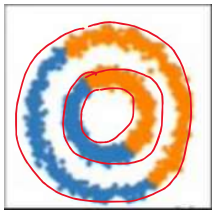
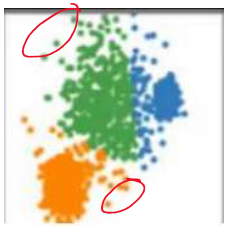
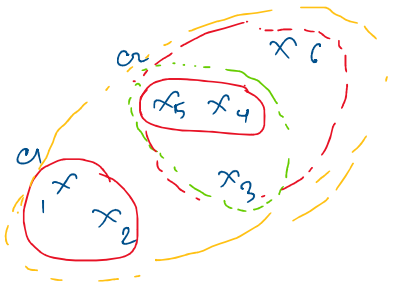
Hierarchical Clustering

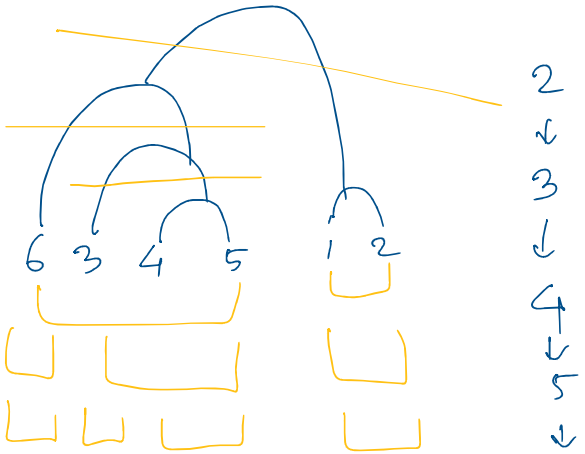
Problem
1) number of cluster specific
2) Elbow method
3) sensitive to outliers ✓
4) non spherical shape





Types
1) Agglomerative clustering → most used
2) Divisive clustering



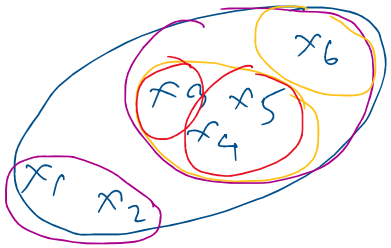


2 ↓ 3 ↓ 4 ↓ 5 ↓ 6

$c_1 - c_3$ ✗ $c_2 - c_3$
$c_1 - c_2$

Divisive clustering

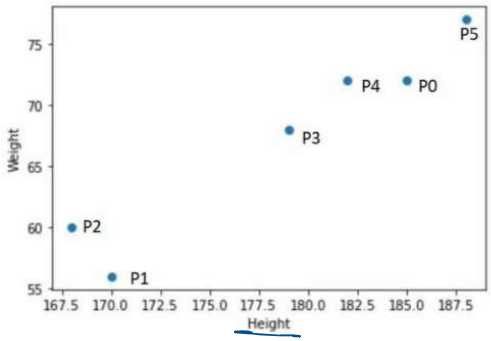
1 ↓ 2



1 ↓ 2 ↓ 3 ↓ 4 ↓ 5 ↓ 6

Agglomerative clustering

	Height	Weight
$P_0 \rightarrow$	185	72
$P_1 \rightarrow$	170	56
$P_2 \rightarrow$	168	60
$P_3 \rightarrow$	179	68
$P_4 \rightarrow$	182	72
$P_5 \rightarrow$	188	77



$A(x_1, y_1)$ $B(x_2, y_2)$

$D(AB)$ = Euclidean Formula $= \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$

(P_0, P_1) $(P_0\ P_2)$ $(P_0\ P_3)$ $(P_0\ P_4)$ $(P_0\ P_5)$

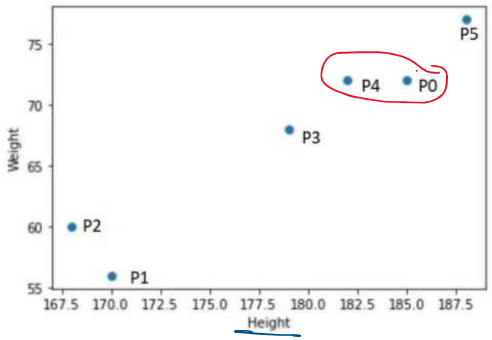
$(P_1\ P_2)$ $(P_1\ P_3)$ $(P_1\ P_4)$ $(P_1\ P_5)$

$(P_2\ P_3)$ $(P_2\ P_4)$ $(P_2\ P_5)$

$(P_3\ P_4)$ $(P_3\ P_5)$

(P4 P5)

	P0	P1	P2	P3	P4	P5
P0	0					
P1	21.93	0				
P2	20.81	4.47	0			
P3	7.21	15	13.6	0		
P4	3	20	18.44	5	0	
P5	5.83	27.66	26.25	12.73	7.81	0



P0, P4 → 3

①

	[P0,P4]	P1	P2	P3	P5
[P0,P4]	0				
P1	20	0			
P2	18.44	4.47	0		
P3	5	15	13.6	0	
P5	5.83	27.66	26.25	12.73	0

After merging P0 and P4

1

	P0	P1	P2	P3	P4	P5
P0	0					
P1	21.93	0				
P2	20.81	4.47	0			
P3	7.21	15	13.6	0		
P4	3	20	18.44	5	0	
P5	5.83	27.66	26.25	12.73	7.81	0

$P_1, [P_0, P_4] \rightarrow \min(d(P_1, P_0), d(P_1, P_4))$

$(21.93, 20)$

$= 20$

$P_2, [P_0, P_4] \rightarrow \min(d(P_2, P_0), d(P_2, P_4))$

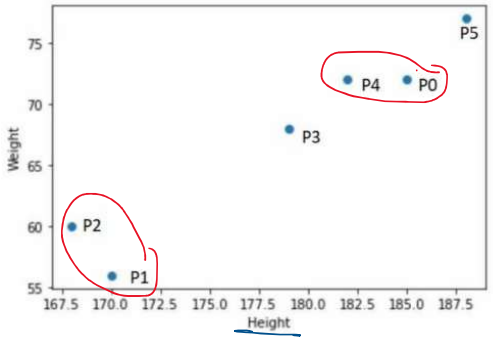
$\rightarrow (20.81, 18.44)$

$= 18.44$

2

	[P0,P4]	[P1,P2]	P3	P5
[P0,P4]	0			
[P1,P2]	18.44	0		
P3	5	13.6	0	
P5	5.83	26.25	12.73	0

Merging P1 and P2



$[P_0,P_4], [P_1,P_2] \rightarrow \min(d(P_0,P_1), d(P_0,P_2), d(P_4,P_1), d(P_4,P_2))$

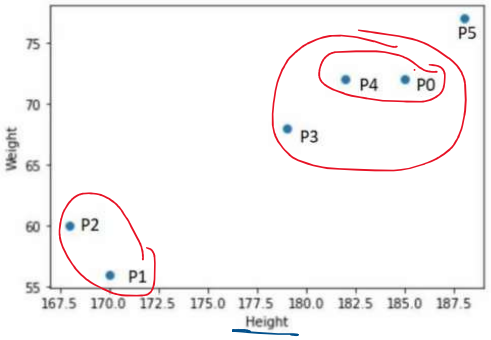
$= \min(21.93, 20.81, 20, 18.44)$

$= 18.44$

	P0	P1	P2	P3	P4	P5
P0	0					
P1	21.93	0				
P2	20.81	4.47	0			
P3	7.21	15	13.6	0		
P4	3	20	18.44	5	0	
P5	5.83	27.66	26.25	12.73	7.81	0

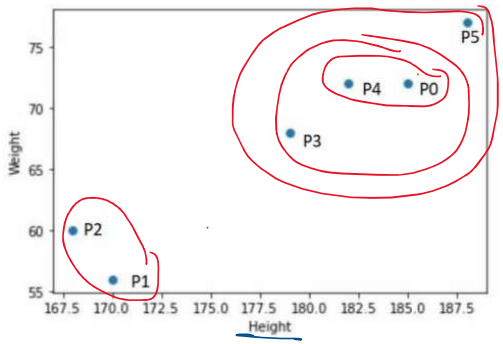
	[P3,[P0,P4]]	[P1,P2]	P5
[P3,[P0,P4]]	0		
[P1,P2]	13.6	0	
P5	5.83	26.25	0

Merging of P3 and [P0,P4]



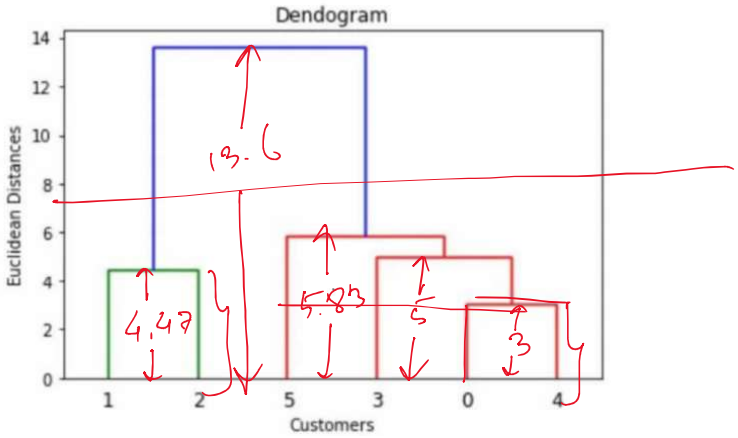
	[P5,[P3,[P0,P4]]]	[P1,P2]
[P5,[P3,[P0,P4]]]	0	
[P1,P2]	13.6	0

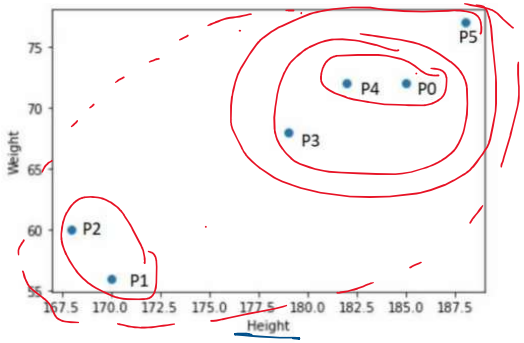
Merging of P5 and [P3,[P0,P4]]

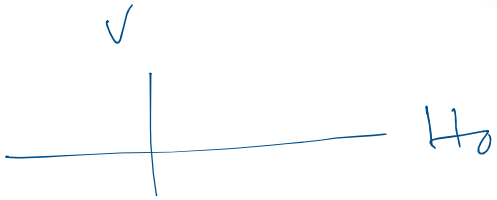


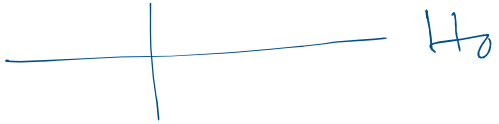
	[[P1,P2],[P5,[P3,[P0,P4]]]]
[[P1,P2],[P5,[P3,[P0,P4]]]]	0

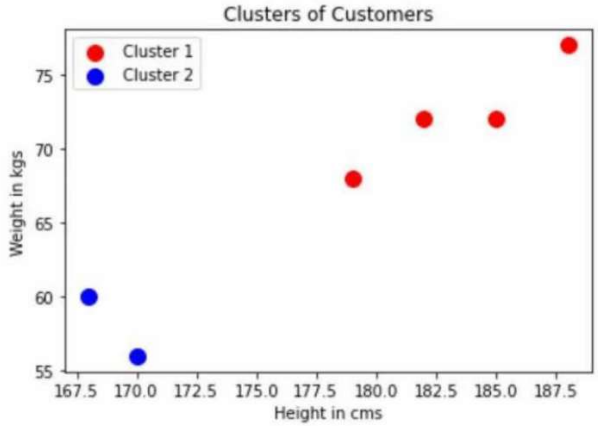
Merging of [P1,P2] and [P5,[P3,[P0,P4]]]





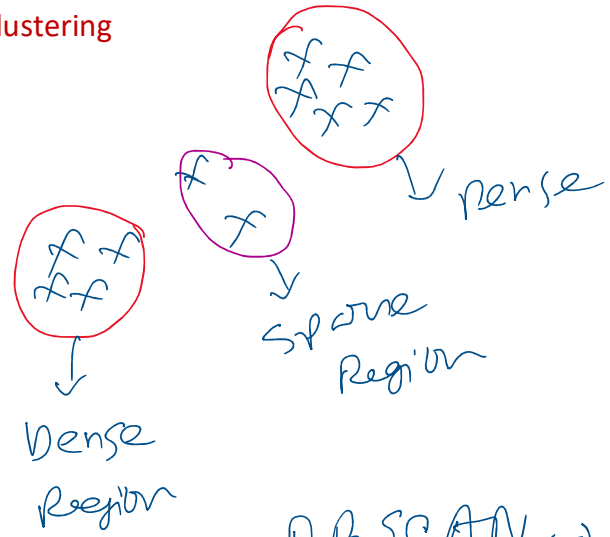






Final Clusters

Density Base Clustering



Dense → घना
Sparse → बिखरा हुआ

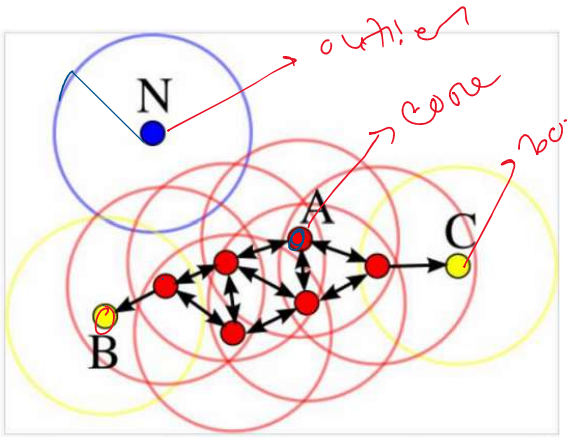
DBSCAN → Density-Based spatial clustering

Region

DBSCAN → Density-Based spatial clustering of Application with noise

OPTICS → Another Desity Based Algorith

DBSCAN



In this diagram, minPts = 4. Point A and the other red points are core points, because the area surrounding these points in an ε radius contain at least 4 points (including the point itself). Because they are all reachable from one another, they form a single cluster. Points B and C are not core points, but are reachable from A (via other core points) and thus

outlier

core

border

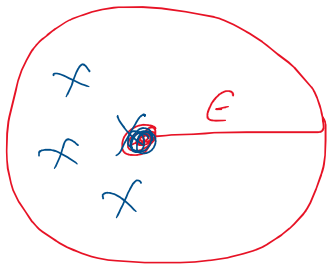
● → core point

● → border point

● → Outliers

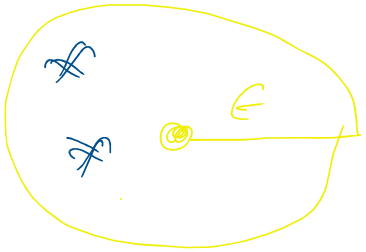
minpts = 4 ϵ = radius

* core point number of point ≥ 4



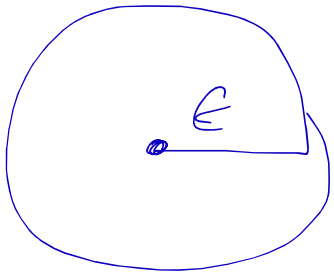
* border point

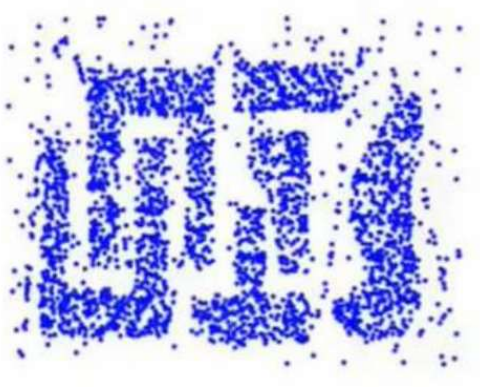
number of point 3 < 4
and
1 atleast = core



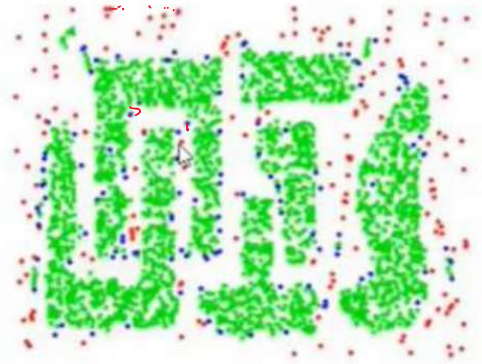
* outlier :

No core point and No Border point



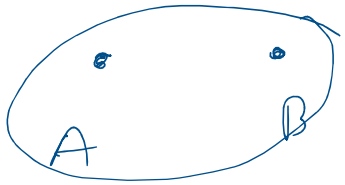


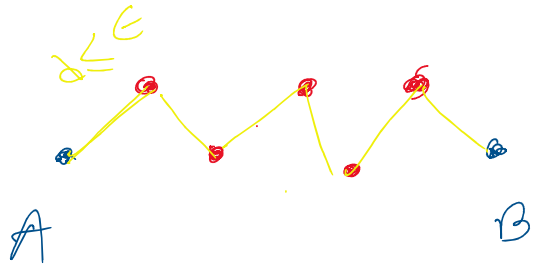
Original Points

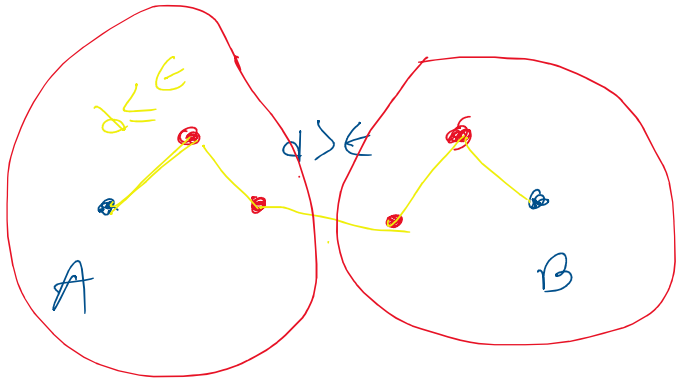


Point types: core, border and noise

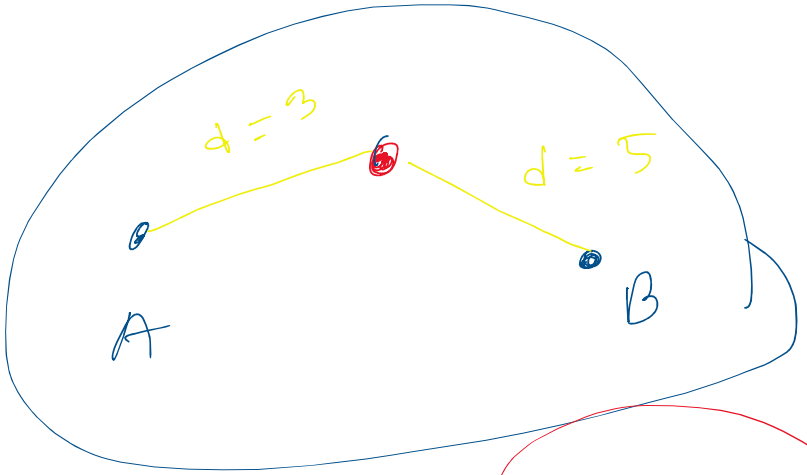
Density connected Points







∈ = 5



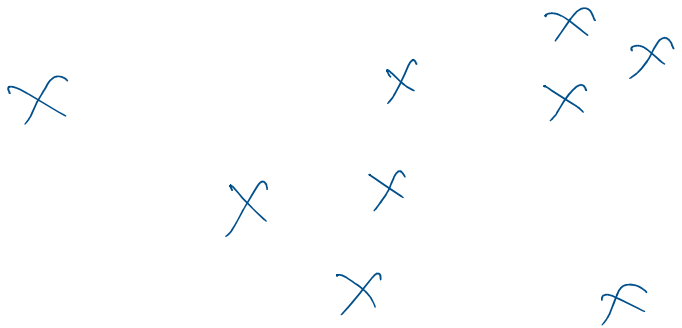
d ≤ ∈

∈ = 5

DBSCAN Algorithm

DBSCAN Algorithm

Step 0: $minpts = 3$ $\epsilon = 5$



Step 1:

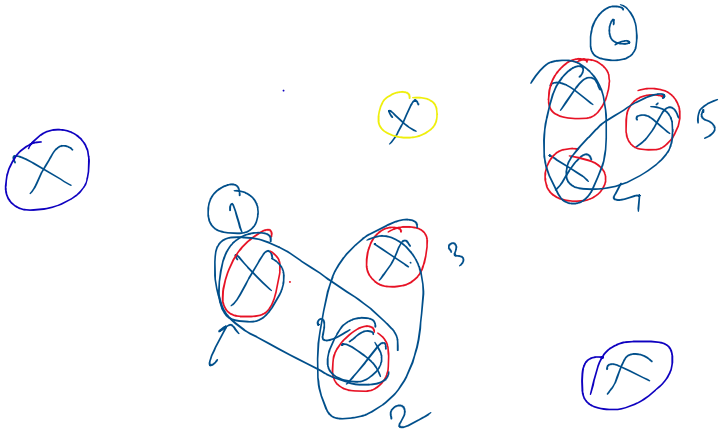
$P_1 P_2 = d = 3 < \epsilon$ ✓

$P_2 P_3 = d = 4 < \epsilon$ ✓

$P_3 P_4 = d = 7 > \epsilon$ ✗

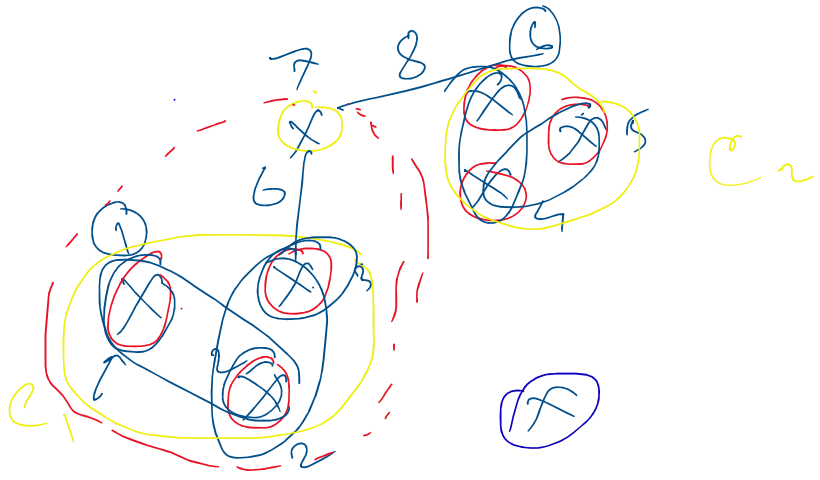
$P_6 P_4 = d = 3 < \epsilon$ ✓

$P_1 P_5 = d = 2 < \epsilon$ ✓



$p_4 p_5 = d = 2 \checkmark \in$

Step 2 :⇒



Border point 7, c_1 ∋ $p_3 = 6$ ✗

Border poin 7, c_2 ∋ $p_6 = 8$ ✗

Step 3 : outliers →