

Bagging & Boosting

In Decision Trees we have a problem of :-

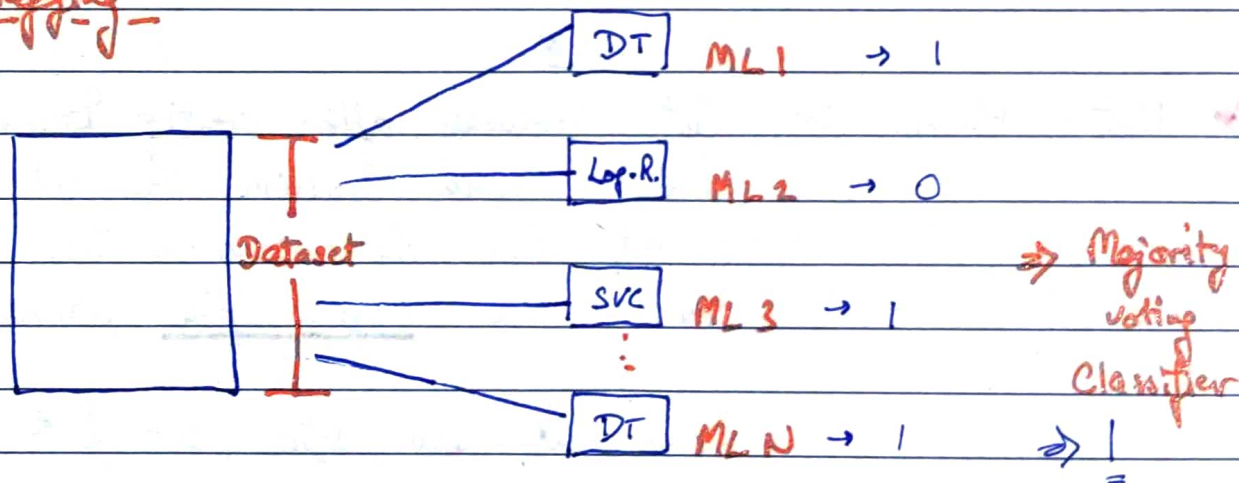
Overfitting $\begin{cases} \text{Low Bias} \\ \text{High variance} \end{cases}$

To overcome this problem we use :-

Pre-pruning & Post-pruning

[Ensembled Techniques]

1. Bagging



Let say we run all these algos on dataset parallelly & we get various output as 0 & 1

↳ Final output will be based on Majority voting Classifier

And this technique is **BAGGING** of algorithms.

Whenever Bagging occurs, there is a concept of something called as Bootstrap Aggregation.



1. We divide dataset.
2. Train with different ML [Bootstrap]
3. Then we aggregate the entire result. [Aggregation]

This technique is called as **Ensemble Techniques**.

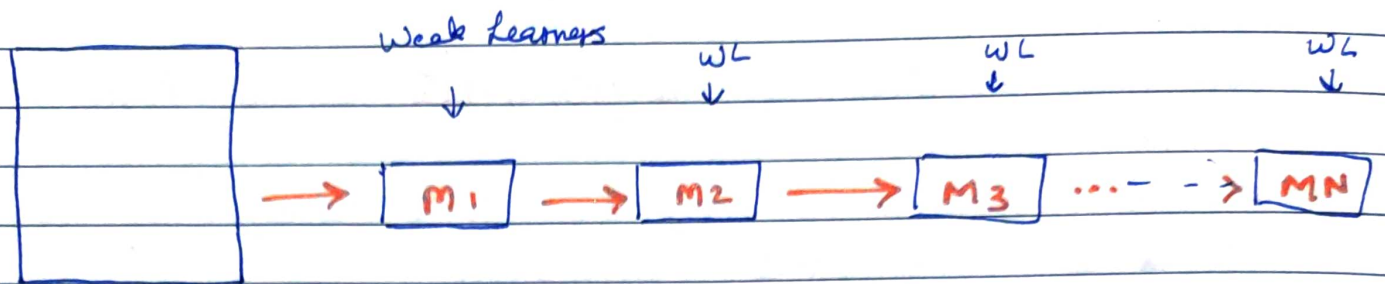
Note:- In case of Regression, we take Avg or Mean.

→ Now in Bagging we use 2 Algorithms :-

1. Random Forest Classifier
2. Random Forest Regressor

2. Boosting-

Here we don't create models parallelly but rather sequentially.



Idea behind Boosting is ... Combine weak learners
to create
Strong learner

~~Keep~~ Keep passing outputs to get even better results.

→ Algorithms in Boosting :-

1. Ada Boost Reg & Classifier
2. Gradient Boost ~ ~ ~
3. Xgboost ~ ~ ~
↳ Extreme gradient boost.