

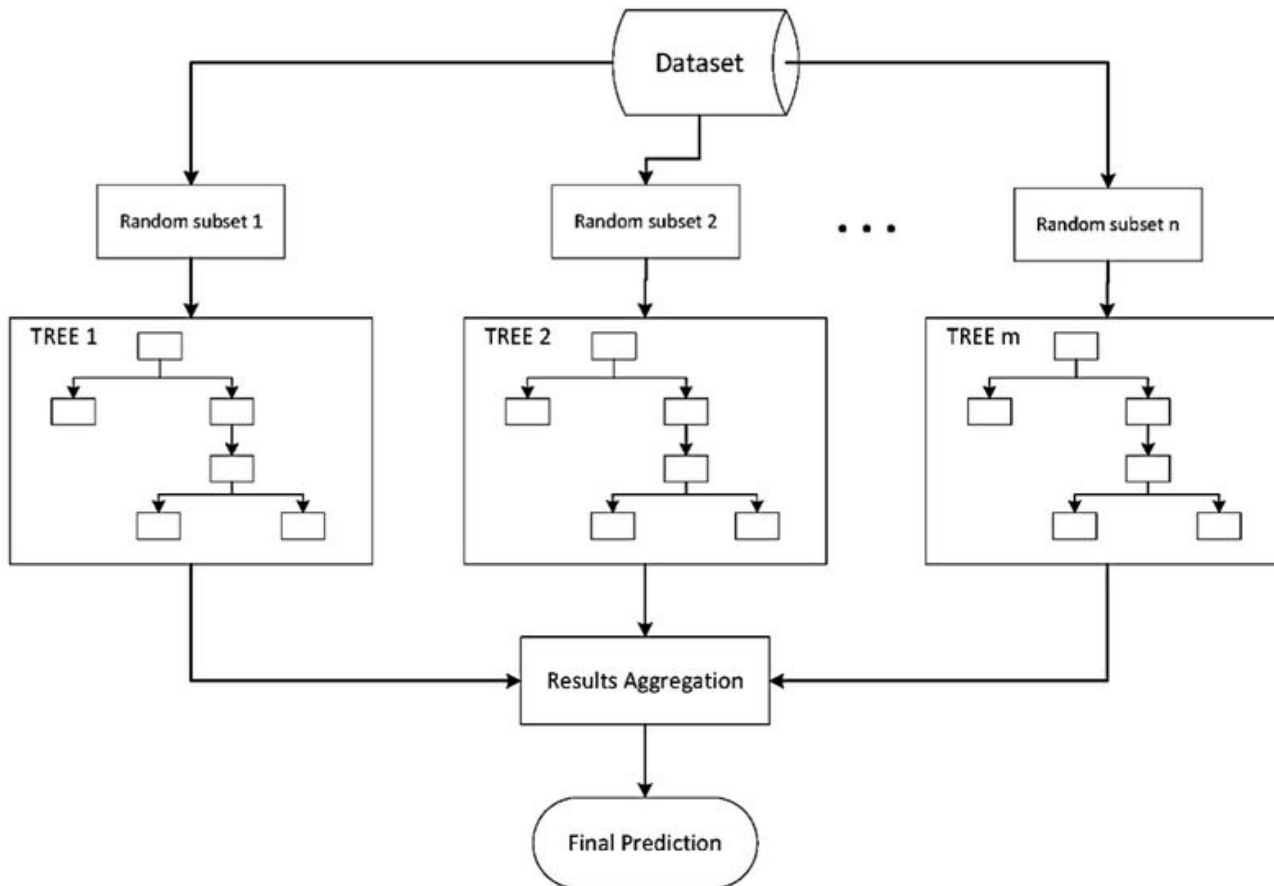
Bagging

Bootstrap AGGregating or BAGGing gets its name because it combines Bootstrapping and Aggregation to form one ensemble model.

- Given a sample of data, multiple subsamples are pulled and a Decision Tree is formed on each of the subsamples.
- After that an algorithm is used to aggregate over the Decision Trees to form the most efficient predictor.
- Once we have a prediction from each model then use a model averaging technique to get the final prediction output.
- One of the famous techniques used in Bagging is Random Forest. In the Random forest, we use multiple decision trees.

**What is
Ensemble Learning ?**

Bagging



Given a Dataset, subsamples are pulled and a Decision Tree is formed on each bootstrapped sample. The results of each tree are aggregated to yield the strongest, most accurate predictor.

Person	Age	Male?	Height > 55"
James	11	1	1
Jessica	14	0	1
Alice	14	0	1
Amy	12	0	1
Bob	10	1	1
Xavier	9	1	0
Cathy	9	0	1
Carol	13	0	1
Eugene	13	1	0
Rafael	12	1	1
Dave	8	1	0
Peter	9	1	0
Henry	13	1	0
Erin	11	0	0
Rose	7	0	0
Iain	8	1	1
Paulo	12	1	0
Margaret	10	0	1
Frank	9	1	1
Jill	13	0	0
Leon	10	1	0
Sarah	12	0	0
Gena	8	0	0
Patrick	5	1	1

⋮



Person	Age	Male?	Height > 55"
Alice	14	0	1
Bob	10	1	1
Carol	13	0	1
Dave	8	1	0
Erin	11	0	0
Frank	9	1	1
Gena	8	0	0

y



$h(x)$

Generalization

$$\text{Error}_{L(h)} = E_{(x,y) \sim P(x,y)} [f(h(x), y)]$$

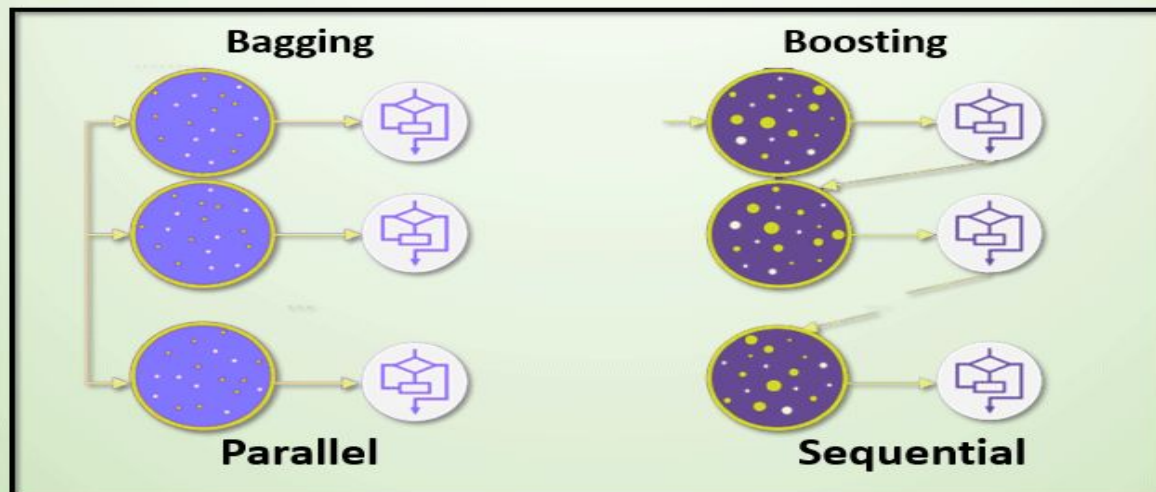
Boosting

- The term 'Boosting' refers to a family of algorithms which converts weak learner to strong learners.
- The weak learner is the classifiers that are correct only up to a small extent with the actual classification, while the strong learners are the classifiers that are well correlated with the actual classification.
- To find weak rule, we apply base learning (ML) algorithms with a different distribution. Each time base learning algorithm is applied, it generates a new weak prediction rule. After many iterations, the boosting algorithm combines these weak rules into a single strong prediction rule.



INTERNSHIPSTUDIO

Bagging and Boosting



Boosting

Choosing different distribution for each round

In boosting we take records from the dataset and pass it to base learners sequentially

- Suppose we have m number of records in the dataset. Then we pass a few records to base learner BL1 and train it and then we pass all the records from the dataset and see how the Base learner works.
- For all the records which are classified incorrectly by the base learner, we only take them and pass it to other base learner say BL2 and simultaneously we pass the incorrect records classified by BL2 to train BL3.
- This will go on unless and until we specify some specific number of base learner models we need.
- Finally, we combine the output from these base learners and create a strong learner, as a result, the prediction power of the model gets improved.

Top advantages and disadvantages

Advantages of Bagging

- Multiple weak learners can work better than a single strong learner.
- It provides stability and increases the accuracy of the ML algorithm that is used in classification and regression.
- It helps in reducing variance i.e. it avoids overfitting.

Disadvantages of Bagging

- It may result in high bias if it is not modelled properly and thus may result in underfitting.
- Since we must use multiple models, it becomes computationally expensive and may not be suitable in various use cases.

Advantages of Boosting

- It is one of the most successful techniques in solving the two-class classification problems.
- It is good at handling the missing data.

Disadvantages of Boosting

- Boosting is hard to implement in real-time due to the increased complexity of the algorithm.
- High flexibility of this techniques results in a multiple number of parameters than have a direct effect on the behaviour of the model.



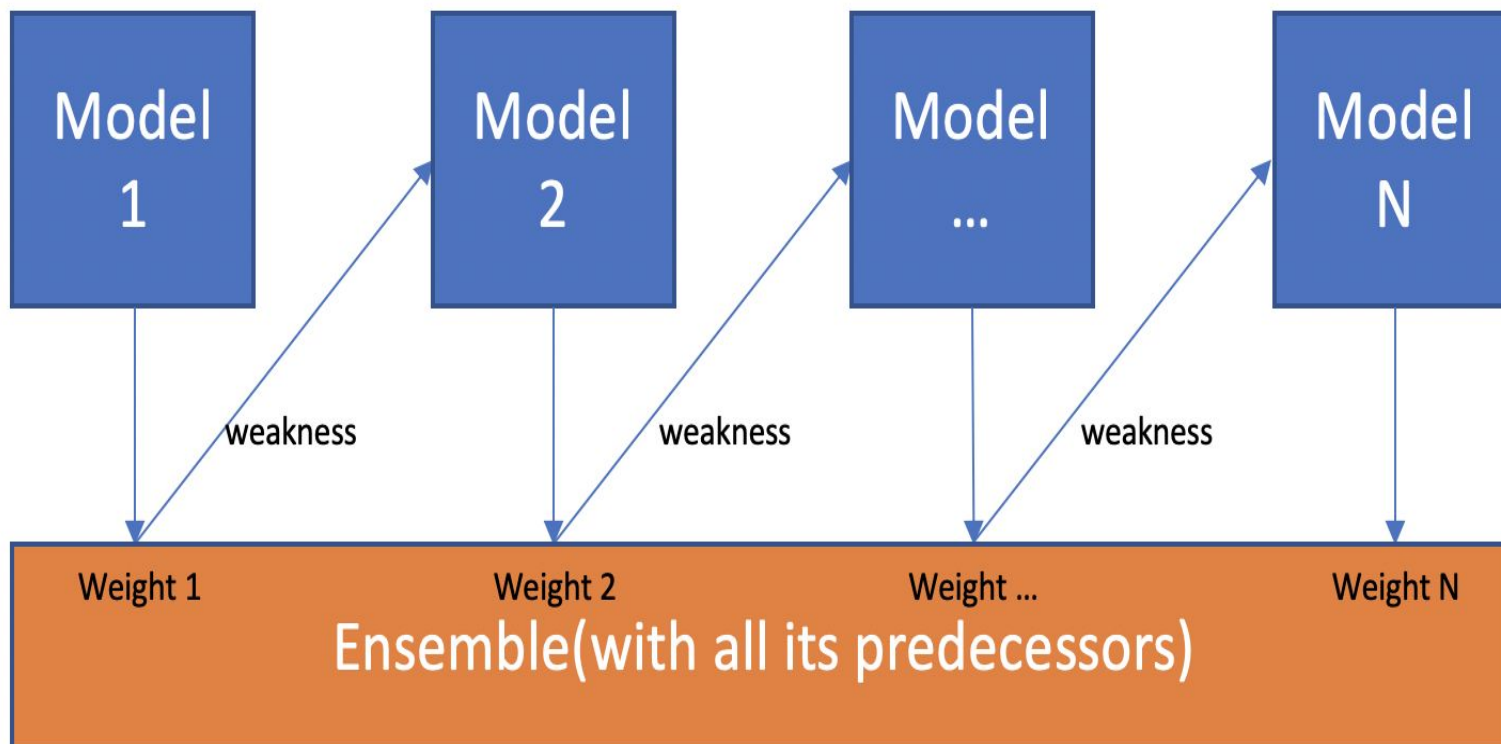
INTERNSHIPSTUDIO

Types of Boosting Algorithms

- Gradient Tree Boosting
- XGBoost

GRADIENT BOOSTING & XGBOOST

Model 1,2,..., N are individual models (e.g. decision tree)





- Q.1 What do you mean by Bagging?
- Q.2 What do you mean by Boosting?
- Q.3 What is the goal of boosting?
- Q.4 What are different methods of Boosting?