- ~~Decision Trees~~
- ~~Random Forest~~
- Completing our project.

CART
Classification AND Regression Trees

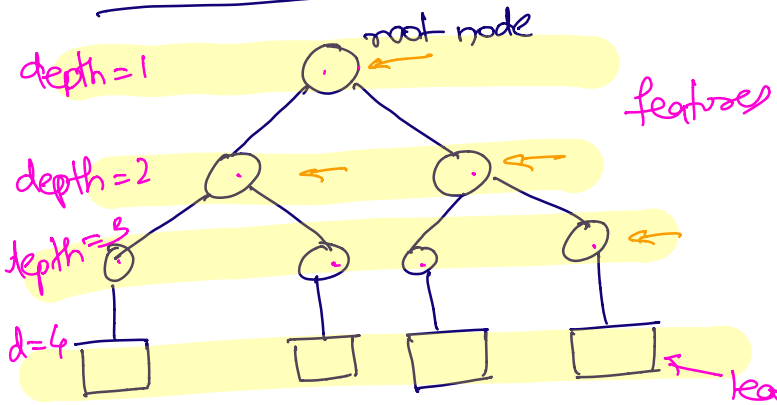
Regression Trees
features : any data
label : numeric.

Error function :
1. MSE *
2. MAE

Classification Trees
features : any data
label : categorical

Error function :
1. Gini impurity *
2. Entropy (Information gain)

Decision Trees



creation of DT depends on Gini / Entropy (IG)

Note : DT will memorize training data if depth is not controlled by user.

Randomforest

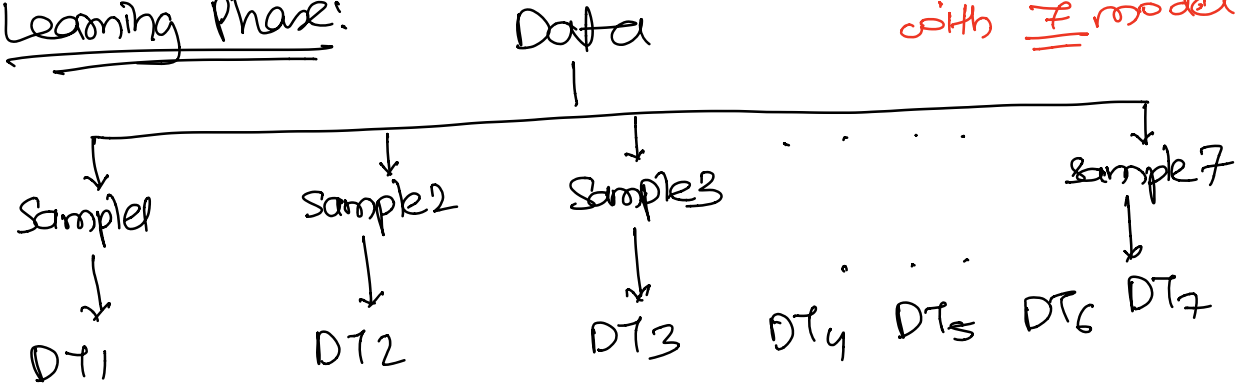
→ Ensemble learning algorithm which uses Bagging (Bootstrap aggregation)

→ In ensemble, multiple models are involved in decision making process.

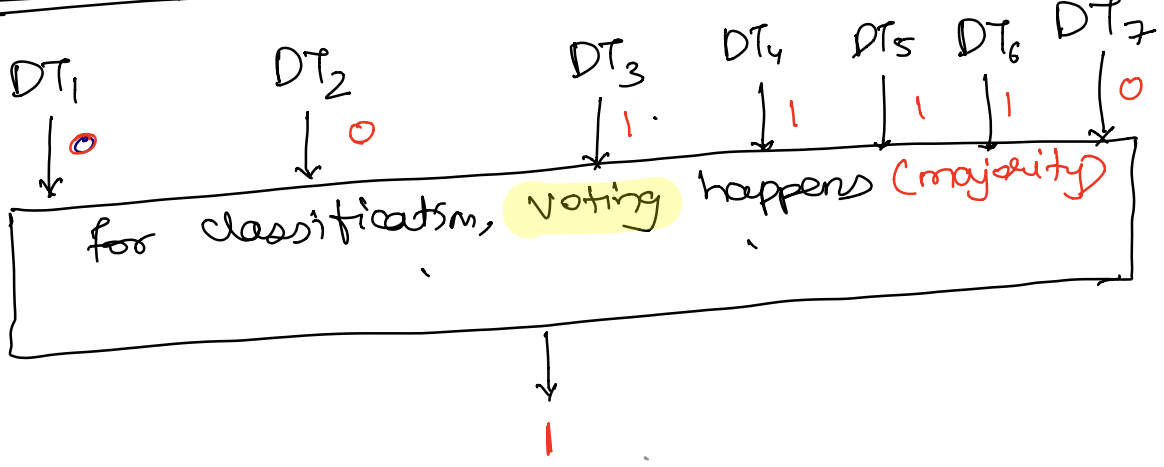
→ Each model is trained with different sample sets.

(DT)
Randomforest with $\neq$ model.

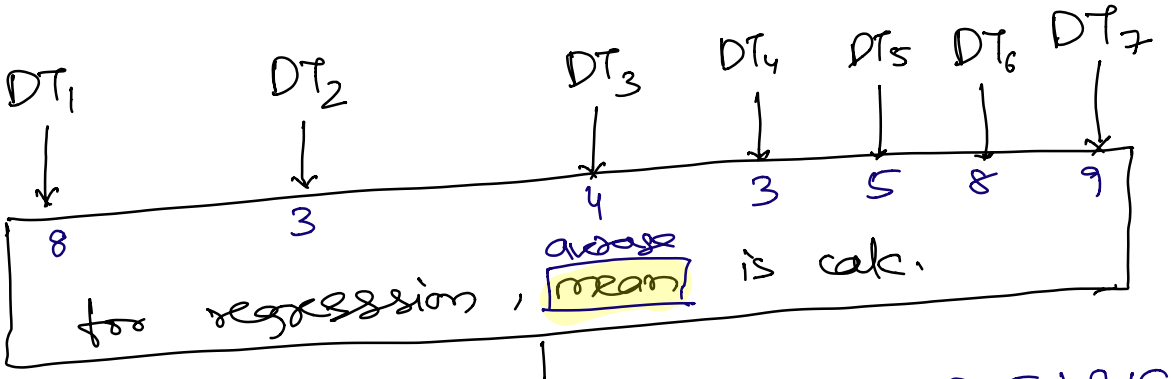
Learning Phase:



Prediction Phase: (classification)



Prediction Phase (Regression)

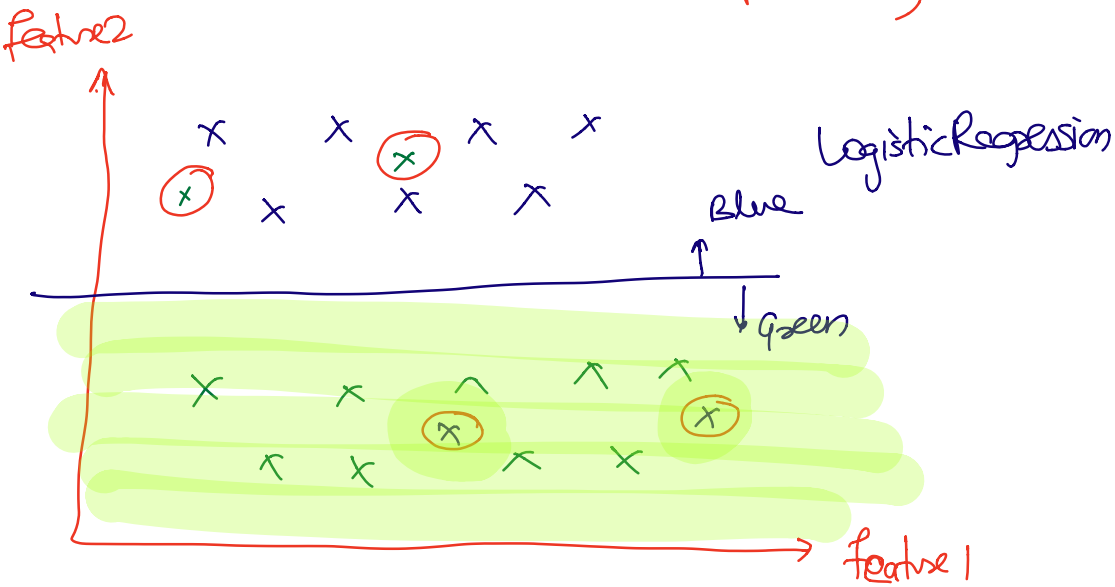


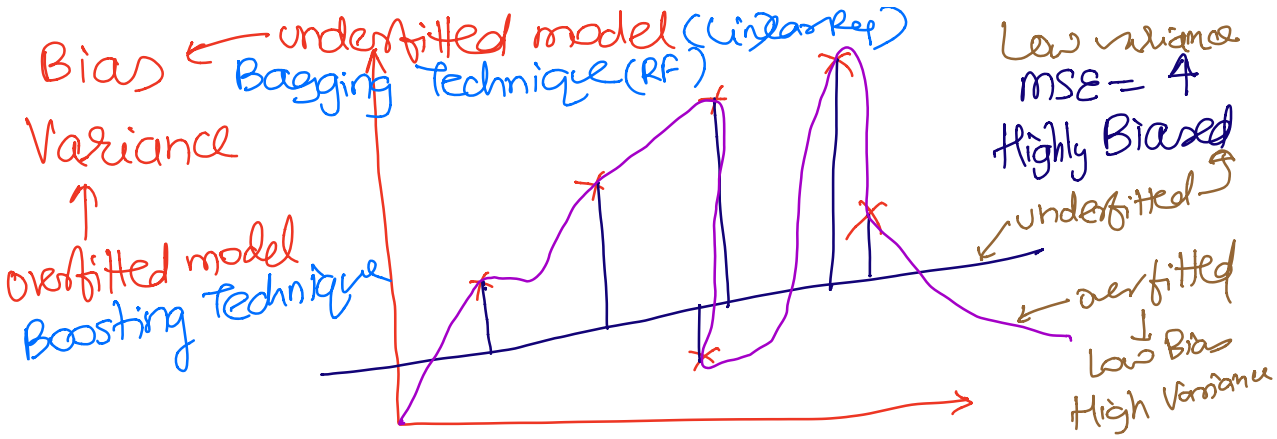
$$\text{mean} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{8+3+4+3+5+8+9}{7} = 5.71$$

n → no. of models

x → result / label

What is mis-classified data? (Classification Problem)




Bias
underfitted model (LinearReg)
Bagging technique (RF)
Low variance
MSE = 4
Highly Biased
underfitted
Variance
overfitted model
Boosting technique
overfitted
Low Bias
High Variance