

Day 5 — Machine Learning Algorithms

Agenda

- ✓ ① Ensemble Techniques
 - Bagging
 - Boosting
- ✓ ② Random Forest
- ✓ ③ AdaBoost
- ④ Xgboost → Youtube channel

{ DJANGO }

FLASK

EDA

Deep Learning

Stats

ML

NLP

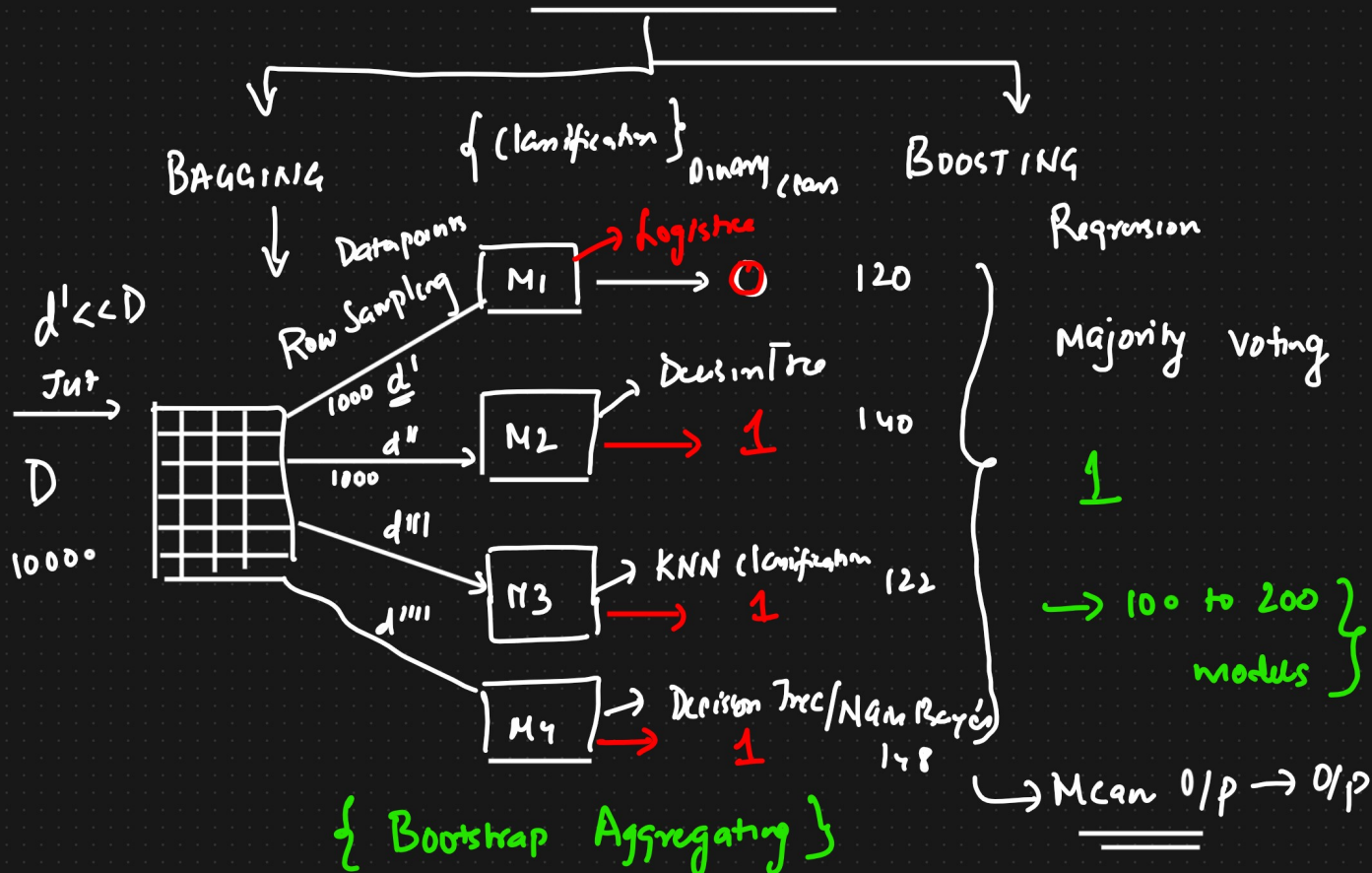
Ensemble Techniques } ✓

① Classification & Regression

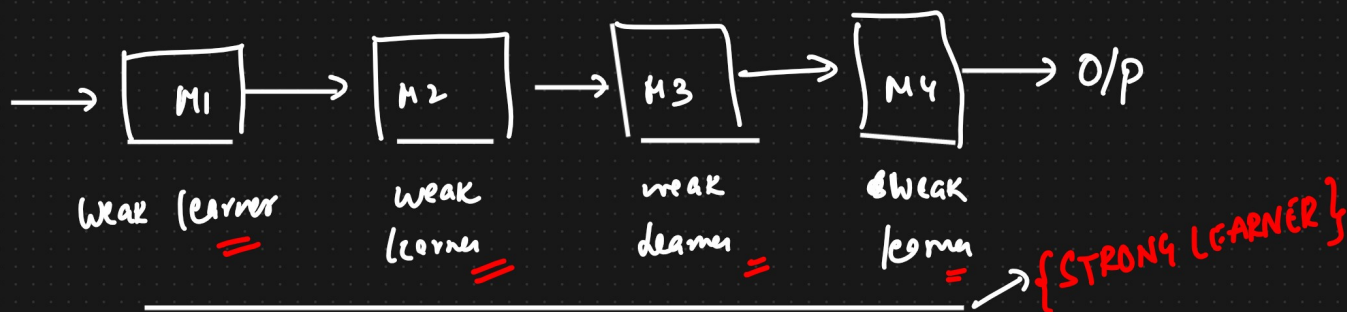
↳ 1 Algorithm $\begin{cases} \rightarrow \text{Cl} \\ \rightarrow \text{Reg} \end{cases}$

Multiple Algorithms to solve a problem?

Ensemble Techniques



Boosting



BAGGING



- ① RANDOM FOREST CLASSIFIER
- ② Random Forest Regression

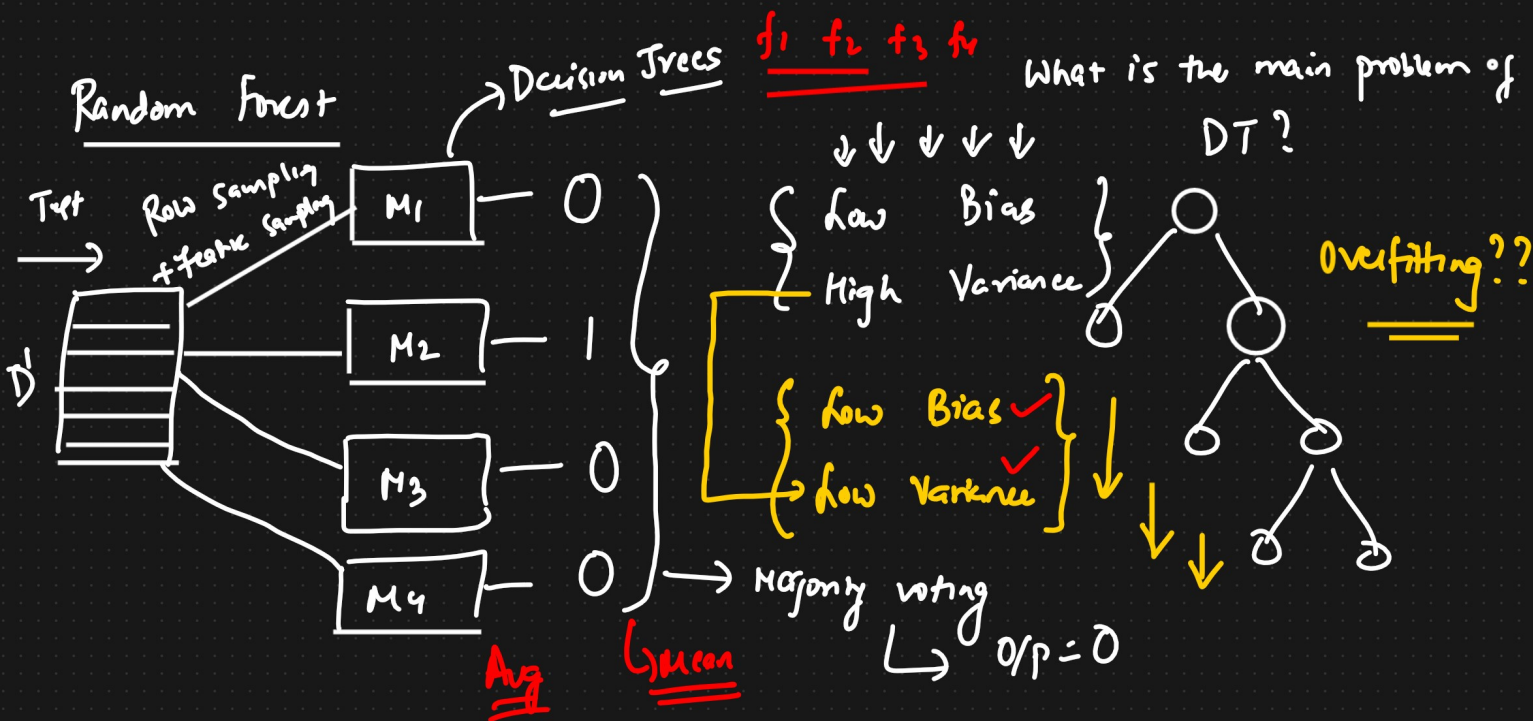
BOOSTING



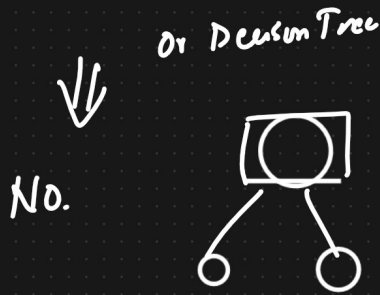
- ① Adaboost
- ② Gradient
- ③ Xgb-boost

① Random Forest classifier and Regressor

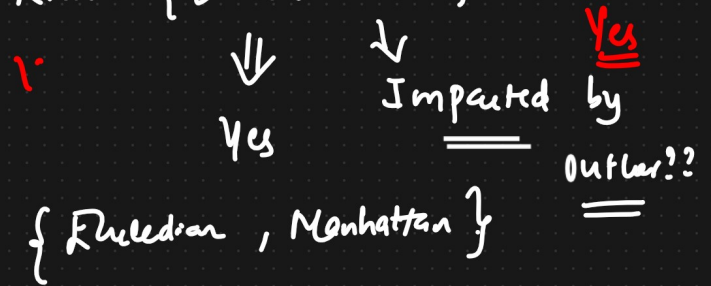
Random Forest



① Normalization ??



② KNN {Standardization}??



③ Random Forest \rightarrow Outliers \Rightarrow NO \rightarrow {check it google}

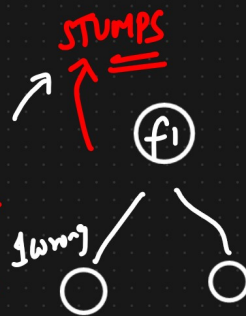
Bagging = Random Forest



② Boosting

i) Adaboost \rightarrow Decision Tree overall = 1

f_1	f_2	f_3	f_4	O/p	Weight
-	-	-	-	Yes	$\checkmark \frac{1}{7}$ 0.05
-	-	-	-	No	$\checkmark \frac{1}{7}$ 0.05
-	-	-	-		$\checkmark \frac{1}{7}$ 0.05
-	-	-	-		$\checkmark \frac{1}{7}$ 0.05
X	-	-	-		$\frac{1}{7}$ 0.349
-	-	-	-		$\checkmark \frac{1}{7}$ 0.05
-	-	-	-		$\checkmark \frac{1}{7}$ 0.05
-	-	-	-		$\checkmark \frac{1}{7}$ 0.05



information gain & entropy

② Performance of Stump

$$= \frac{1}{2} \log_e \left(\frac{1 - TE}{TE} \right)$$

$$= \frac{1}{2} \log_e \left(\frac{1 - \frac{1}{7}}{\frac{1}{7}} \right)$$

$$= 0.895$$

{ weak learner }

Total Error = $\frac{1}{7}$ (TE)

③ New Sample =

Weight = Weight $\times e^{-p_s} = \frac{1}{7} \times e^{-0.895} = 0.05$

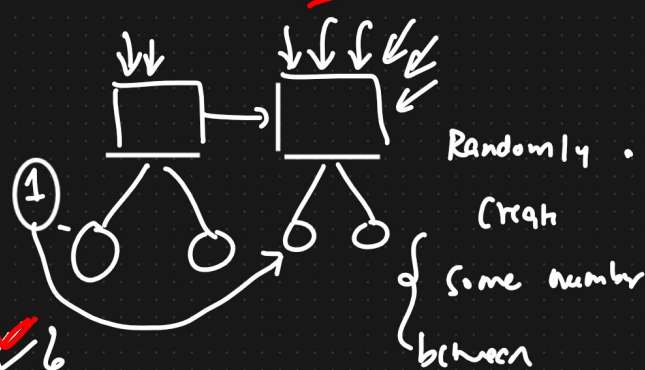
Incorrect record = Weight $\times e^{p_s} = \frac{1}{7} \times e^{0.895} = 0.349$

Is it 1??

New Weight

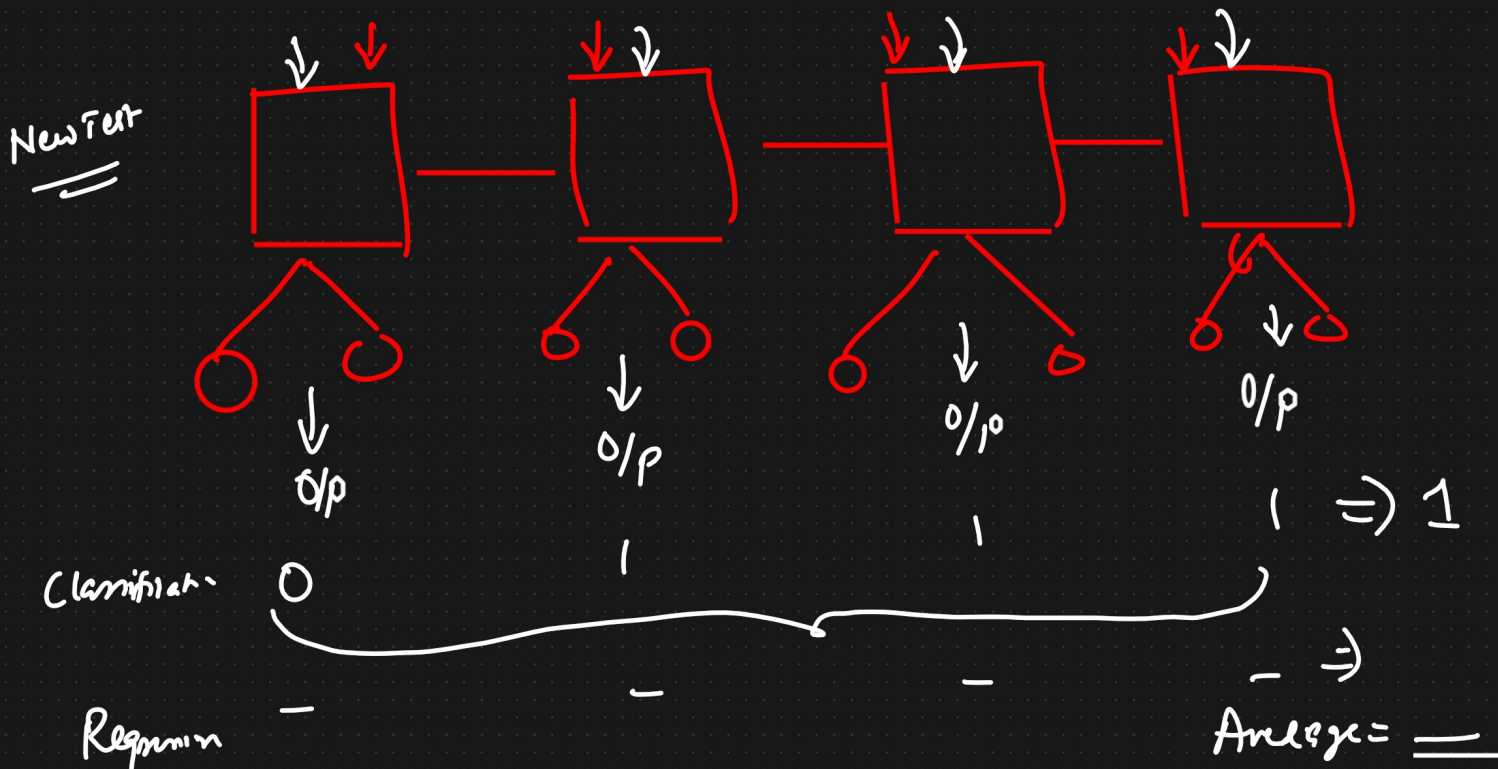
Normalized weight

Buckets



0.05 ÷ 0.649	0.07	[0 - 0.07]
0.05 ÷ 0.649	0.07	[0.07 - 0.14]
0.05 ÷ 0.649	0.07	[0.14 - 0.21]
→ 0.349	0.537	{ [0.21 - 0.747] ✓ }
0.05	0.07	[0.747 - 0.751]
0.05	0.07	[—]
0.05	0.07	[—]
<u>0.649</u>	<u>~ 1</u>	

$$\frac{0.537 [0-1]}{0.21} = 0.747$$



Black models VS White box Models

ANN → Black Box

→ Linear Regression → White box

Random Forest → Black box

Decision Tree → White box