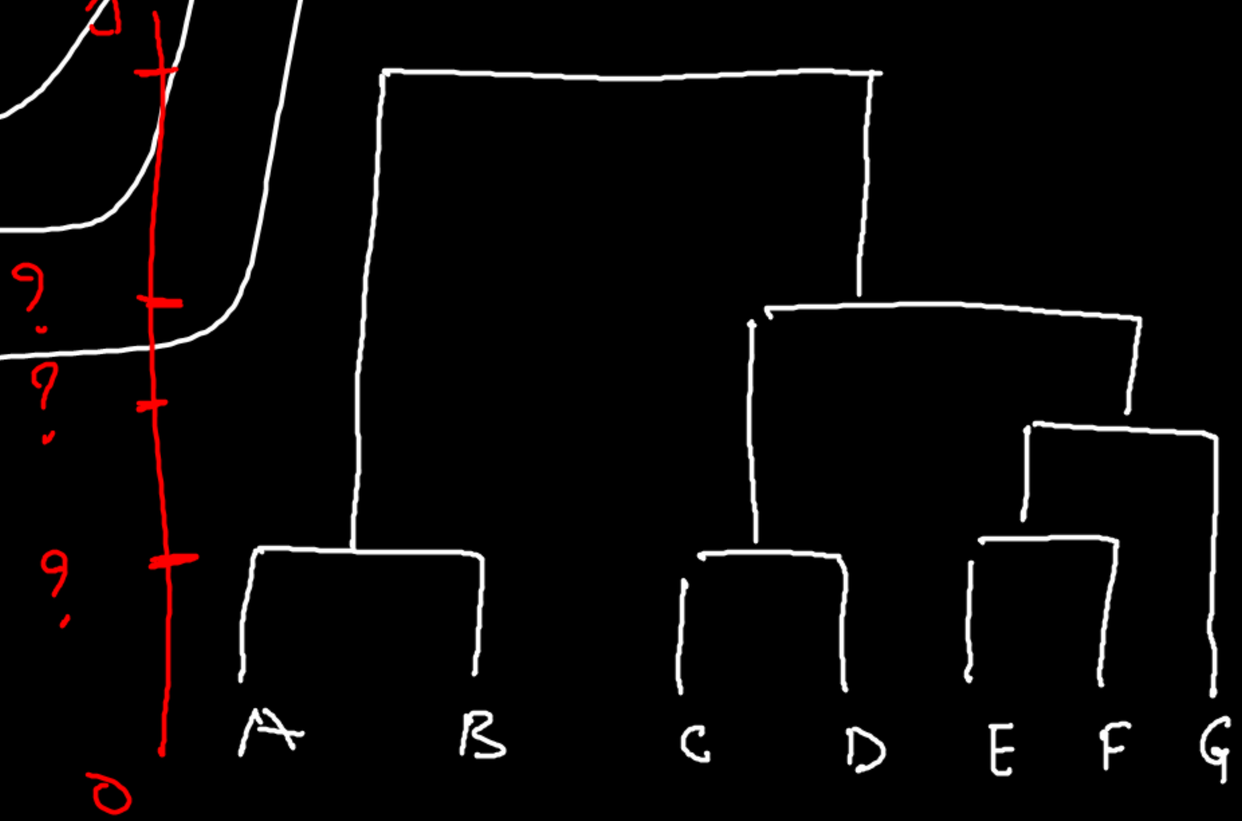


1. 15 Points – 15 Cluster
2. 8 Points – 8 cluster
3. 5 Points – 5 Cluster
4. 3 Points – 3 Cluster
5. 2 points – 2 cluster
6. 1 Point – 1 cluster

Final

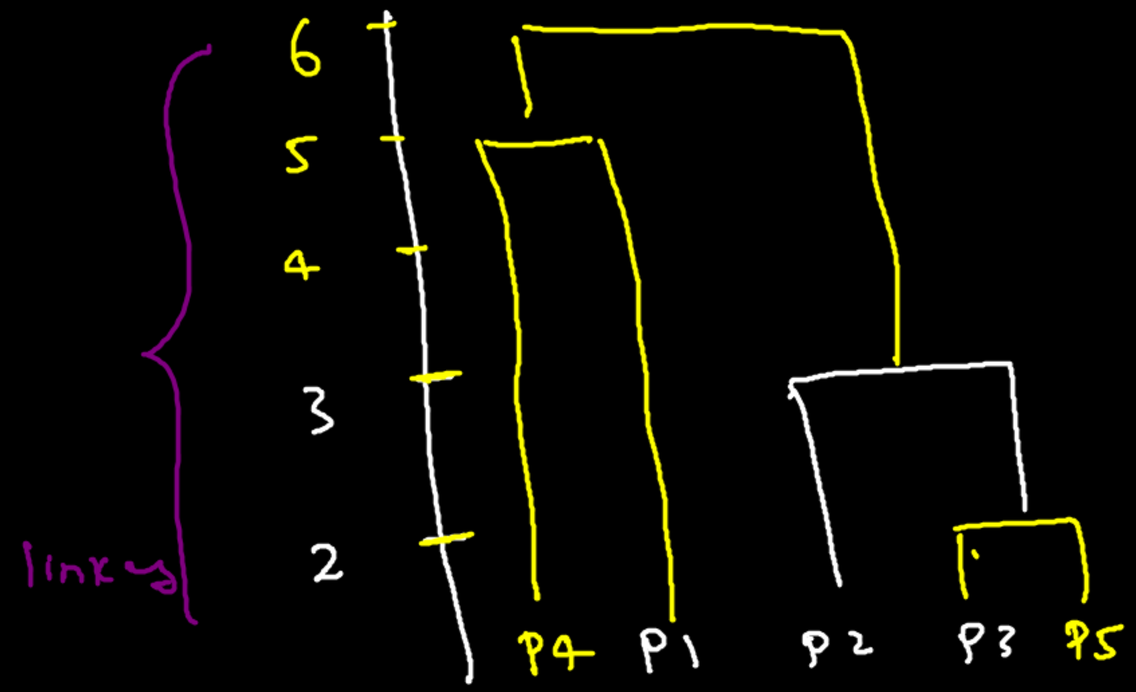


Dendrogram



	P1	P2	P3	P4	P5
P1	0				
P2	9	0			
P3	7	3	0		
P4	5	6	9	0	
P5	11	10	2	8	0

linkage



Bottom-Top approach

Assignment

	P1	P2	(P3,P5)	P4
P1	0			
P2	9	0		
P3,P5	7	3	0	
P4	5	6	8	0

$$dist\{min(P4,(P3,P5))\} = dist\{min(P4P3,P4P5)$$

$$dist\{min(9,8)\} = 8$$

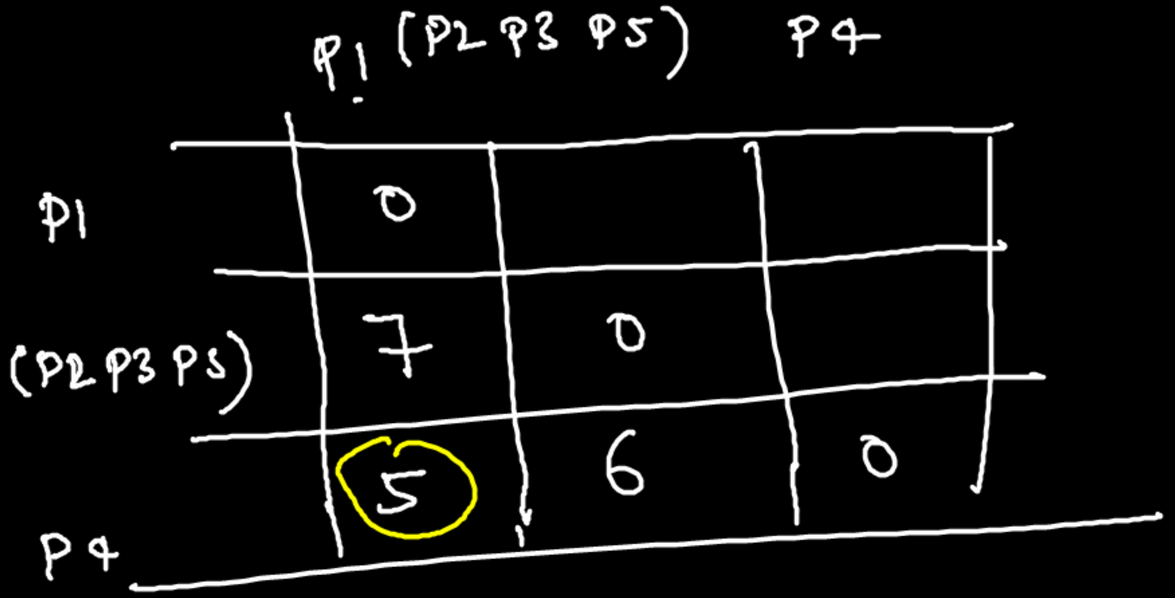
Similarly,

$$dist\{min(P2,(P3P5))\} = dist\{min(P2P3,P2P5)$$

$$= dist\{min(3,10)\} = 3$$

$$d[min(P1,(P3,P5))] = dist\{min((P1,P3),(P1,P5))\}$$

$$= dist\{min(7,11)\} = 7$$



$$dist\{min(P_1(P_2\,P_3\,P_5))\}$$

$$dist\{min((P_1,P_2)(P_1P_3)(P_1P_5))\}$$

$$dist\{min(9,7,11)\} = 7$$

$$dist\{min(P_2\,P_3\,P_5), P_4\}$$

$$dist\{min(P_2P_4, P_3P_4, P_4P_5)\}$$

$$= dist\{min\{6,9,8\}\} = 6$$

	(P1 P4)	(P2 P3 P5)
(P1 P4)	0	
(P2 P3 P5)	6	0

(P1 P'2 P3 P4 P5)

dist { min { (P1 P4) (P2 P3 P5) } }

dist { min (P1 P2, P1 P3, P1 P5, P2 P4, P3 P4, P4 P5) }

dist { min (9, 7, 11, 6, 9, 8) } = 6

Silhoutte distance (-1 to +1)

Cohesion (a)

measure the dist of one point to other point in same cluster

very low / less

Separation (b)

measure the dist of one point to other point in diff cluster

very high.

Intra - within

Inter - outside

-1 = Cluster is wrong

0 = diff is same

1 (more than 0.35) = both cluster / all clust — are far away

$$\frac{b-a}{\max(b,a)}$$

$$\frac{0.8-0.2}{\max(0.8,0.2)} = \frac{0.6}{0.8} = \frac{6}{8} = 0.7$$

case 1 $a = 0.2$ $b = 0.8$

case 2 $a = 0.8$ $b = 0.2$ $\rightarrow$ $\frac{0.2-0.8}{\max(0.2,0.8)} = \frac{-0.6}{0.8} = -0.7$

DBSCAN

Clust

Noise

• → outl

Density Based Spatial Clustering & Application with Noise

Core philosphy

Q: how to measure density ?

Hierarchical cluste - dentrogram [dist
K-means — Elbow meth] Euclidn dw-

key-idea :— , minpoints , Epsilon (Eps) , core Points ,

border/boundry point , Noise

DBSCAN

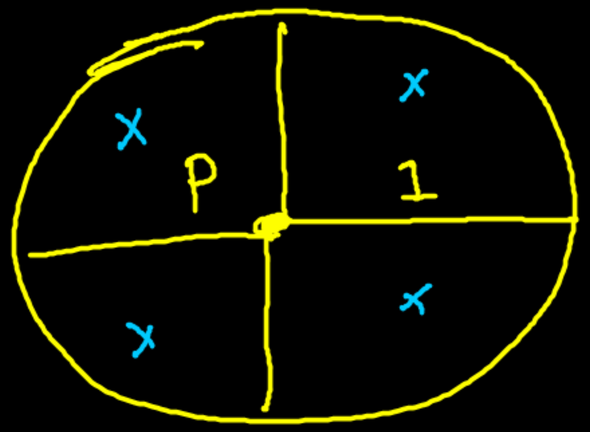
4

Min Points & Epsilon (Eps) - Density

Eps → radius

↓

Hyperparameter of DBSCAN



① Density at a "p" :- # Points within a hypersphere (in D) of radius (Eps) around 'p'

Circle = 2D

Sphere = 3 D

condition :-

Density = 4 (minpts)

Eps (radius) = 1

eps = 1 - satisfied

minpts = 4 - Sati.

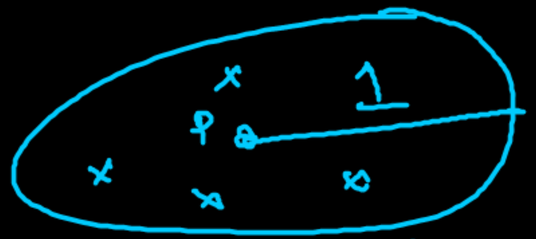
dense region → cluster

* Dense Region :— a hypersphere/sphere/Circle of radius eps (1) that contains atleast minpts (4) points

Core Point, Border point & Noise

$D = \{x_i\}$ No $\underline{y}$, minpts = 4 & EPS = 1

① Core points :- if p has $\geq$ minpts in an EPS reading around it. it is called core point.
'p'



↳ core point

↳ Not core point
X

(2) Border Point

(a) P is not a Core Point

⇒ P has < minPts in EPS radius

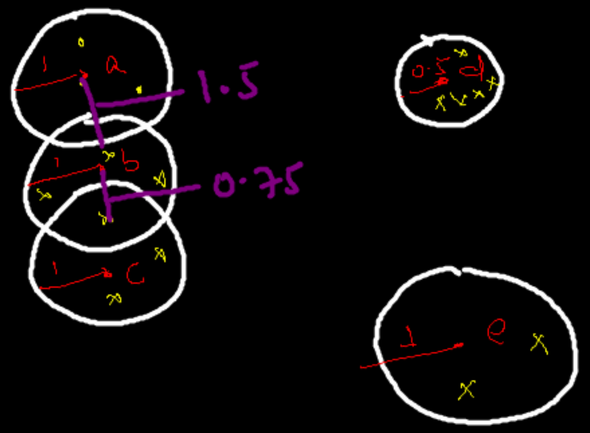
(b) P ∈ Neighborhood (Core Point)

Q = Core Point

dist (P, Q) ≤ EPS

③ Noise / outlier / sparse region

→ Neither core point nor a border point



hyperparameters

minpts = 4

Eps = 1

	Eps	Minpts	Core	Border	Noise
a	1	2	X	X	✓
b	1	4	✓	X	X
c	1	3	X	✓	X
d	0.5	5	X	X	✓
e	1	2	X	X	✓

dist (a,b) = 1.5

dist (b,c) = 0.75