5 Easy questions on Ensemble Modeling everyone should know

Introduction

If you've ever participated in data science competitions, you must be aware of the pivotal role that ensemble modeling plays. In fact, it is being said that ensemble modeling offers one of the most convincing way to build highly accurate predictive models. The availability of bagging and boosting algorithms further embellishes this method to produce awesome accuracy level.

So, next time when you build a predictive model, do consider using this algorithm. You would definitely pat my back for this suggestion. And, if you've already mastered this method, great. I'd love to hear your experience about ensemble modeling in the comments section below.

For the rest, I am sharing some of the most commonly asked questions on ensemble modeling. If you ever wish to evaluate any person's knowledge on ensemble, you can daringly ask these questions & check his / her knowledge. In addition, these are among the easiest questions, hence you can't dare to get them wrong!
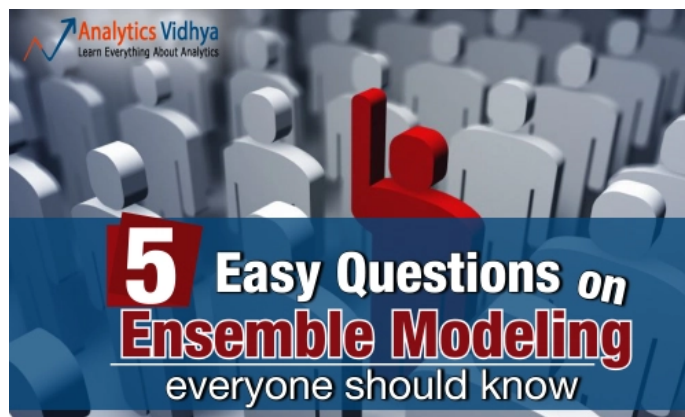
Which are the common questions (related to Ensemble Models)?

After analyzing various data science forums, I have identified the 5 most common questions related to ensemble modeling. These questions are highly relevant to data scientists new to ensemble modeling. Here are the questions:

1. What is an ensemble model?
2. What are bagging, boosting and stacking?
3. Can we ensemble multiple models of same ML algorithm?
4. How can we identify the weights of different models?
5. What are the benefits of ensemble model?

Let's discuss each question in detail.

1. What is an Ensemble Model?

Let's try to understand it by solving a classification challenge.

Problem: Set rules for classification of spam emails

Solution: We can generate various rules for classification of spam emails, let's look at the some of them:

- Spam
  - Have total length less than 20 words
  - Have only image (promotional images)
  - Have specific key words like "make money and grow" and "reduce your fat"
  - More miss spelled words in the email
- Not Spam
  - Email from Analytics Vidhya domain
  - Email from family members or anyone from e-mail address book

Above, I've listed some common rules for filtering the SPAM e-mails. Do you think that all these rules individually can predict the correct class?

Most of us would say no – And that's true! Combining these rules will provide robust prediction as compared to prediction done by individual rules. This is the principle of Ensemble Modeling. Ensemble model combines multiple 'individual' (diverse) models together and delivers superior prediction power.

If you want to relate this to real life, a group of people are likely to make better decisions compared to individuals, especially when group members come from diverse background. The same is true with machine learning. Basically, an ensemble is a supervised learning technique for combining multiple *weak learners/ models* to produce a *strong learner*. Ensemble model works better, when we ensemble models with low correlation.

A good example of how ensemble methods are commonly used to solve data science problems is the *random forest* algorithm (having multiple CART models). It performs better compared to individual CART model by classifying a new object where each tree gives "votes" for that class and the forest chooses the classification having the most votes (over all the trees in the forest). In case of regression, it takes the average of outputs of different trees.

You can also follow this article "Basics of Ensemble Learning Explained in Simple English" for more knowledge on ensemble modeling.

2. What are Bagging, Boosting and Stacking?

Let's look at each of these individually and try to understand the differences between these terms:

Bagging (Bootstrap Aggregating) is an ensemble method. First, we create random samples of the training data set (sub sets of training data set). Then, we build a classifier for each sample. Finally, results of these multiple classifiers are combined using average or majority voting. Bagging helps to reduce the variance error.
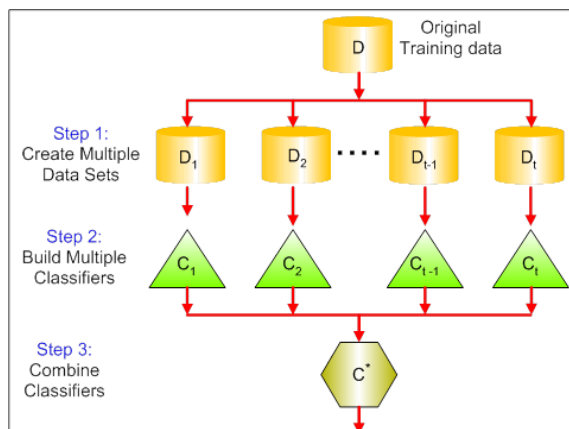

Boosting provides sequential learning of the predictors. The first predictor is learned on the whole data set, while the following are learnt on the training set based on the performance of the previous one. It starts by classifying original data set and giving equal weights to each observation. If classes are predicted incorrectly using the first learner, then it gives higher weight to the missed classified observation. Being an iterative process, it continues to add classifier learner until a limit is reached in the number of models or accuracy. Boosting has shown better predictive accuracy than bagging, but it also tends to over-fit the training data as well.
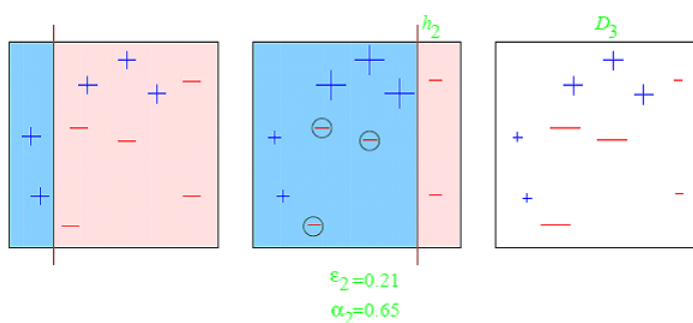
Most common example of boosting is AdaBoost and Gradient Boosting. You can also look at these articles to know more about boosting algorithms.

- [Getting smart with Machine Learning – AdaBoost and Gradient Boost](#)

- [Learn Gradient Boosting Algorithm for better predictions (with codes in R)](#)

Stacking works in two phases. First, we use multiple base classifiers to predict the class. Second, a new learner is used to combine their predictions with the aim of reducing the generalization error.
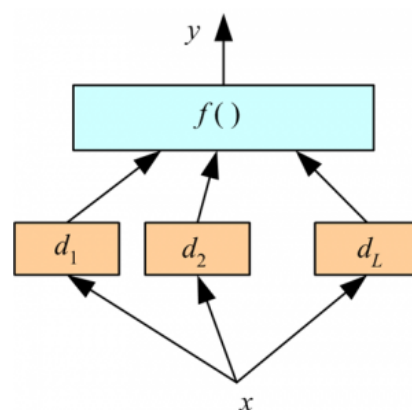
3. Can we ensemble multiple models of same ML algorithm?

Yes, we can combine multiple models of same ML algorithms, but combining multiple predictions generated by different algorithms would normally give you better predictions. It is due to the diversification or independent nature as compared to each other. For example, the predictions of a random forest, a KNN, and a Naive Bayes may be combined to create a stronger final prediction set as compared to combining three random forest model. The key to creating a powerful ensemble is model diversity. An ensemble with two techniques that are very similar in nature will perform poorly than a more diverse model set.

Example: Let's say we have three models (A, B and C). A, B and C have prediction accuracy of 85%, 80% and 55% respectively. But A and B are found to be highly correlated where as C is meagerly correlated with both A and B. Should we combine A and B? No, we shouldn't, because these models are highly correlated. Hence,we will not combine these two as this ensemble will not help to reduce any generalization error. I would prefer to combine A & C or B & C.

4. How can we identify the weights of different models for ensemble?

One of the most common challenge with ensemble modeling is to find optimal weights to ensemble base models. In general, we assume equal weight for all models and takes the average of predictions. But, is this the best way to deal with this challenge?

There are various methods to find the optimal weight for combining all base learners. These methods provide a fair understanding about finding the right weight. I am listing some of the methods below:

- Find the collinearity between base learners and based on this table, then identify the base models to ensemble. After that look at the cross validation score (ratio of score) of identified base models to find the weight.

- Find the algorithm to return the optimal weight for base learners. You can refer article Finding Optimal Weights of Ensemble Learner using Neural Network to look at the method to find optimal weight.

- We can also solve the same problem using methods like:
  - Forward Selection of learners

You can also look at the winning solution of Kaggle / data science competitions to understand other methods to deal with this challenge.

5. What are the benefits of ensemble model?

There are two major benefits of Ensemble models:

- Better prediction
- More stable model

The aggregate opinion of a multiple models is less noisy than other models. In finance, we called it "Diversification" a mixed portfolio of many stocks will be much less variable than just one of the stocks alone. This is also why your models will be better with ensemble of models rather than individual. One of the caution with ensemble models are over fitting although bagging takes care of it largely.

End Note

In this article, we have looked at the 5 frequently asked questions on Ensemble models. While answering these questions, we have discussed about "Ensemble Models", "Methods of Ensemble", "Why should we ensemble diverse models?", "Methods to identify optimal weight for ensemble" and finally "Benefits". I would suggest you to look at the top 5 solutions of data science competitions and see their ensemble approaches to have better understanding and practice a lot. It will help you to understand what works or what doesn't.

Did you find this article useful? Have you tried anything else to find optimal weights or identify the right base learner? I'll be happy to hear from you in the comments section below.

If you like what you just read & want to continue your analytics learning, [subscribe to our emails](#), [follow us on twitter](#) or like our [facebook page](#).

Article Url - https://www.analyticsvidhya.com/blog/2015/09/questions-ensemble-modeling/

Sunil Ray

I am a Business Analytics and Intelligence professional with deep experience in the Indian Insurance industry. I have worked for various multi-national Insurance companies in last 7 years.