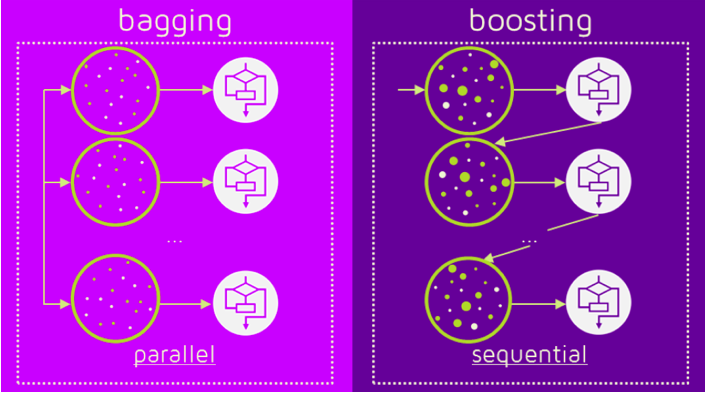
**Difference between Bagging & Boosting**

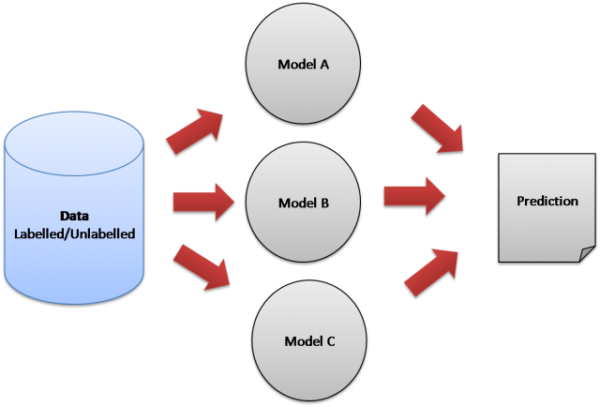
Ensemble methods: Bagging & Boosting



Difference between bagging and boosting

Bagging and boosting are commonly used terms by various data enthusiasts around the world. But what exactly bagging and boosting mean and how does it help the data science world. From this post you will learn about bagging, boosting and how they are used.

Both Bagging and boosting are part of a series of statistical techniques called ensemble methods.



Ensemble model

# Introduction to Ensemble Learning

Let’s understand the concept of ensemble learning with an example. Suppose you are a story writer and you wrote a story on some interesting topic. Now, you want to take preliminary feedback (reviews) on the story before posting it online. What are the possible ways by which you can do that?

**A:** You may ask one of your friends to rate the story for you. Now it’s entirely possible that the person you have chosen loves you very much and doesn’t want to break your heart by providing a bad review to the horrible story you have written.

**B:** Another way could be by asking 5 colleagues of yours to give reviews to your story. This should provide a better understanding of the story. This method may provide honest reviews for your story. But a problem still exists. These 5 people may not be “Subject Matter Experts” on the topic of your story. Sure, they might understand the essence, but at the same time may not be the best judges.

**C:** How about asking 50 people to review your story? Some of which can be your friends, some of them can be your colleagues and some may even be total strangers.

The responses, in this case, would be more generalized and diversified since now you have people with different sets of skills. And as it turns out — this is a better approach to get honest reviews than the previous cases we saw.

With these examples, you can infer that a diverse group of people are likely to make better decisions as compared to individuals. Similar is true for a diverse set of models in comparison to single models. This diversification in Machine Learning is achieved by a technique called Ensemble Learning. The idea here is to train multiple models, each with the objective to predict or classify a set of results.

These two are the most important terms describing the ensemble (combination) of various models into one more effective model.

* **Bagging** to decrease the model’s variance.
* **Boosting** to decreasing the model’s bias.

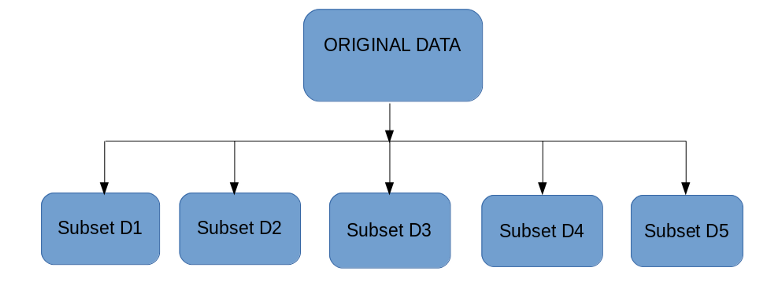
Most of the errors from a model’s learning are from three main factors: variance, noise, and bias. By using ensemble methods, we’re able to increase the stability of the final model and reduce the errors mentioned previously. By combining many models, we’re able to (mostly) reduce the variance, even when they are individually not great, as we won’t suffer from random errors from a single source.

# Bagging

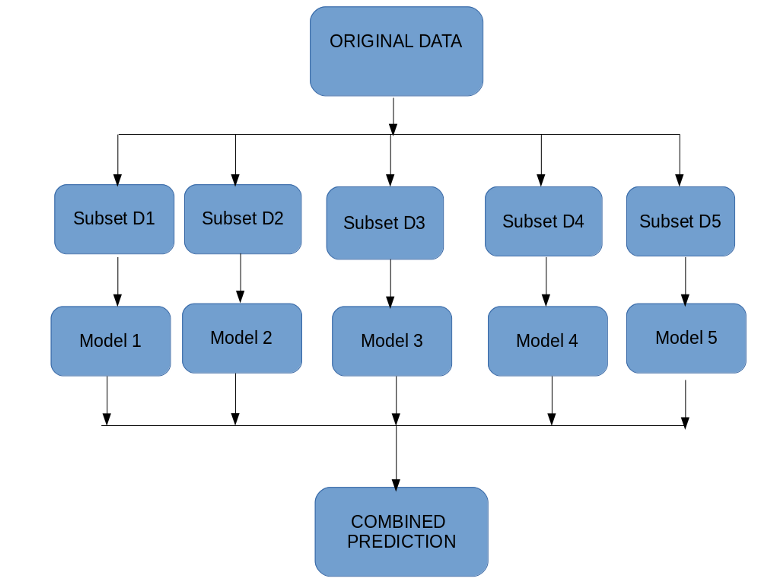
The idea behind bagging is combining the results of multiple models (for instance, all decision trees) to get a generalized result. Here’s a question: If you create all the models on the same set of data and combine it, will it be useful? There is a high chance that these models will give the same result since they are getting the same input. So how can we solve this problem? One of the techniques is bootstrapping.

Bootstrapping is a sampling technique in which we create subsets of observations from the original dataset, **with replacement**. The size of the subsets is the same as the size of the original set.

Bagging (or Bootstrap Aggregating) technique uses these subsets (bags) to get a fair idea of the distribution (complete set). The size of subsets created for bagging may be less than the original set.



1. Multiple subsets are created from the original dataset, selecting observations with replacement.
2. A base model (weak model) is created on each of these subsets.
3. The models run in parallel and are independent of each other.
4. The final predictions are determined by combining the predictions from all the models.

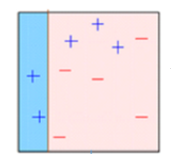


# Boosting

Before we go further, here’s another question for you: If a data point is incorrectly predicted by the first model, and then the next (probably all models), will combining the predictions provide better results? Such situations are taken care of by boosting.

Boosting is a sequential process, where each subsequent model attempts to correct the errors of the previous model. The succeeding models are dependent on the previous model. Let’s understand the way boosting works in the below steps.

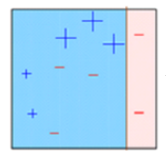
1. A subset is created from the original dataset.
2. Initially, all data points are given equal weights.
3. A base model is created on this subset.
4. This model is used to make predictions on the whole dataset.



5. Errors are calculated using the actual values and predicted values.

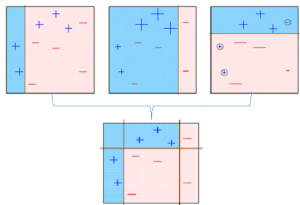
6. The observations which are incorrectly predicted, are given higher weights. (Here, the three misclassified blue-plus points will be given higher weights)

7. Another model is created and predictions are made on the dataset. (This model tries to correct the errors from the previous model)

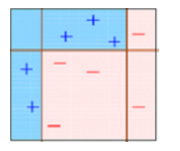


8. Similarly, multiple models are created, each correcting the errors of the previous model.

9. The final model (strong learner) is the weighted mean of all the models (weak learners).



Thus, the boosting algorithm combines a number of weak learners to form a strong learner. The individual models would not perform well on the entire dataset, but they work well for some part of the dataset. Thus, each model actually boosts the performance of the ensemble.



# Which is the best, Bagging or Boosting?

There’s not an outright winner; it depends on the data, the simulation and the circumstances. Bagging and Boosting decrease the variance of your single estimate as they combine several estimates from different models. So the result may be a model with **higher stability**.

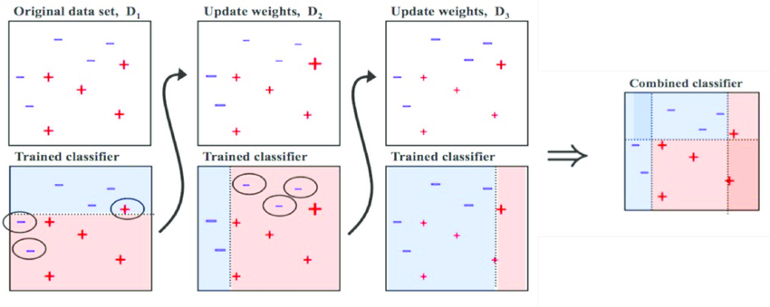
If the problem is that the single model gets a very low performance, Bagging will rarely get a **better bias**. However, Boosting could generate a combined model with lower errors as it optimises the advantages and reduces the pitfalls of the single model.

By contrast, if the difficulty of the single model is **over-fitting**, then Bagging is the best option. Boosting for its part doesn’t help to avoid over-fitting; in fact, this technique is faced with this problem itself. For this reason, Bagging is effective more often than Boosting.

# Boosting and Bagging explained with examples !!!

# ****Boosting****

Boosting algorithms is the family of algorithms that combine weak learners into a strong learner.



Working of boosting

## Main Steps involved in boosting are :

* Train model A on the whole set
* Train the model B with exaggerated data on the regions in which A performs poorly
* …

Instead of training the models in parallel, we can train them sequentially. This is the main idea of Boosting!

# ****Idea behind boosting algorithms?****

The idea behind boosting algorithms is to learn weak classifiers that are only slightly correlated with the true classification, they combine them into a strong classifier that is well-correlated with the true classification.

In order to provide you with an idea about mail spam detection problem. The spam detection problem can be divided into the following steps.

* Detect if an email contains ‘how have earned the prize’?
* If the email contains only an image?
* Who is the sender?
* How caps lock was used in the email?
* Check the subject line in the email

All these steps are the weak classifiers to detect spam. Individually they cannot answer the question if the email is spam or not. However, when they are used together, they can detect spam with high probability and accuracy.

**Working of boosting algorithm?**

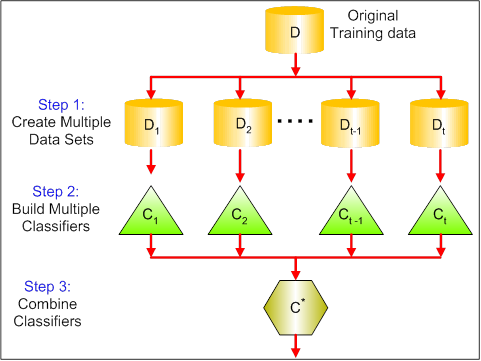
Boosting algorithm iteratively learns weak classifiers and add them to a final strong classifier. The added weak classifiers are usually weighted according to their accuracy. After each iteration, the training data is reweighted in a way that misclassified instances gain weight and correctly classified instances lose weight. On the next iteration, the future weak learner concentrates mostly on wrongly classified instances. The boosting algorithms mostly differ in reweighting approach applied to the training set.

Common Boosting algorithms:

* AdaBoost
* GBM
* XGBM
* Light GBM
* CatBoost

# Bagging( ****B****ootstrap ****Agg****regat****ing****)

As we discussed before bagging is an ensemble technique mainly used to reduce the variance of our predictions by combining the result of multiple classifiers modelled on different sub-samples of the same data set.



**Working of bagging**

## Main Steps involved in bagging are :

* **Creating multiple datasets**: Sampling is done with a replacement on the original data set and new datasets are formed from the original dataset.
* **Building multiple classifiers:**On each of these smaller datasets, a classifier is built, usually, the same classifier is built on all the datasets.
* **Combining Classifiers:**The predictions of all the individual classifiers are now combined to give a better classifier, usually with very less variance compared to before.

Bagging is similar to Divide and conquer. It is a group of predictive models run on multiple subsets from the original dataset combined together to achieve better accuracy and model stability.

# Bagging Sampling Example

N = {18,20,24,30,34,95,62,21,14,58,26,19} — Original sample with 12 elements

Bootstrap sample A: {20, 34, 58, 24, 95,18}

Bootstrap sample B: {62, 21, 19, 30, 14,26}

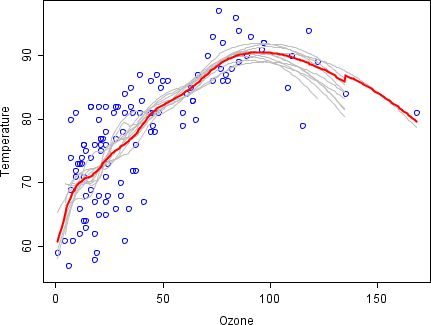
Bootstrap sample C: {58, 24, 18, 24, 34,20}

Why do we create bootstrap samples? Bootstrap samples are created to estimate and validate models for improved accuracy, reduced variance and bias, and improved stability of a model.

Once bootstrap samples are created, a model classifier is used for training or building a model and then selecting model based on popularity votes. In the classification model, a label with maximum votes will be assigned to the observations. The average value is used in case of a regression model.

## To explain the basic scenario of bagging. Below is an analysis of the relationship between ozone and temperature (taken from Wikipedia)

The relationship between temperature and ozone in this data set is apparently non-linear, based on the scatter plot. Instead of building a single prediction model from the complete data set, 100 samples of the data were drawn. Each sample is different from the original data set, yet resembles it in distribution and variability. Predictions from these 100 were then samples made across the range of the data. The first 10 predicted smooth fits appear as grey lines in the figure below. The lines are clearly very wigglyand they overfit the data — a result of the bandwidth being too small.



By taking the average of 100 smoothers, each fitted to a subset of the original data set, we arrive at one bagged predictor (red line). Clearly, the mean is more stable and there is less overfit.

Common Bagging algorithms:

* Bagging meta-estimator
* Random forest

Thanks for reading!