

ML Interview Question

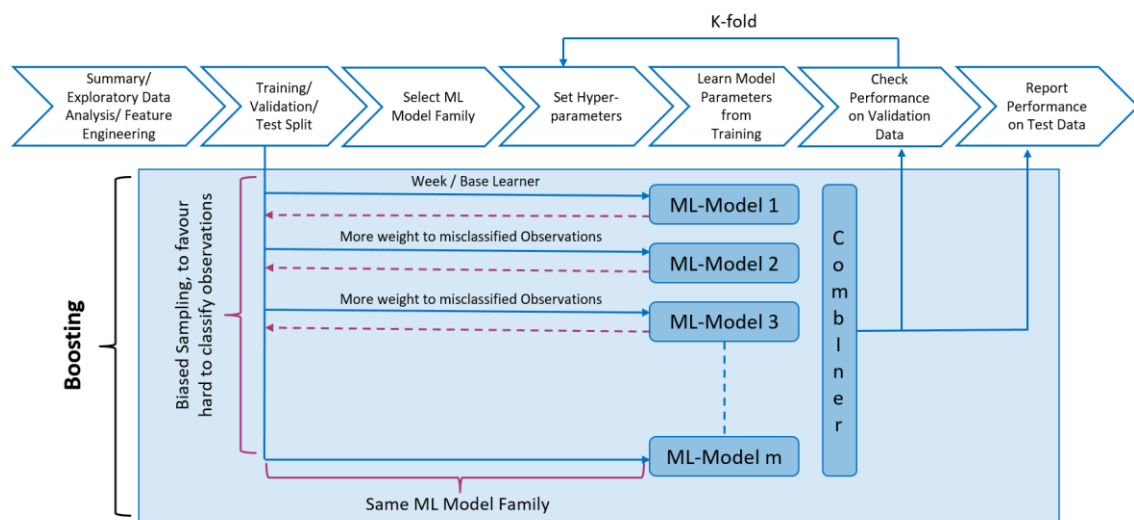
What is Boosting in Ensemble Learning

www.ashutoshtripathi.com

While Bagging helps us in reducing the model variance, Boosting helps us in improving the model accuracy. Let me ask you a question. Suppose you have tried all the possible models and none of them performing as expected. So now what you will do? I will go with Boosting. Got the point? This article explains this in detail. Before deep-diving into the Boosting let me tell you that like Bagging, Boosting is not used to reduce model variance however, it is used to improve the accuracy of the model. Also, this is one of the major difference between Bagging and Boosting.

Boosting is very simple. Just start with the weak learner also is known as the base learner. Weak learner I mean, one of those models which you have tried and not given the desired accuracy. Next Steps perform many iterations by assigning more weights to those observations which were classified wrongly in the previous iteration. In this way, we keep on boosting the model on each step. There is also the possibility that the observations which were correctly classified in previous steps might be misclassified in the next step. However, this will be taken care once you have performed many iterations and combine all the weak learners and do the final prediction.

Boosting in Ensemble Learning



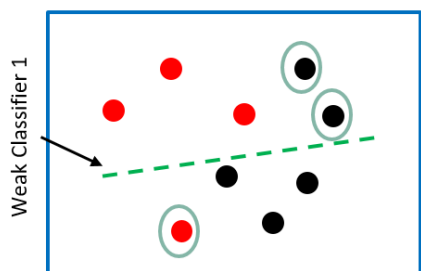
I hope now, you must have got the basic idea on What is Boosting, why and when to use it. So let's move ahead and discuss each of the point associated with boosting in detail.

Boosting

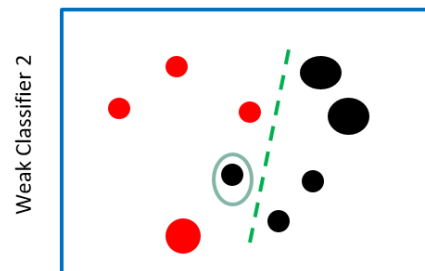
- Boosting helps to improve the accuracy of any given machine learning algorithm.
- It is algorithm independent so we can apply it with any learning algorithms.
- It is not used to reduce the model variance.
- Iterations happen sequential not parallel.
- However, as it involves many iterations to strengthen the model accuracy, it is computationally costly, so it is important to understand when to apply it.

So I would say apply only when your base model is not giving the desired level of accuracy. With the desired level I mean, you need to know when you have to stop training. Means if you are getting 90% accuracy and the business case is ok with 90% so you can neglect the 10% misclassified observations. This will be more clear once you have enough experience of working in ML domain with business understanding. So if you are facing it little difficult to understand, you are good to ignore it. It will not have much effect on your algorithm understanding.

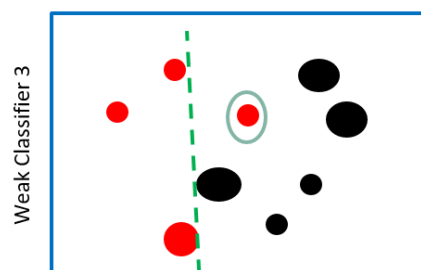
Boosting – Key Intuition



Initially weights are uniformly distributed on training data set. Circled one's are misclassified.



Assign more weight to misclassified observations



Assign more weight to misclassified observations

Note:

Final Classifier will be the weighted combination of these weak classifiers

$$H(x) = \text{sign}(a_1 h_1(x)) + \text{sign}(a_2 h_2(x)) + \text{sign}(a_3 h_3(x))$$

Steps:

1. Start with one of the ML algorithm. And train model on training data. Here if you see the above diagram, weak classifier 1 is trained on training data which has uniformly distributed weight to each observation. This means every observation in training data has equal preference.
2. Check the observations (also known as rows / items) which are misclassified.
3. Assign more weight to misclassified observations. This means, when we train the next time, algorithm will pay more attention to those classifier which were misclassified in previous step.
4. Again train the model on new distribution. This will generate the weak classifier 2. Refer the above diagram.
5. Iterate steps 2 to 4 enough no of times. This no is not fixed, and hence we call this one of the hyper parameter for boosting algorithm.
6. We have to combine all these weak classifier and generate the final prediction. So after certain no of iteration we combine and check if accuracy is met then we stop otherwise we generate more weak classifier.

How should the weak rules be combined into a single rule?

Take a (weighted) majority vote of their predictions is natural and effective way of combining and getting the final prediction.

A more intuitive visual representation of how Boosting works.

Adaptive Boosting – Adaboost

