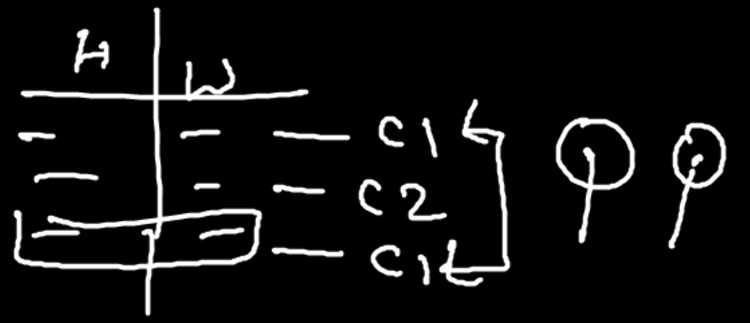
DBSCAN — Density Based Spatial Clustering of Applications with Noise

k-Means → centroid based model

Hierarchical — Agglomerative



DBSCAN — Dense regions of sparse — min(dist(- —)) = 1 clu

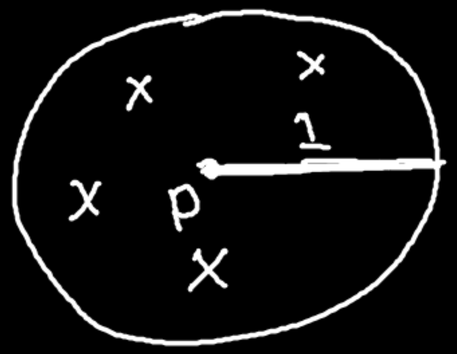
Seperate dataset into Dense region from Sparse region.

Q:- how to measure Density ?

Key Idea :- { Minpoints , Epsilon , core points , Border Point , noise point }

Minpoints & Eps — Density(Hyperparam)

(1) Density at a "p" : # points within a Circle / hypersphere / sphere of radius (Eps) around p .



Eps / radius = 1 - fixed hyperplan

MinPts = 4 - fixed - hyperplan

↳ Density Cluster

* Core Point :- if p ≥ MinPts in an Eps / radius around it

p = cluster

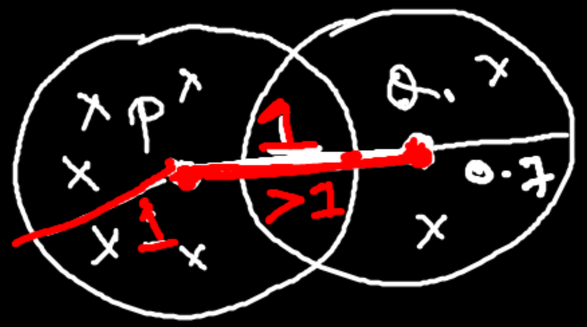
MinPts = 5

Eps = 1

p ≥ minPts

p = Core Points → Dense region

Border / Boundry Point



2.(ii) Q ≥ minpts & eps < 1 or
Q < mm & eps = 1,
Q ∈ Neighbour (P)
dist(P, Q) ≤ eps - border points

(1) P is core point, P ≥ Minpts & radius (eps) = 1

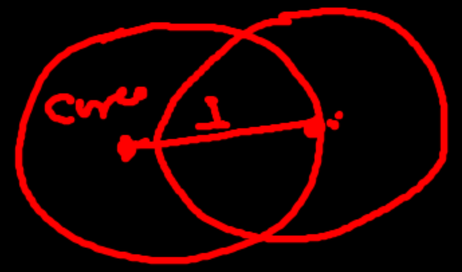
2.(i) Q < Minpts & eps < 1, also Q ∈ Neighborhood (P)
P = core point
dist(P, Q) ≤ eps - Border

* Minpts & Eps doesn't match both the hyperparam
* Either one of them (minpts/Eps) does not match it

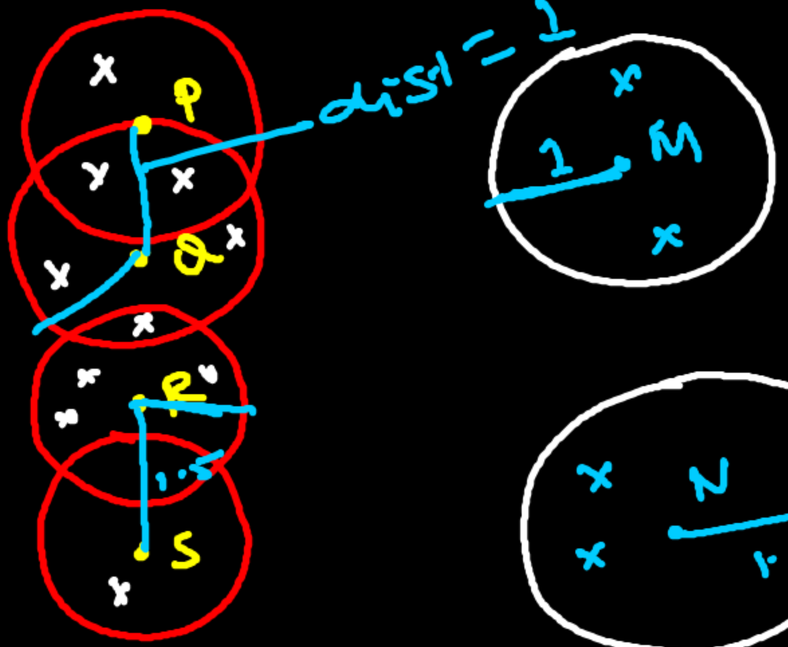
→ Q is neighbour of P — Core points

dist (P, Q) ≤ Eps

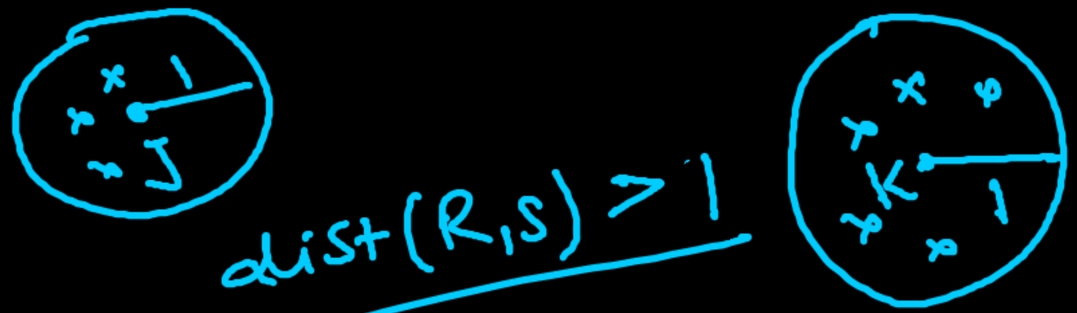
↳ Q = Border points



(3) Noise Point → Neither core point nor Border point



	Eps	Minpts	conclusion
P	1	3	Border Points
Q	1	5	Core Points
R	1	4	Core Points
S	1.5	1	Noise
M	1	2	Noise
N	1.5	2	Noise
K	1	5	Core p



NOTE :- remove all noise points from your data

↳ don't belong to any cluster

Case Study -

Association Rule Mining / Market Basket Analytics

Transaction

items - Products

		OJ	soda	milk	WC	Dete
T1	OJ	4	2	1	1	2
T2	soda	2	3	0	1	1
T3	milk	1	0	1	1	0
T4	WC	1	1	1	2	0
T5	Dete	2	1	0	0	2

Custer	
1	OJ, soda
2	milk, OJ, WC
3	OJ, Det
4	OJ, Det, Soda
5	WC, Soda

LHS → RHS

OJ → Soda

Support :- it is all about finding out the most frequent combination that are there in data Support (OJ → Soda)

Rule :- $\frac{P(OJ \cap Soda)}{\#\ Transaction} = \frac{2}{5} = 40\%$ Support

confidence :- What is the probability of RHS happening given LHS has already happened.

OJ → Soda $\quad \frac{P(OJ \cap Soda)}{P(OJ)} = \frac{2}{4} = \frac{50\%}{1}$

Rule: Lift :- Even after having acceptable occurance (high support) & acceptable confidence, it can happen that a particular combination shows occurance just because of random chance. we have to see if they are really correlated or the cooccurance is just random by chance.

pow lift
Funct - Fun
Fear
BOX-COX
0-0
400 - Jun

$$\frac{\frac{P(OJ \cap soda)}{\text{Total No. of Tran}}}{\frac{P(OJ)}{\# Tran} * \frac{P(soda)}{\# Trans}} = \frac{\frac{2}{5}}{\frac{4}{5} \times \frac{3}{5}}$$

$$\frac{2}{5} \times \frac{5}{4} \times \frac{5}{3} = \frac{5}{6} = 0.83$$

A - 2000
B - 1800
C - 2000

Rule:- Lift > 1 - Offer