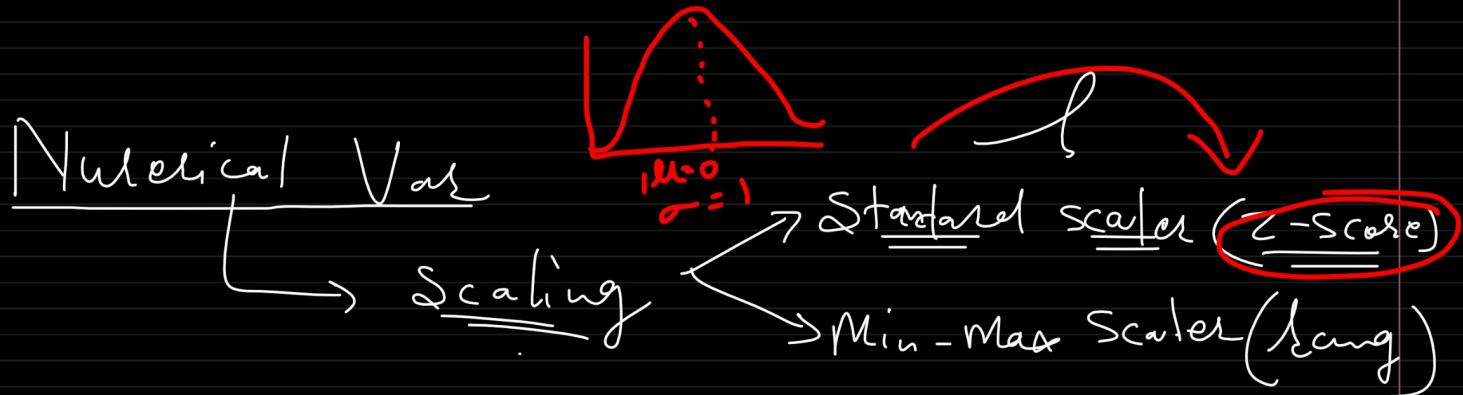


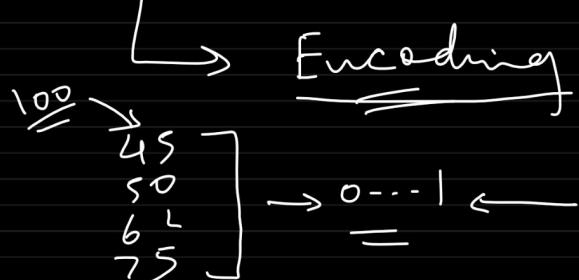
Reg

Ec

Ge



Categorical Var



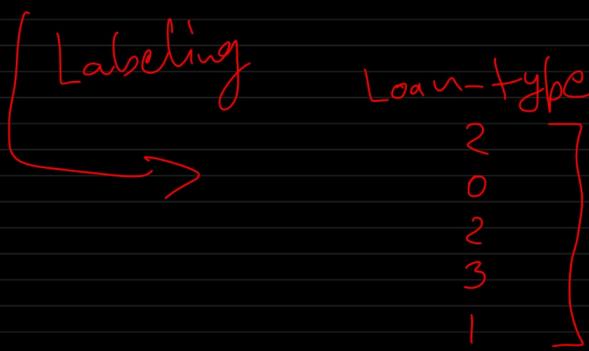
Numerical equivalent

$(\text{String}) \rightarrow Z =$

① One-hot encoding
② Labeling

One-hot encoding

Custid	Loan-type	Encoding	Custid	Loan-home	Loan-credit	Loan-person	Loan-education
1001	2 home		1001	1	0	0	0
1002	0 Credit		1002	0	1	0	0
1003	2 home		1003	1	0	0	0
1004	3 Person		1004	0	0	1	0
1005	1 Education		1005	0	0	0	1



Types of Variables

- ① Independent (S)
- ② Dependent (Target)

Year	Adverexp	Sales
2021	200	1000
2022	400	1800
2023	650	2250
2024	850	2500
2025	1200	???

M-L Types

✓ Supervised

✓ Regression Classification

UnSupervised

- Clustering
- Dim Reduction

Dep → Continuous → Regression → Predict Temp
Val 41°C , 48Lb Tomorrow?
 38°C ($35-38$)

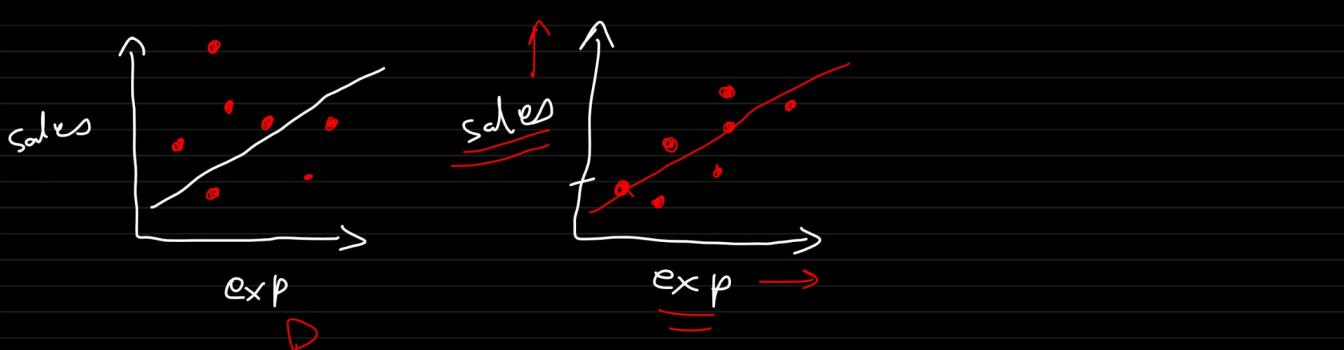
Dep → Categorical → Classification → Tomorrow will
Val True/false, M/F, Hot/Cold be hot/cold/being?

Regression

- ① Linear Regression
- ② DT & RF

Classification

- ① Logistic Reg
- ② DT & RF
- ③ NB
- ④ SVM

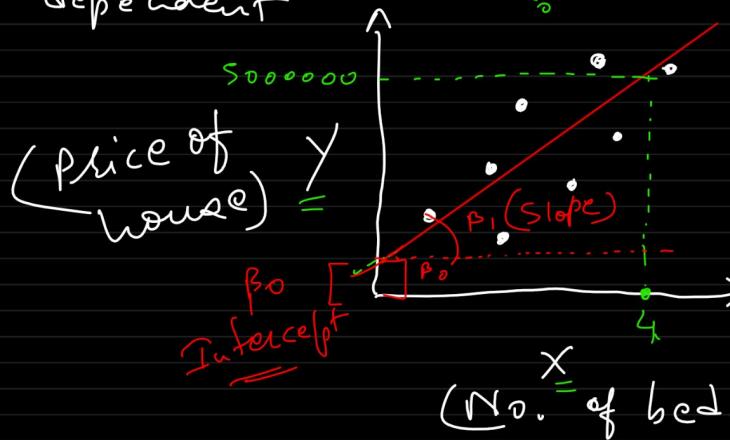


Simple LR

$$Y = \beta_0 + \beta_1 X$$

$$Y = mx + c$$

Dependent Independent



Multiple-LR

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$$



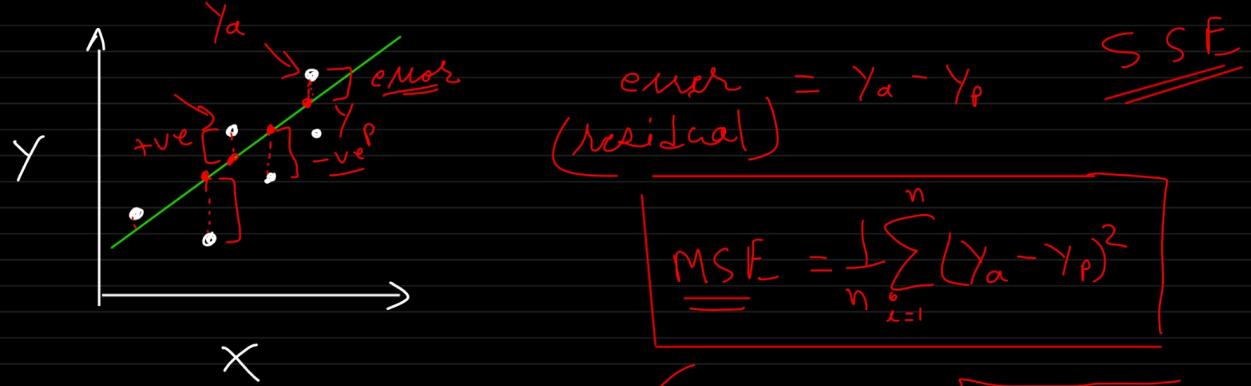
Error ? $\downarrow \text{error} = Y_{\text{actual}} - Y_{\text{predicted}}$

SSE (Sum of Squared Errors)

$\hookrightarrow \underline{\text{MSE}}$, MAE, RMSE

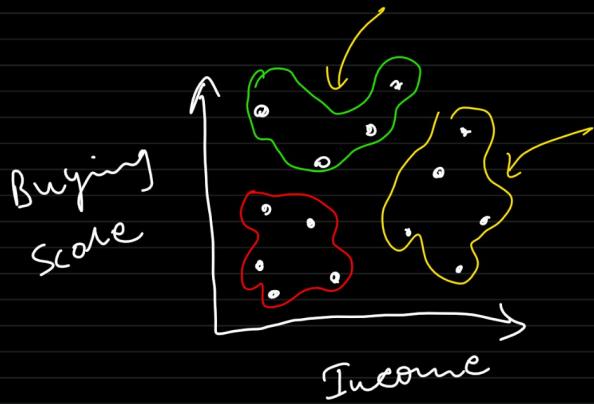
Accuracy Metrics (Performance)

$\hookrightarrow R^2$, Adj. R^2



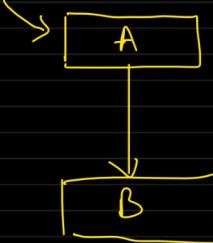
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_a - Y_p)^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_a - Y_p|$$

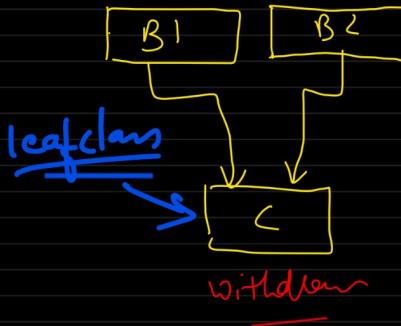


Types of Inheritance

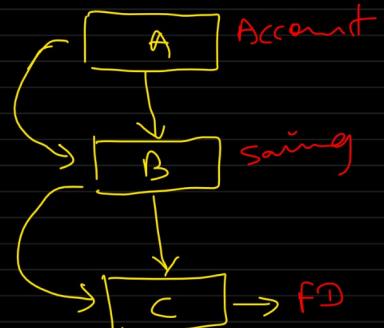
① Single Simple



② Multiple Inheritance
saving credit

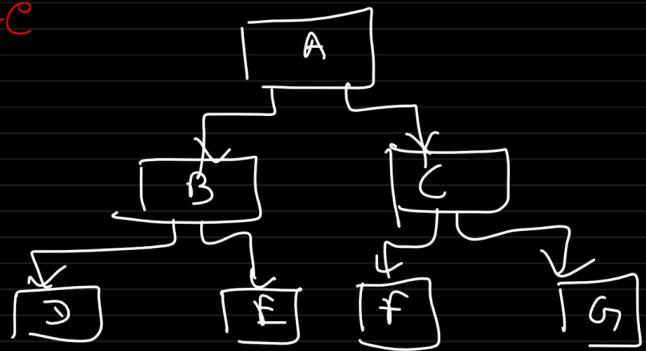
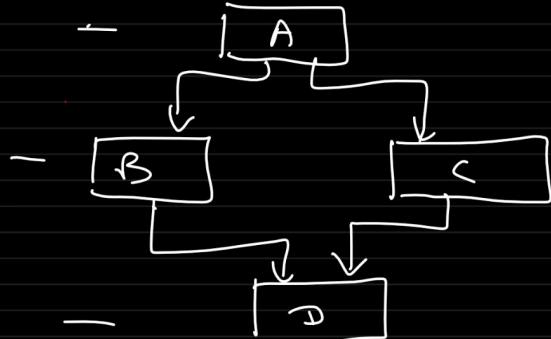


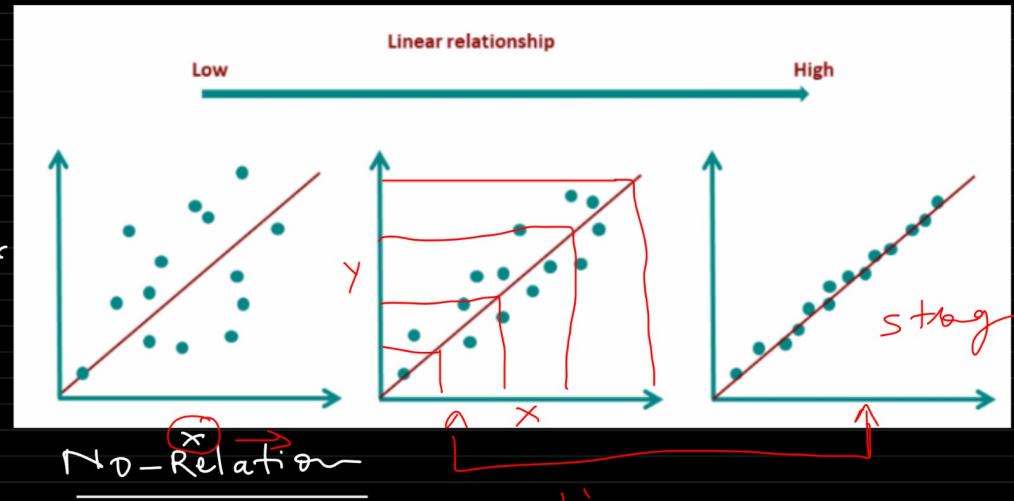
③ Multilevel Inheritance



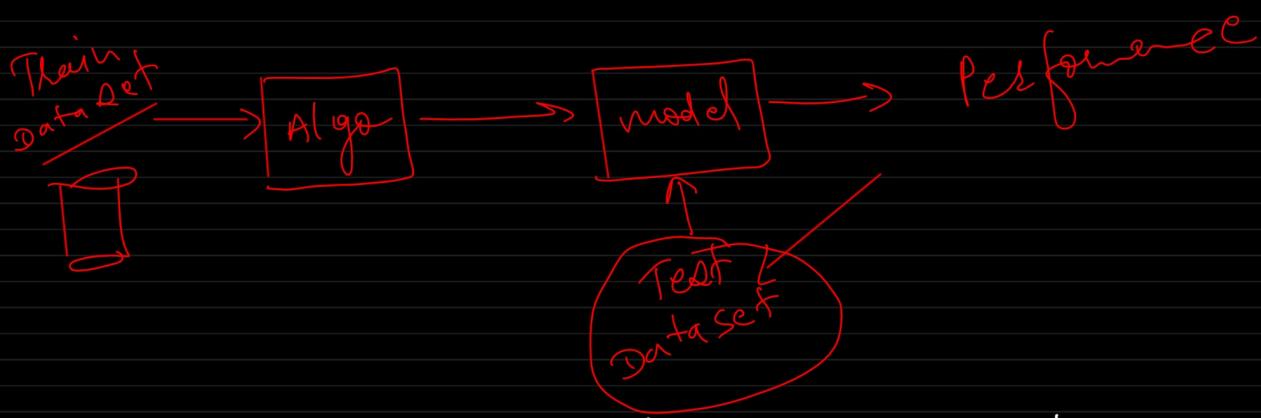
④ Hybrid Inheritance

⑤ Hierarchical Inheritance



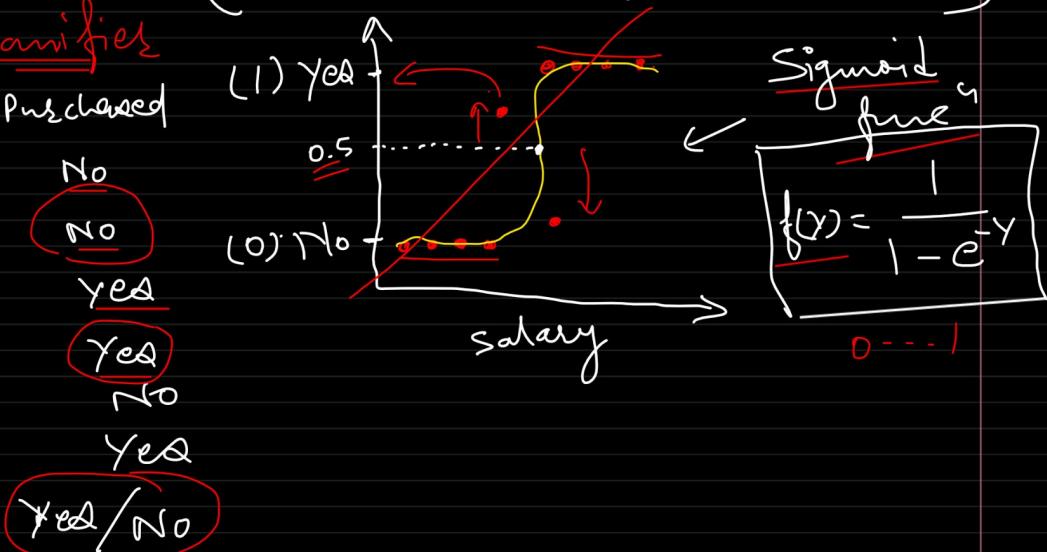


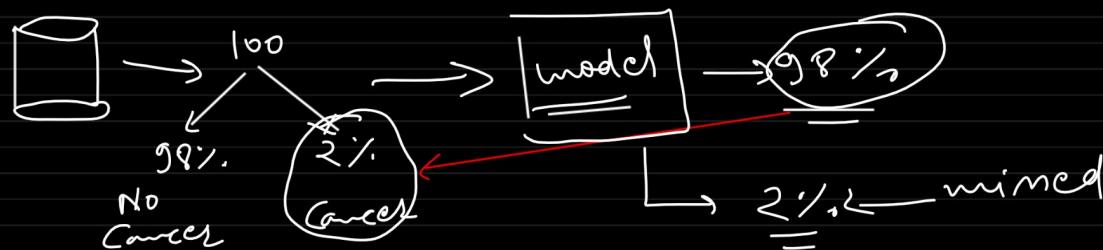
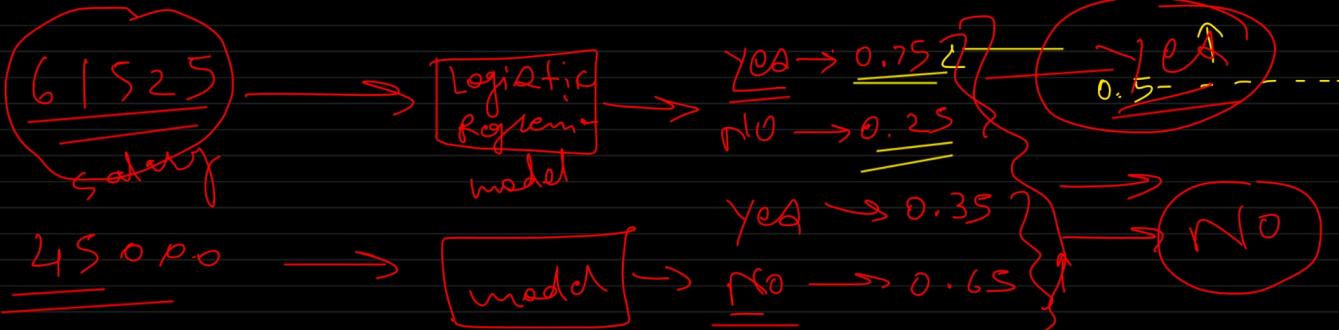
$$\rightarrow \begin{cases} 80 \\ 70 \\ 75 \\ 60 \end{cases} \begin{cases} 20 \\ -30 \\ -25 \\ -40 \end{cases}$$



[Logistic Regression (Solves classification problem)
 ↳ Probabilistic classifier]

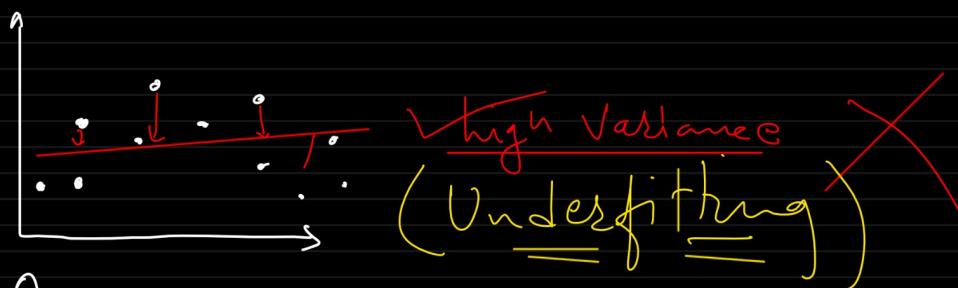
Cust_id	Salary	Purchased
1001	50000	No
1002	25000	No
1003	75000	Yes
1004	62000	Yes
1005	61500	No
1006	63000	Yes





Accuracy

Precision }
Recall } $\frac{\text{Precision} + \text{Recall}}{2}$ = f1-score

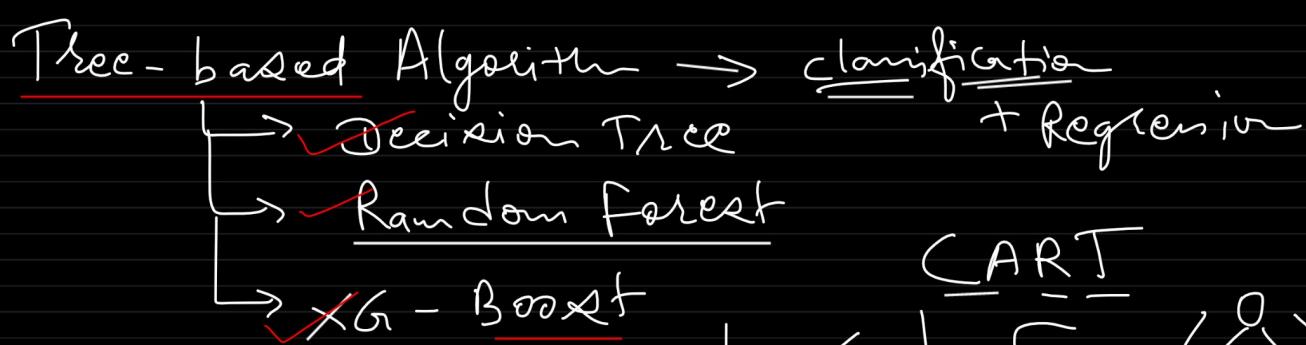


Bias - Variance Trade off

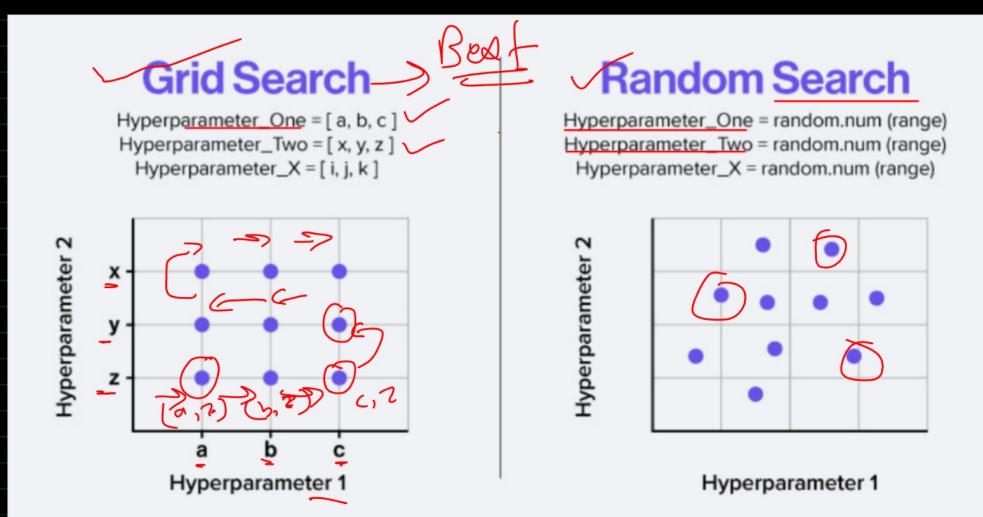
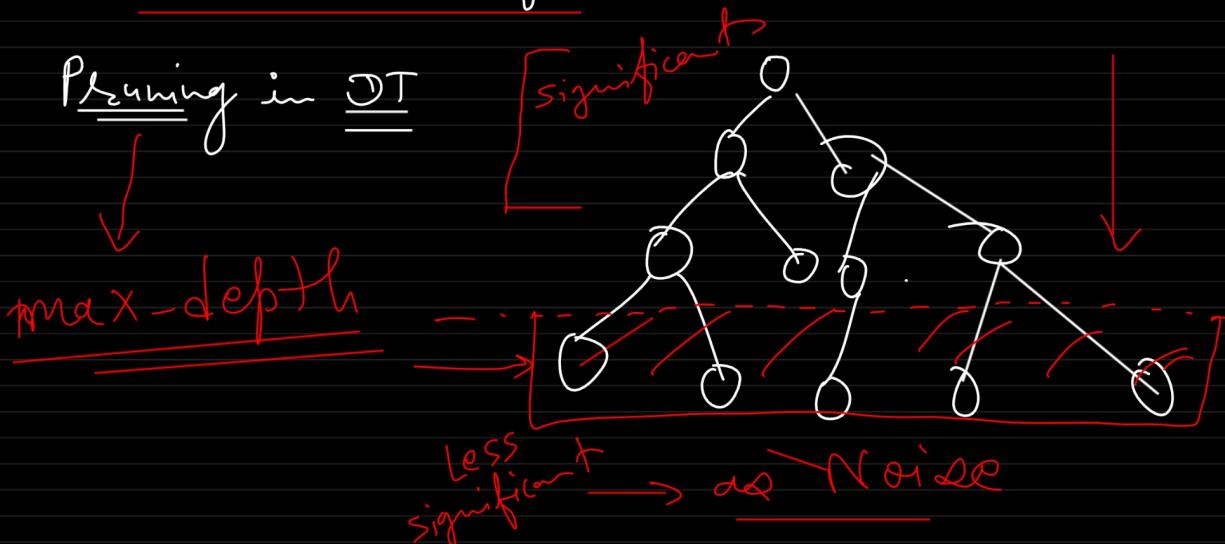
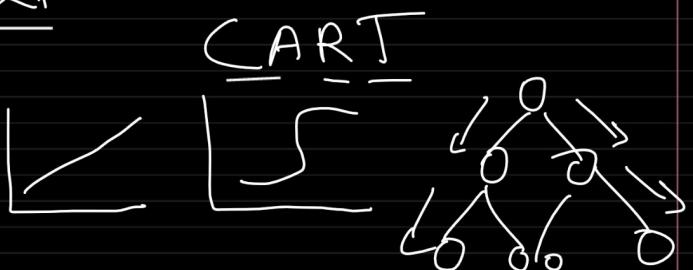
Balance
Best-fit

Overfitting

- ① hyperparameter Tuning
- ② Cross Validation
- ③ Ensemble Learning



- * Easy to Interpret
- * prone to overfit



$\left\{ \begin{array}{l} X \rightarrow 1, 3, 5 \\ Y \rightarrow a, b, c \end{array} \right.$
 model (1, b) \rightarrow best.
 model (3, c) \rightarrow best

