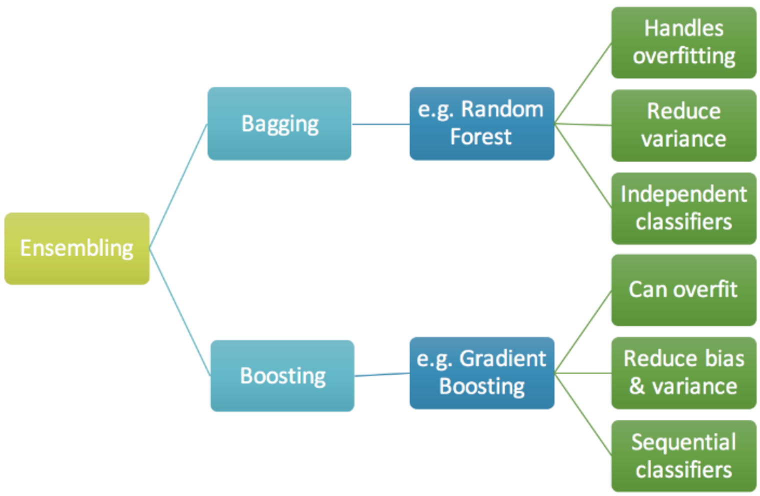
ENSEMBLE

Ensemble learning is a machine learning paradigm where multiple models, often called weak learners are trained to solve the same problem and combined to get better results. The main hypothesis is that when weak models are combined, we can obtain robust model out of it. We create several subsets of data from training sample chosen randomly with replacement and each collection of subset data is used to train a model and we end up with different models. Average of all the predictions from different models are used which is more robust than a single decision tree.



Bagging

Bagging (Bootstrap Aggregation) is designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. It reduces variance and helps to avoid overfitting. Although it is usually applied to decision tree methods, it can be used with any type of method.

The principle behind bagging is that first we create multiple bootstrap samples so that each new bootstrap sample will act as an independent dataset. Then, we can fit a weak learner for each of these dataset and finally aggregate them using average or most frequent prediction.

How it works in both regression and classification

- Create many random sub-samples (size will be the same) of our dataset with replacement (some elements of the training set will present multiple times in the generated sample and some will be absent).

- Train a model on each sample, calculate the average prediction from each model in case of regression and most frequent prediction in case of classification.

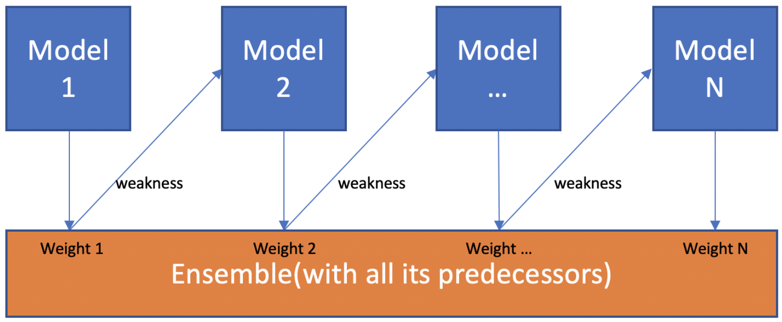
Boosting

Boosting is a technique in which the models are not made independently but sequentially i.e. each of the model is added sequentially one by one on the top of each other.

Boosting is a weighted ensemble method. Weights are updated to allow subsequent classifiers to pay more attention on training tuples that were misclassified by previous classifier. By adding models on top of each other iteratively, the errors of the previous model are corrected by the next predictor.

The models that form the ensemble (also known as base learners) could be either from the same learning algorithm or different learning algorithms. Each of these weak learners contributes some vital information for prediction enabling the boosting technique to produce a strong learner by effectively combining these weak learners. The final strong learner brings down both the bias and the variance.

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



How it works in classification

1. Assign equal weights to all observations. The weight assigned to a data point is the probability that a particular data point will be selected to be fed to the model.

2. Then it uses the Bagging (Bootstrap Aggregating) algorithm to create random samples. Given a data set D1 (m rows and p columns), it creates a new dataset D2(n rows and p columns) by row sampling at random with replacement from the original data where m>n.

3. First weak classifier is made and the data points which are predicted wrongly are now assigned higher weights.

4. Now we again create random samples. But the data points which were predicted wrongly by last week classifier has higher weights and thus have higher chances of getting selected to be fed to next weak classifier. The next weak classifier will work to predict it rightly.

5. Several classifiers specified by number of estimators hyperparameter are built and the accumulated predicted value is the final prediction value

How it works in regression

1. Fit a decision tree regressor on data.

2. Calculate error residuals which is the difference between actual target value and predicted target value

3. Fit a new model on error residuals as target variable with same input variables

4. Add the predicted residuals multiplied by learning rate to the previous predictions i.e. y_predicted2 = y_predicted1 + learning_rate * e1_predicted

5. Repeat steps 2 to 4 until the number of iterations matches the number specified by the hyperparameter i.e. number of estimators

Similarities

Both are ensemble methods to get 1 learner from N learner

Both generate several training data sets by random sampling

Both are good at reducing variance and provide higher stability

Differences

Bagging models have deep decision trees and boosting models have shallow decision trees

All models are built independently and simultaneously in bagging while in boosting models are built one after the other

Bagging models have high variance and boosting models have high bias.

Although both can produce models with low bias and low variance, bagging mainly tries to reduce variance (over-fitting) while boosting mainly tries to reduce bias (under-fitting).

So if you have good infrastructure to train the boosting model, then go for it else go for bagging. Boosting models take more time but will outperform bagging models.