Day-6

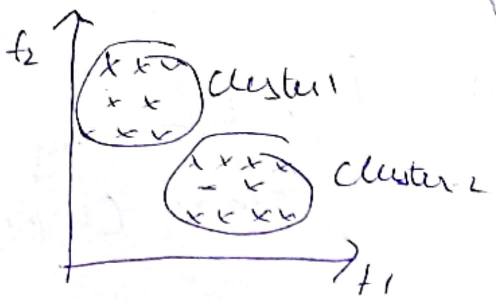
Agenda

1) K-Means cluster
2) Hicarchial clustering
3) Silhouste Score
4) DBScan clustiring

Unsupervised ml :- it doesn't have specific O/P

O/P   f1   f2

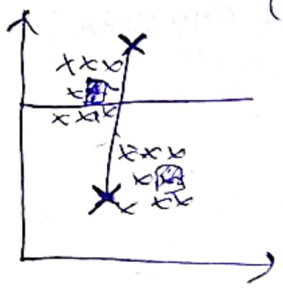
{ Similar kind of O/P }   -   -



Custom Ensemble technique

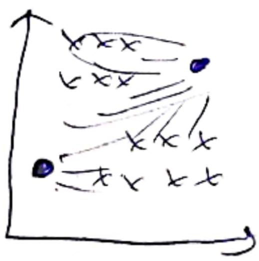
DSet → Clustering → (2 (or) 3 group) → for Each

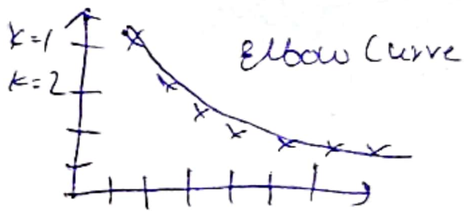
K-means → k = Centriods
(Diagram Chusi k value Chepalemu 30 C ways to solve)



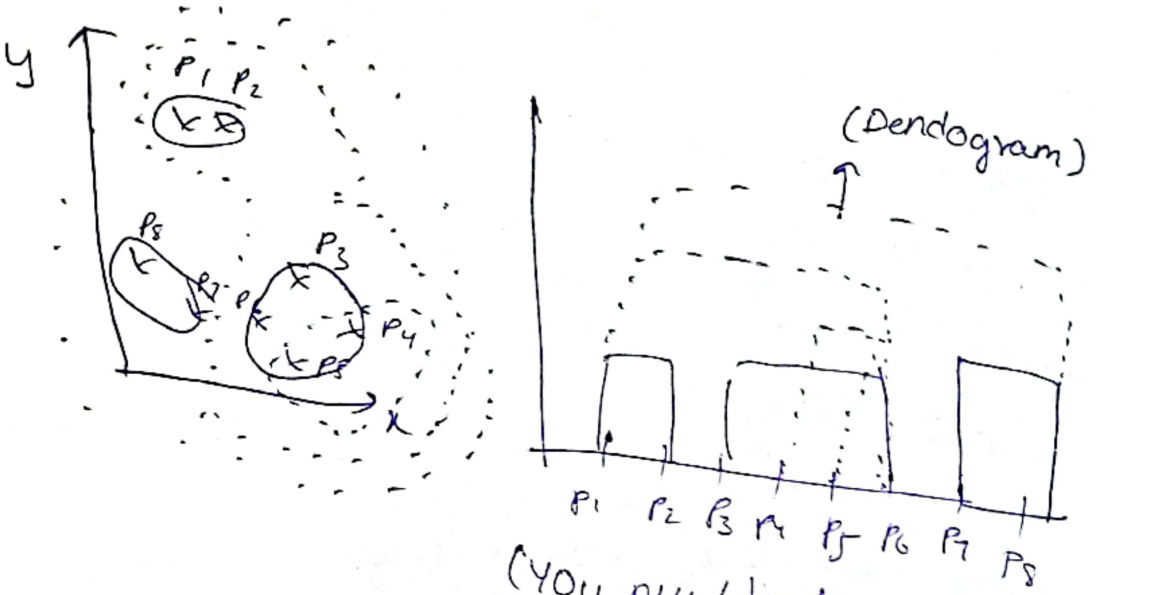
k=2
i) we try with different k-Values
ii) we intilize k no-of Centriods
(By shortest Dist, Euclidean Dist)
iii) Compute - the avg to Update Centriod

Elbow method (K - value)
We use iteration for k-values
for i=1, 10





③ Hierarchial clustering



(You need to find the longest Vertical line that has no horizontal line passed through it)

Main time taken by k-means (a) Hierarchial clustering?

Ans:- Hierarchial (used to find Dendograms)

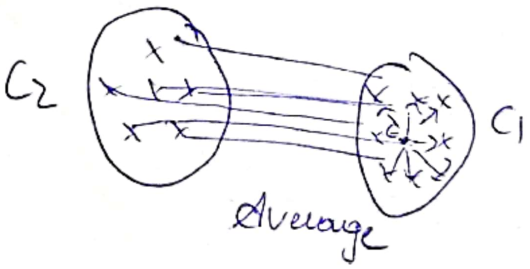
Validating Clustermodel

Silhouete Model

The data have been clustered via any technique such as k-means into k clusters

$$a(i) = \frac{1}{|C_r|-1} \sum_{j=C_r, i \neq j} d(i,j)$$

ranges:- −1 to +1



DBScan Clustering

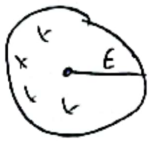
(a) Epsilon (radius of circle)

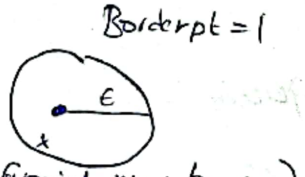
(i) Min pts

(ii) Core pts

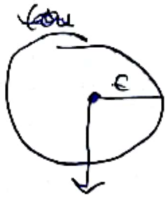
(iii) Border pts

(iv) Nase pts





(point near border)

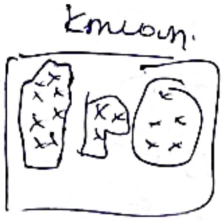


Noise pts (or) outlier

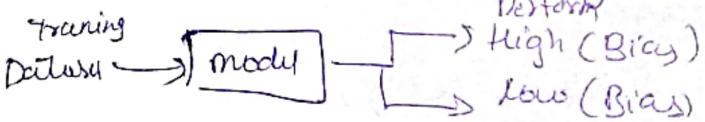
(no points found)

[DBScan >> kmean] while solving pro6 DBScan solves in the problem in different group-wise but k-mean all one solved in

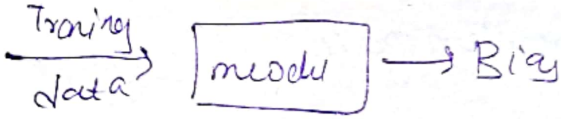
group wise

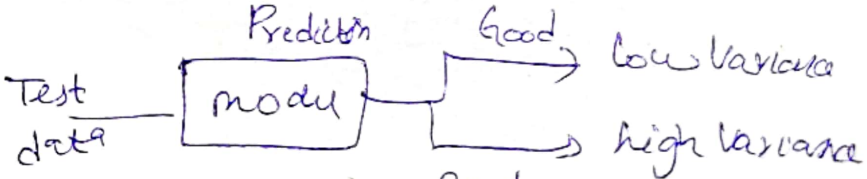


Bias:- it's a phenomenon that steers the result of a algorithm in favour or against an idea (dataset)



Variance :- refers to changes in model when using different portions of the training (or) test data





model-1
Train = 90%
Test = 71%
{ low Bias, high Variance }

model-2
Train = 60%
Test = 55%
{ high Bias, high Variance }

model-3
Train = 90%
Test = 92%
{ low Bias, low Variance }