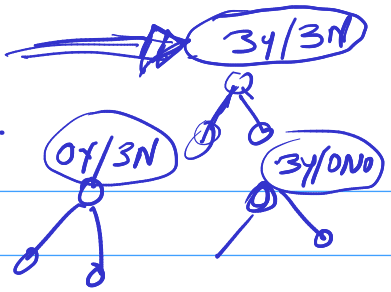
DT — Yes-No check for Decision

↳ overfitting → Pre-pruning & Post pruning.
Big data + small data.



Weak learner →

Ensemble technique

(Grouping) → weak learner

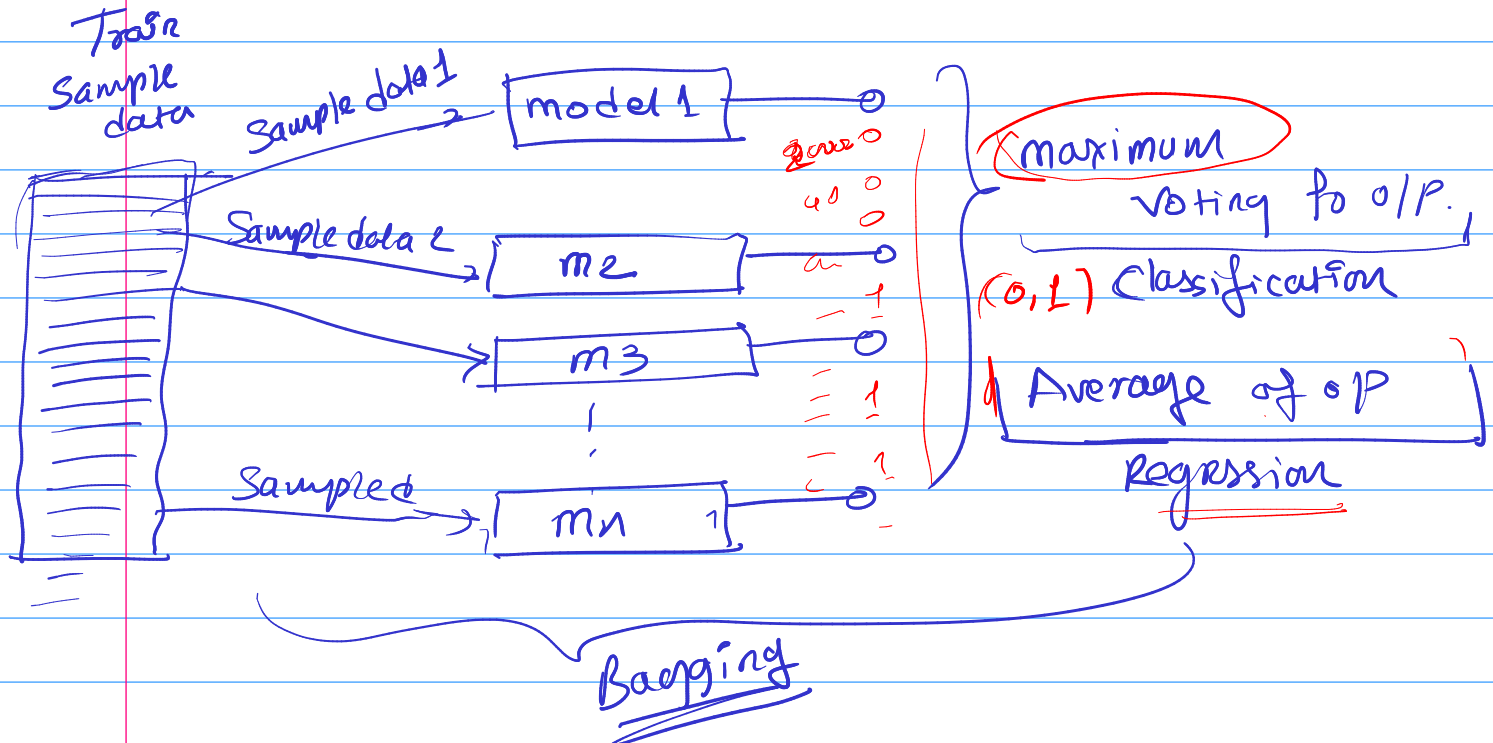
Bagging

- Random forest

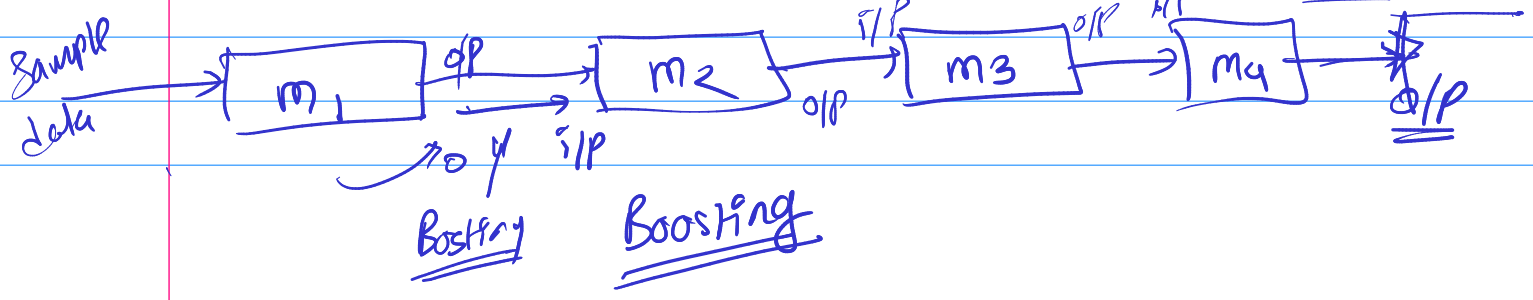
Boosting

- Adaboost
- Gradient Boost
- xg-boost
- CatBoost

Decision Tree - overfitting → ensemble technique.



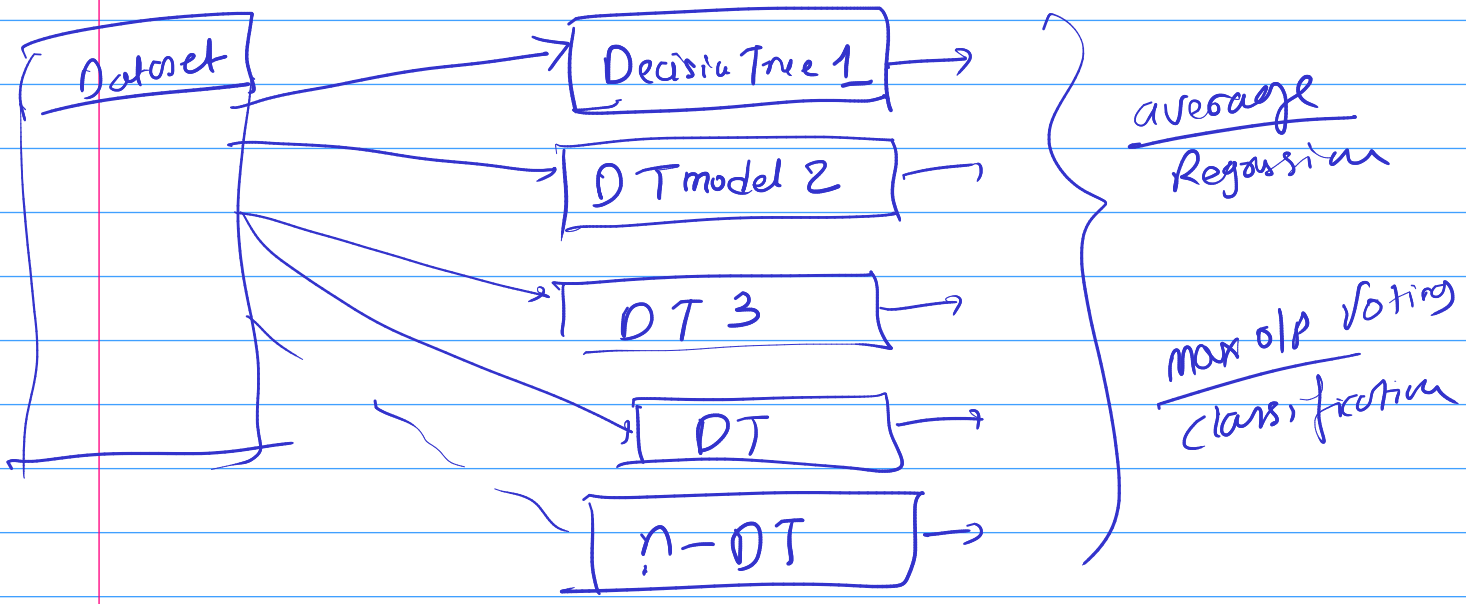
Bagging



Bosting

Boosting

Bagging → Random forest

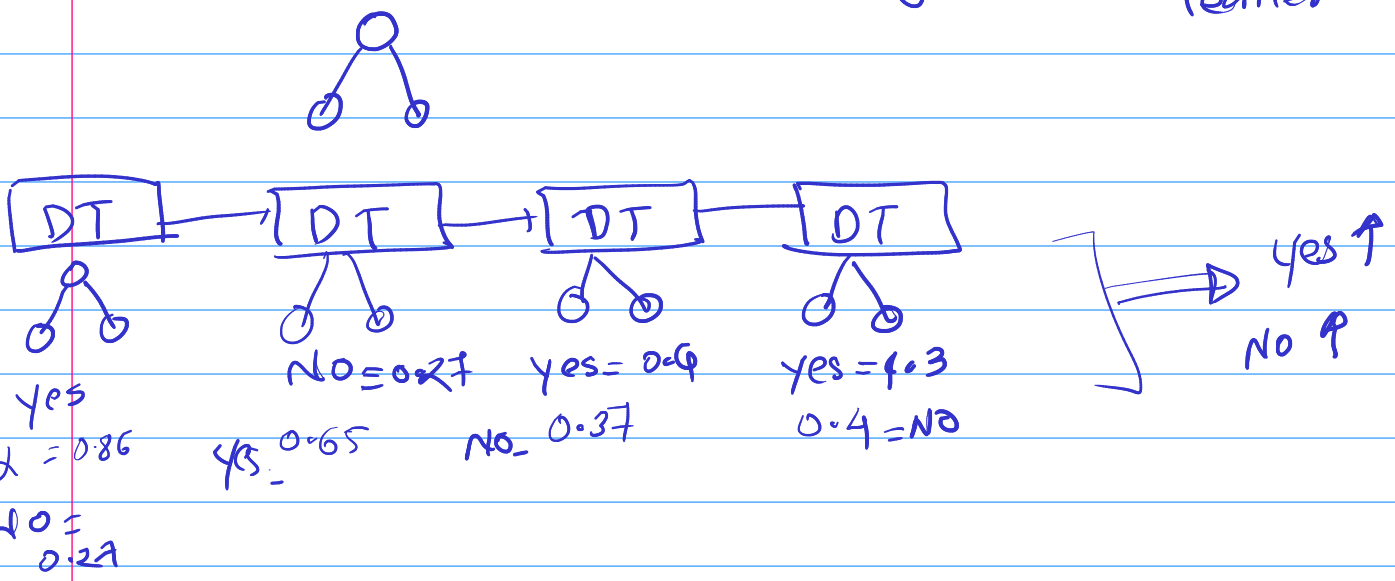


Hyperameter = (n-estimator)

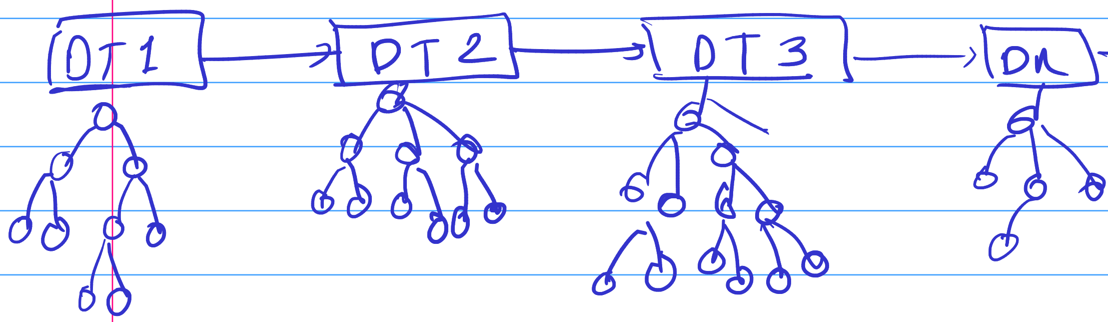
Boosting

Adaboost (classification)

DT (Decision tree) → depth only 1 level i.e. weak learner



Gradient Boosting → class - Reg (max depth)



(all output average)
Reg
class
voting

Supervised Machine Learning

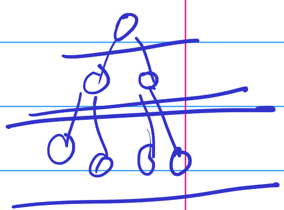
Random for - n estimator = (10), 8
= [2, 4, 5, 6, 8, 10, 12, 20]

GridSearchCV = Best Params

Ridge & lass = λ → error
Parameter = (alpha = 0.1, 0.2, 0.01, 0.02 ...)

Best pa

alpha
elastic alpha =



pree-prun post

dT = (max_depth) = ...

rad - max-depth, n-esti...

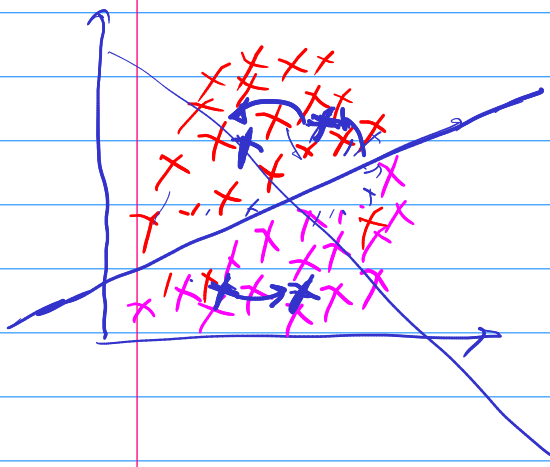
Unsupervised (No Lable data)

↳ i/p → finding the pattern → clustering

↳ ① k-mean clustring
② Hirencical clustering
③ DBSCAN

---

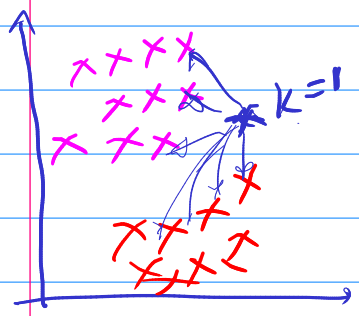
① k-mean clustring

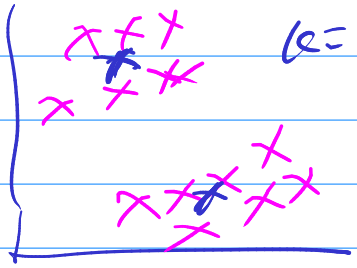


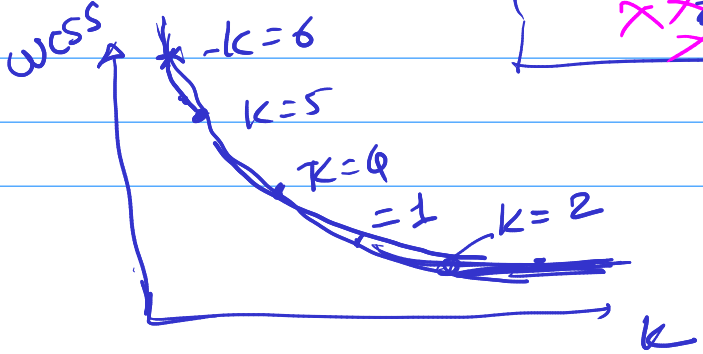
① k-centroid initialize (k=2)
② points that are nearest we need to group.
③ more centroid by calculating mean
④ repeat step 2 & 3 again & again
⑤ final centroid is in middle

k ? = within cluster sum of square (WCSS)

$$WCSS = \sum_{i=1}^{k} (\text{distance bet}^{n} \text{ point to centroid})^2$$

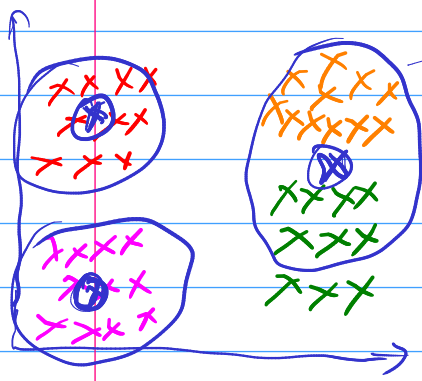






k=

Hirarical Clustring



small data = Trap - K value

K = centroid

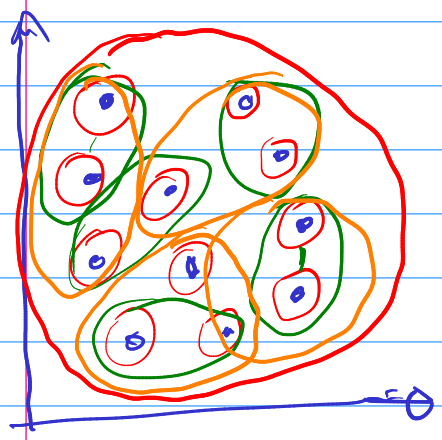
No centroid

Hiranical clustring

① Agglomerative (Combine करना)

② Divisive (divide करना)

No centroid



Step1 → for each point consider as a seperate cluster

Step2 → find nearst point & create new cluster

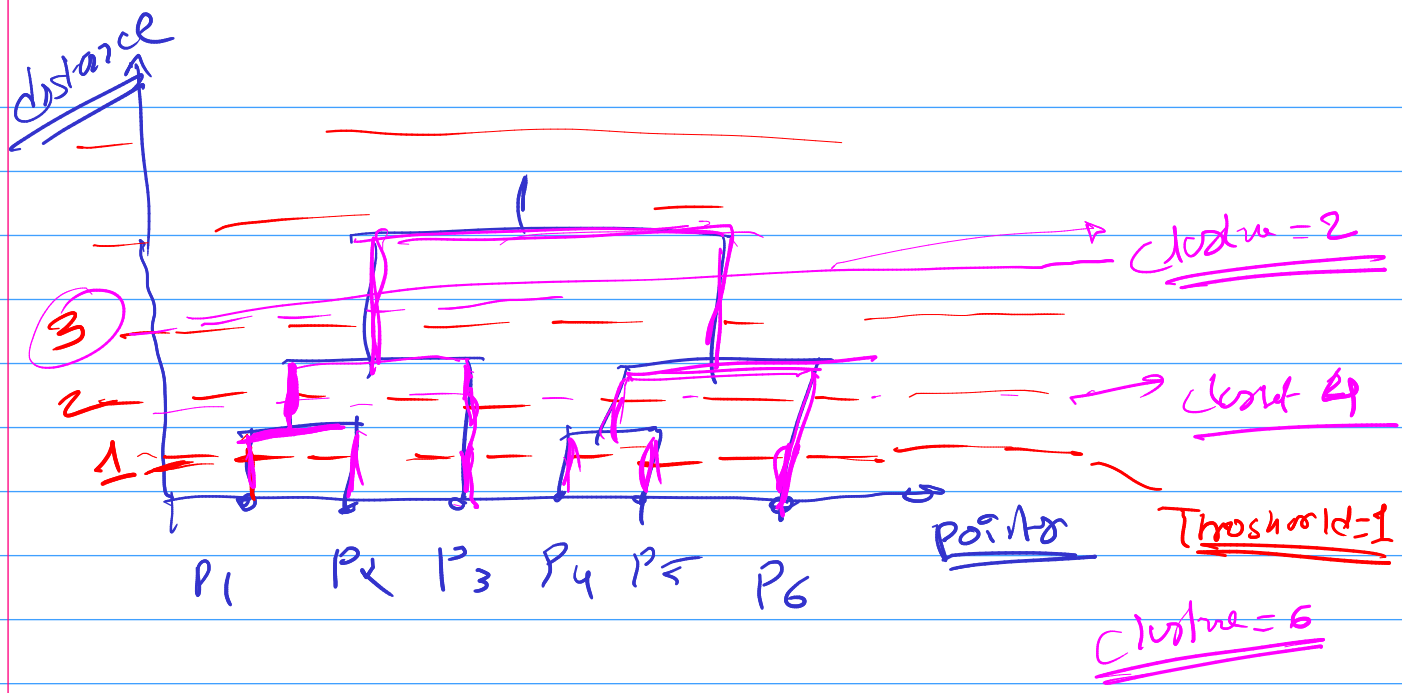
Step 3 → keep doing Step 2 until we get single cluster

for 6 point = 6 clusters → 5 clust → 4 clust → 3 clust → 2 clust → 1 cluster

6 point ⟶ 1 point → Agglomerative combing

Just opposite 1 → ---6 cluster = Divisive

How many clust? in Hirarical ⇒ Dendogram



Threeshold Value Consider

Select longest vertical line such that none of Horizontal pass through it.

Hyperparameter = Threshold

DBSCAN → Density Based clustering application with noise