Day-5

Agenda

1) Ensemble Techniques [Bagging / Boosting

2) Random forest

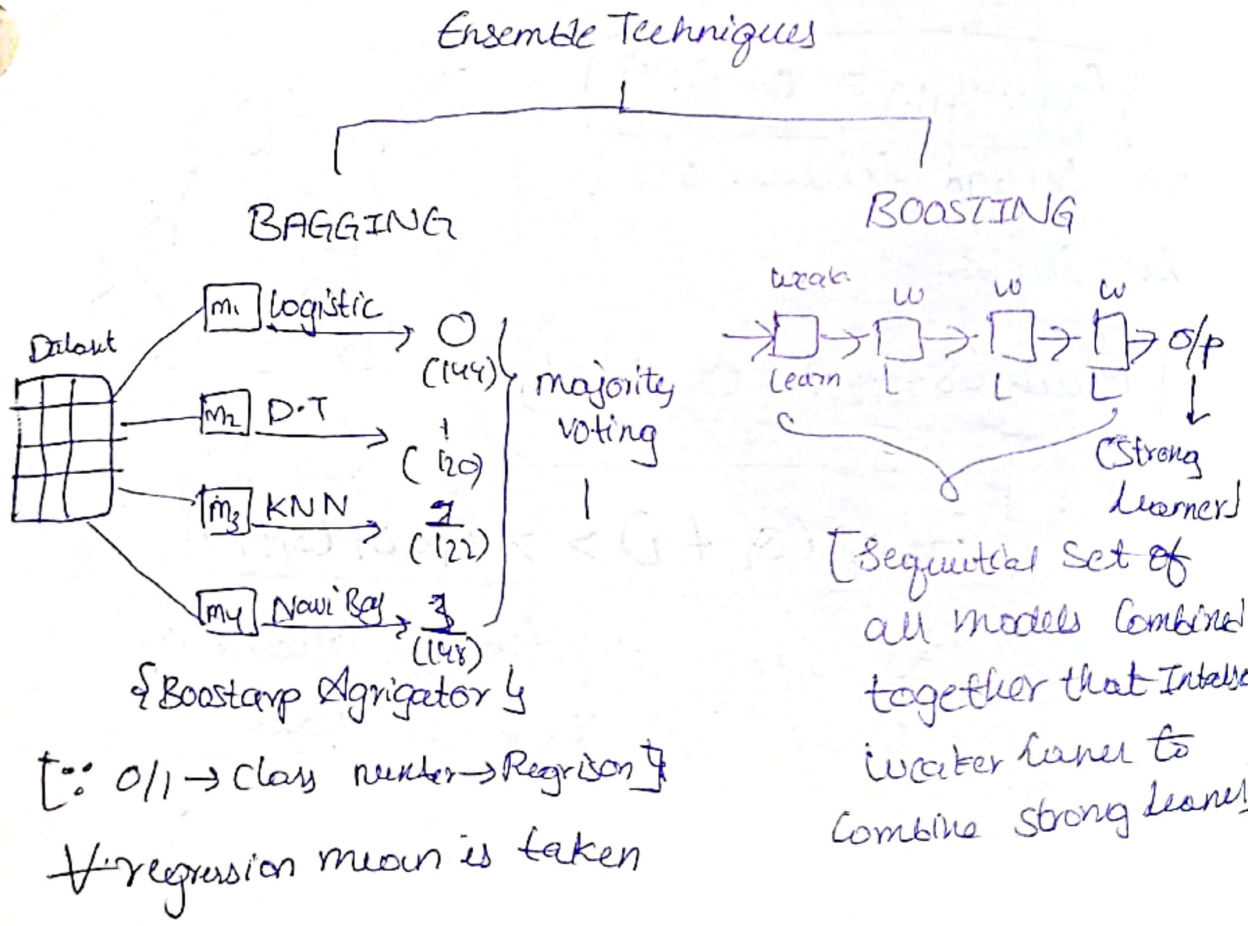
3) Adaboost

4) Xgboost

Ensemble Techniques

i) Algorithm < Classification Algorithm

multiple algorithms to solve a problem?

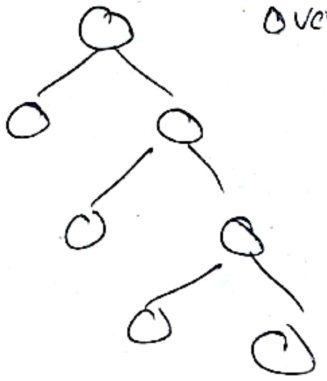


[∵ 0/1 → Class number → Regrison]

∴ regression mean is taken

[Sequential set of all models combined together that intake weaker learner to combine strong learner]

Random forest



Overfitting { low Bias high Variance }

{ low Bias low Variance }

Prop

1. Normalisation [No]
2. KNN {Standarzation} [Yes]
3. Random forest does Impact by outlier? [No]

Bagging = Random forest [used]

② Boosting

(1) Adaboost

t_1 t_2 t_3 t_4 O/p weight

- - - - Yes 0.05 1/5 (t1) (t2) (t3)
- - - - No 0.05 1/5 wrong
- - - - 1/5 {weak learn}
- - - 0.349 1/5
- - - 0.05 1/5

① Total Error = $\frac{1}{5}$

② Performance of stump

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

$= \frac{1}{2}\log_e\left(\frac{1-1/5}{1/5}\right)$

$= 0.895$

③ New Sample Weight (after updation removing error)

= Weight of e^{-Ps}

$= \frac{1}{5} \times e^{-0.895}$

④ Incorrect record = weight $\times e^{Ps}$

Correct record = 0.05

Black Box model VS White Box Model

Linear regression → White Box

Random Forest → Black Box

Decision Tree → White Box

ANN → Black Box

New weight	Normalised weight	Bucket
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] (large)
0.05 : "	0.07	[0.747-0.751]
0.05 : "	0.07	
0.05 : "	0.07	
0.649	≈ 1	

avg