

Unsupervised ML

1. K-means

2. Hierarchical

3. DBSCAN

Silhouette Scoring

+ Practical

{ PCA

+

Flash

dB

File handling

Git

}

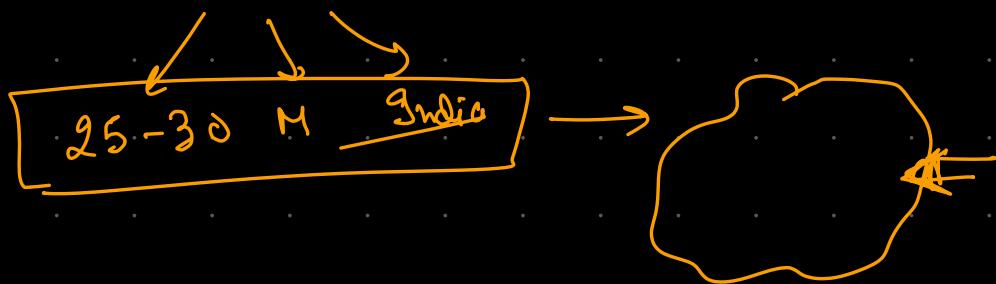
+

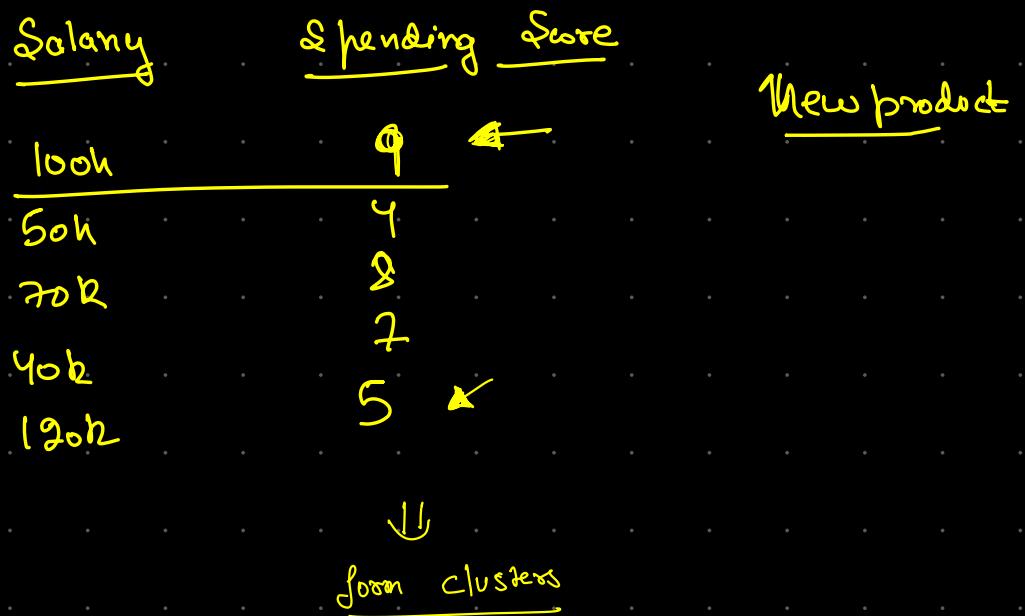
Lots of projects

Unsupervised ML



Instagram \Rightarrow Sharks / Extreme \Rightarrow Really Bad





Netflix / Prime



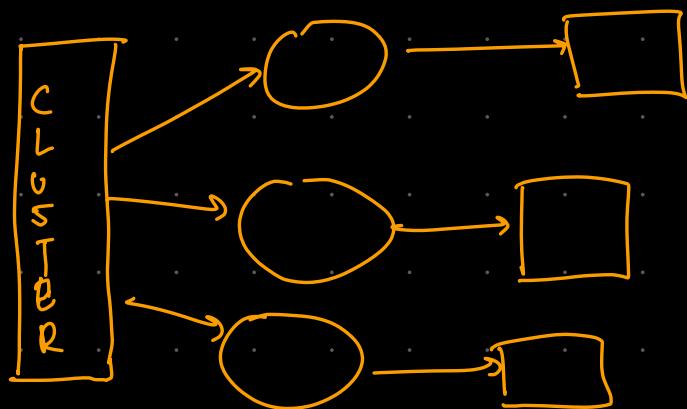
We can make algorithm specifically catered to our clusters.

USA/Can \vdash XGBoost

Europe \vdash Logistic

India \vdash SVM

Africa \vdash Decision



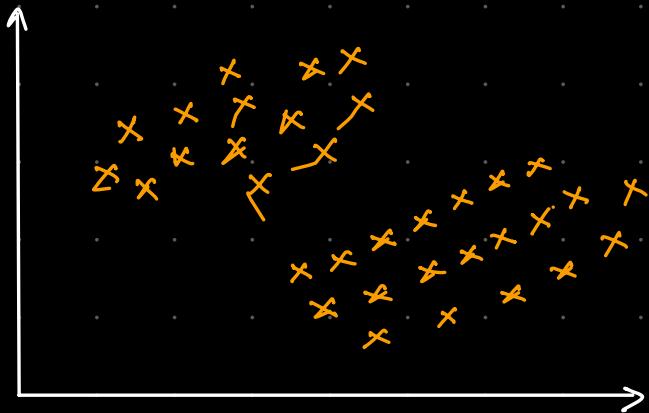
Custom ensemble technique

Salary

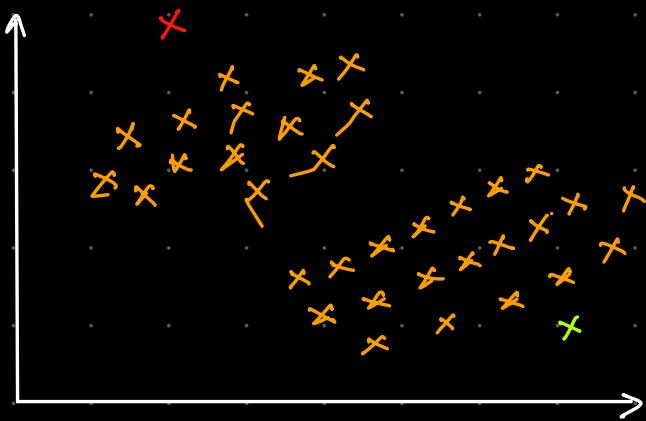
Spending

Score of discount

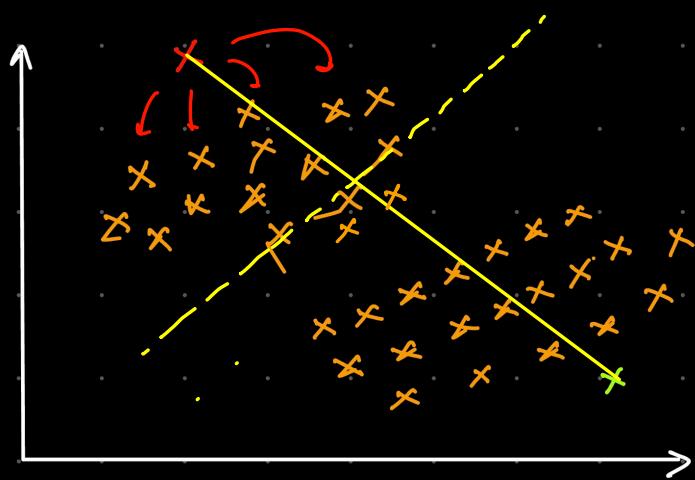
K-means clustering

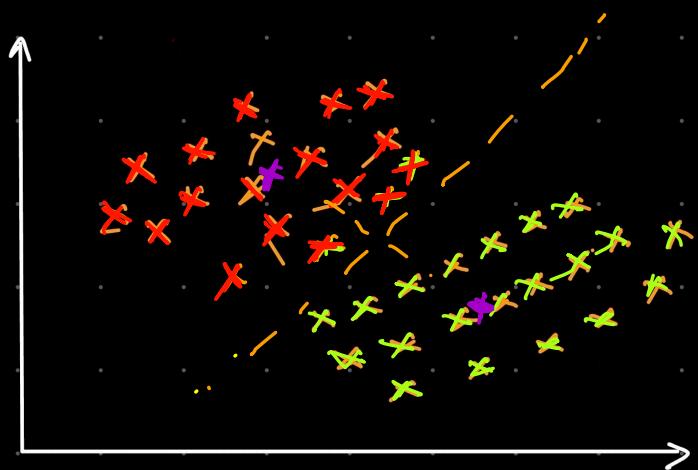
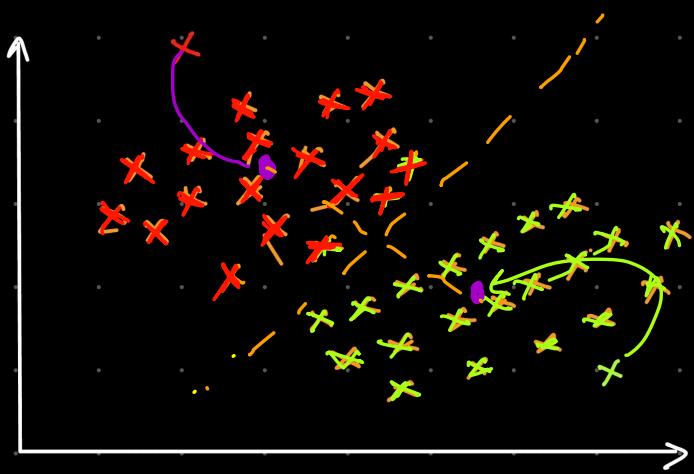


1st Step : we are going to take K centroids



2.





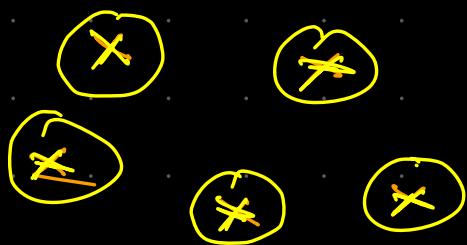
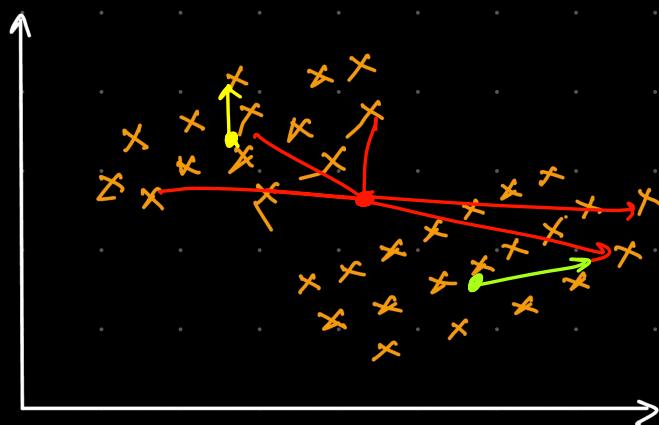
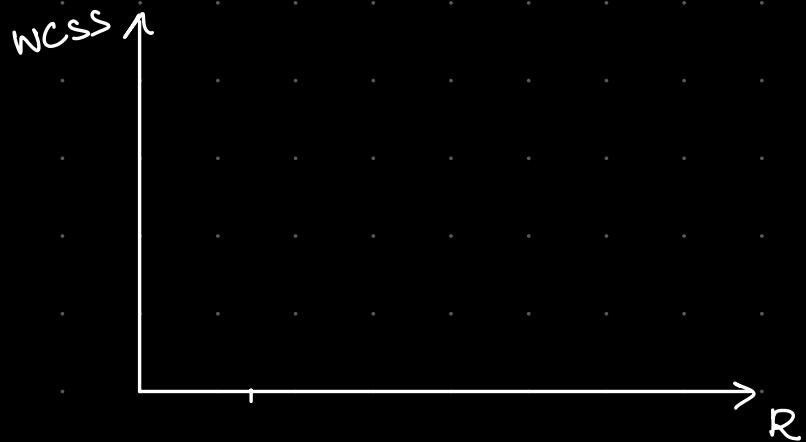
 *

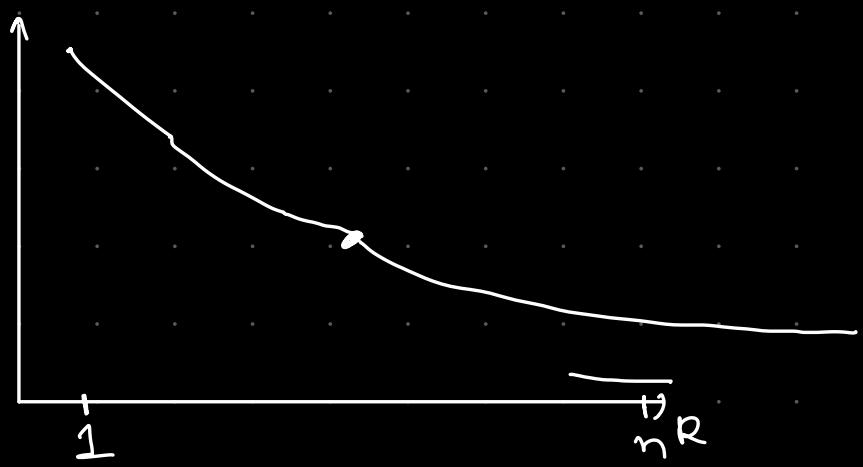
 *

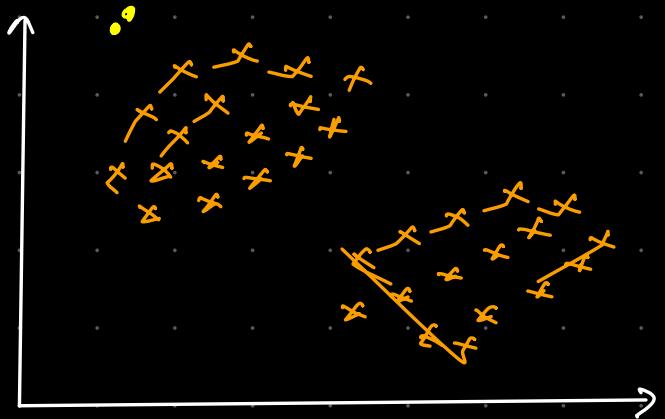
 *

 *

Elbow method

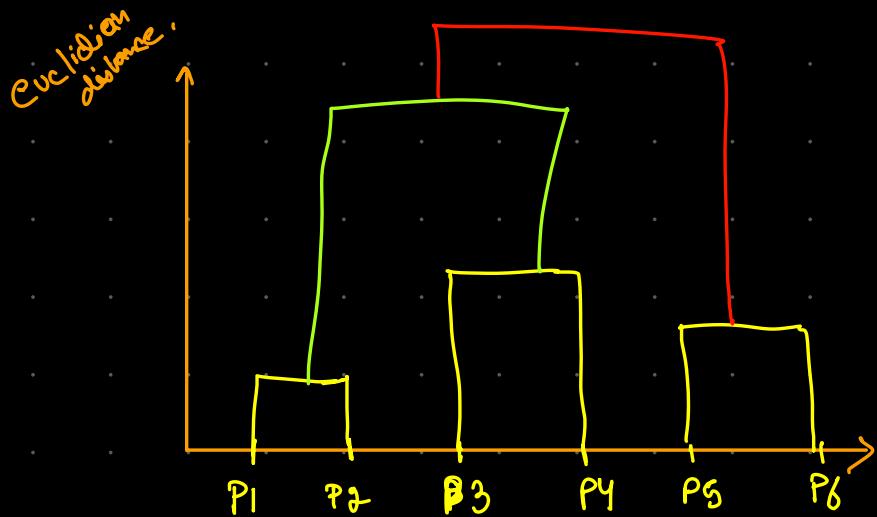
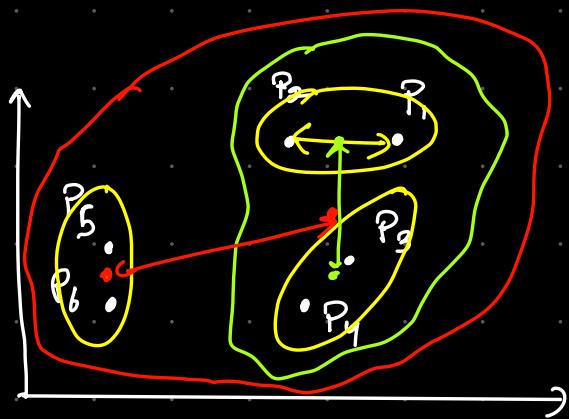


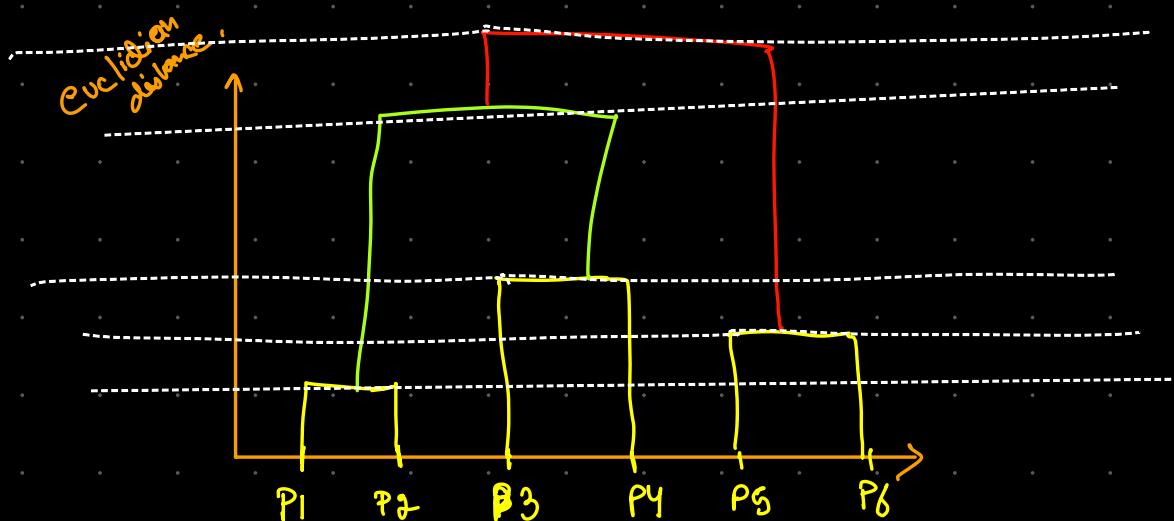




Rmeans ++

Hierarchical Clustering





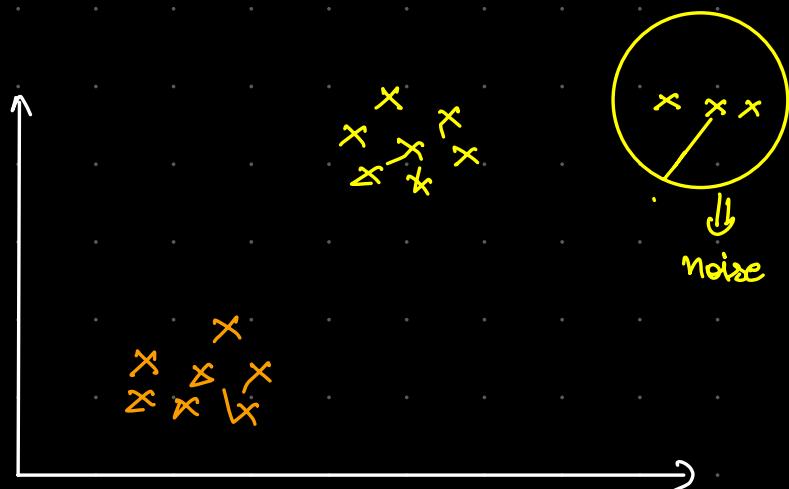
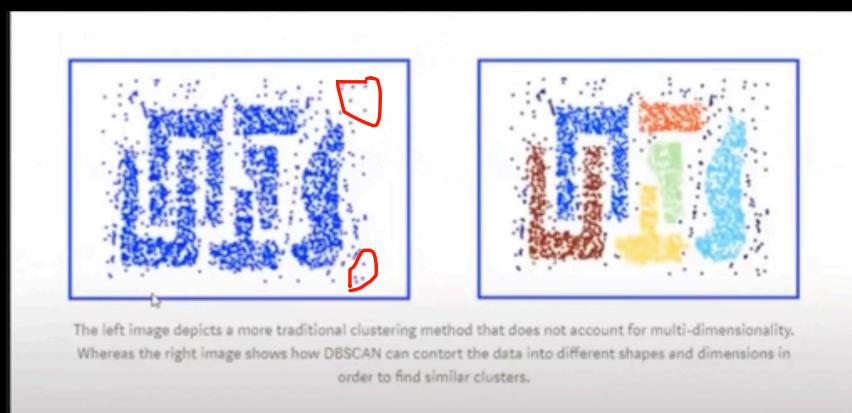
Number of clusters?

Try to cut the longest vertical line such
that no horizontal line passes through

DB SCAN

density based, spatial clustering of application
with noise

What is the problem with R-means clustering?



Some constraints / hyperparameters.

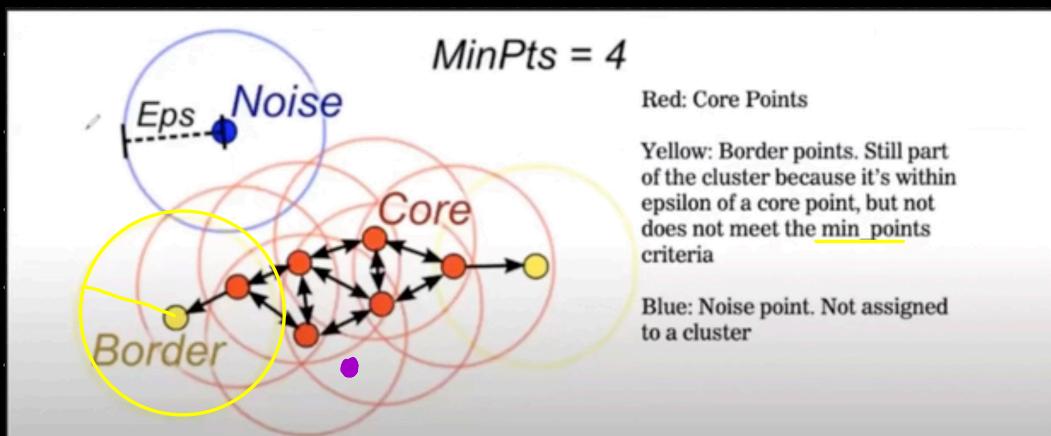
1. \in

2. min^m point

Red

Yellow

Blue



1st step select a random point & draw a ϵ radius circle

2nd step atleast min_points are inside our circle than we call this point red / core.

3rd move to next point ∵

2 condn

1. Points inside \leq min-hulls
2. Either of this point should
be a core point
(lies in epsilon of core
point)

4th step

Blue \Rightarrow noise points,

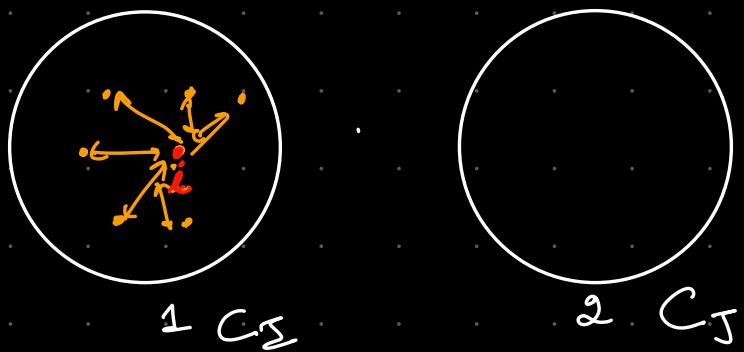
not assigned
to any cluster

Silhouette scoring

Is a method of validation of points in data

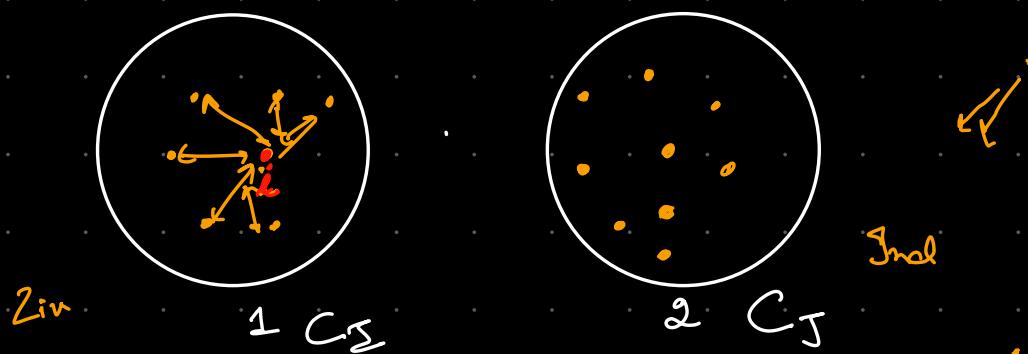
→ SS value is a measure of how similar
an object is to its own cluster
compared to other clusters.

$$-1 \leq \text{SS} \leq 1$$



for a data point $i \in C_\Sigma$,

$$\alpha(i) = \frac{1}{|C_\Sigma| - 1} \sum_{\substack{i \neq j \\ j \in C_\Sigma}} d(i, j)$$



$$\beta(i) = \min_{j \neq i} \frac{1}{|C_J|} \sum_{j \in C_J} d(i, j)$$

$$\delta(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, |C_i| > 1$$

$$\delta(i) = \begin{cases} 1 - a/b & b > a \\ 0 & a = b \\ \frac{b}{a} - 1 & a > b \end{cases}$$