



Basics of Ensemble Learning Explained in Simple English

Introduction

Ensemble modeling is a powerful way to improve the performance of your model. It usually pays off to apply ensemble learning over and above various models you might be building. Time and again, people have used ensemble models in competitions like Kaggle and benefited from it.

Ensemble learning is a broad topic and is only confined by your own imagination. For the purpose of this article, I will cover the basic concepts and ideas of ensemble modeling. This should be enough for you to start building ensembles at your own end. As usual, we have tried to keep things as simple as possible.

Let's quickly start with an example to understand the basics of Ensemble learning. This example will bring out, how we use ensemble model every day without realizing that we are using ensemble modeling.

Example: I want to invest in a company XYZ. I am not sure about its performance though. So, I look for advice on whether the stock price will increase more than 6% per annum or not? I decide to approach various experts having diverse domain experience:

1. *Employee of Company XYZ:* This person knows the internal functionality of the company and have the insider information about the functionality of the firm. But he lacks a broader perspective on how are competitors innovating, how is the technology evolving and what will be the impact of this evolution on Company XYZ's product. **In the past, he has been right 70% times.**

2. *Financial Advisor of Company XYZ:* This person has a broader perspective on how companies strategy will fair of in this competitive environment. However, he lacks a view on how the company's internal policies are fairing off. **In the past, he has been right 75% times.**

3. *Stock Market Trader:* This person has observed the company's stock price over past 3 years. He knows the seasonality trends and how the overall market is performing. He also has developed a strong intuition on how stocks might vary over time. **In the past, he has been right 70% times.**

4. *Employee of a competitor:* This person knows the internal functionality of the competitor firms and is aware of certain changes which are yet to be brought. He lacks a sight of company in focus and the external factors which can relate the growth of competitor with the company of subject. **In the past, he has been right 60% of times.**

5. *Market Research team in same segment:* This team analyzes the customer preference of company XYZ's product over others and how is this changing with time. Because he deals with customer side, he is unaware of the changes company XYZ will bring because of alignment to its own goals. **In the past, they have been right 75% of times.**

6. *Social Media Expert:* This person can help us understand how has company XYZ has positioned its products in the market. And how are the sentiment of customers changing over time towards company. He is unaware of any kind of details beyond digital marketing. **In the past, he has been right 65% of times.**

Given the broad spectrum of access we have, we can probably combine all the information and make an informed decision.

In a scenario when all the 6 experts/teams verify that it's a good decision(assuming all the predictions are independent of each other), we will get a combined accuracy rate of

$$\begin{aligned} 1 & - 30\% \times 25\% \times 30\% \times 40\% \times 25\% \times 35\% \\ & = 1 - 0.07875 = 99.92125\% \end{aligned}$$

Assumption: The assumption used here that all the predictions are completely independent is slightly extreme as they are expected to be correlated. However, we see how we can be so sure by combining various predictions together.

Let us now change the scenario slightly. This time we have 6 experts, all of them are employee of company XYZ working in the same division. Everyone has a propensity of 70% to advocate correctly.

What if we combine all these advice together, can we still raise up our confidence to >99% ?

Obviously not, as all the predictions are based on very similar set of information. They are certain to be influenced by similar set of information and the only variation in their advice would be due to their personal opinions & collected facts about the firm.

Halt & Think : What did you learn from this example? Was it abstruse ? Mention your arguments in the comment box.

What is Ensemble Learning?

Ensemble is the art of combining diverse set of learners (individual models) together to improvise on the stability and predictive power of the model. In the above example, the way we combine all the predictions together will be termed as Ensemble Learning.

In this article, we will talk about a few ensemble techniques widely used in the industry. Before we get into techniques, let's first understand how do we actually get different set of learners. Models can be different from each other for a variety of reasons, starting from the population they are built upon to the modeling used for building the model.

Here are the top 4 reasons for a model to be different. They can be different because of a mix of these factors as well:

1. Difference in population



2. Difference in hypothesis



3. Difference in modeling technique



4. Difference in initial seed



Error in Ensemble Learning (Variance vs. Bias)

The error emerging from any model can be broken down into three components mathematically. Following are these components:

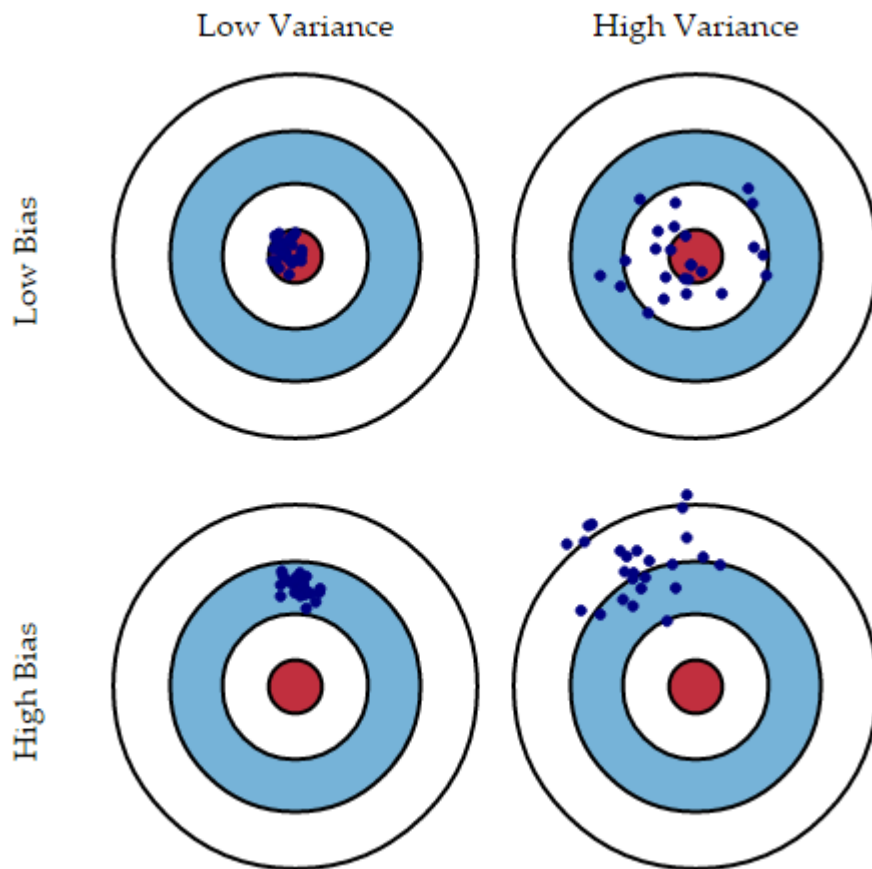
$$Err(x) = \left(E[\hat{f}(x)] - f(x) \right)^2 + E \left[\hat{f}(x) - E[\hat{f}(x)] \right]^2 + \sigma_e^2$$

$$Err(x) = \text{Bias}^2 + \text{Variance} + \text{Irreducible Error}$$

Why is this important in the current context? To understand what really goes behind an ensemble model, we need to first understand what causes error in the model. We will briefly introduce you to these errors and give an insight to each ensemble learner in this regard.

Bias error is useful to quantify how much on an average are the predicted values different from the actual value. A high bias error means we have a under-performing model which keeps on missing important trends.

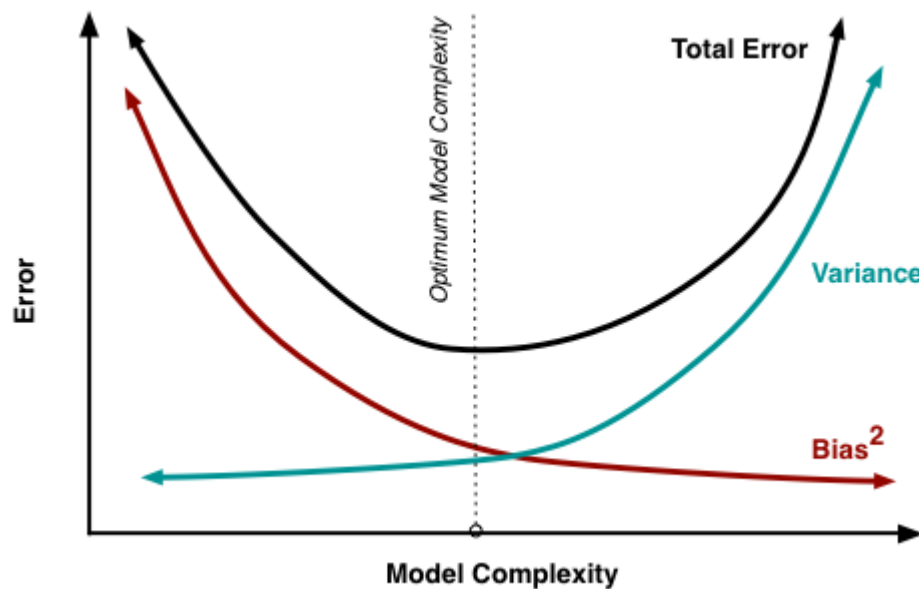
Variance on the other side quantifies how are the prediction made on same observation different from each other. A high variance model will over-fit on your training population and perform badly on any observation beyond training. Following diagram will give you more clarity (Assume that red spot is the real value and blue dots are predictions):



Credit : Scott Fortman

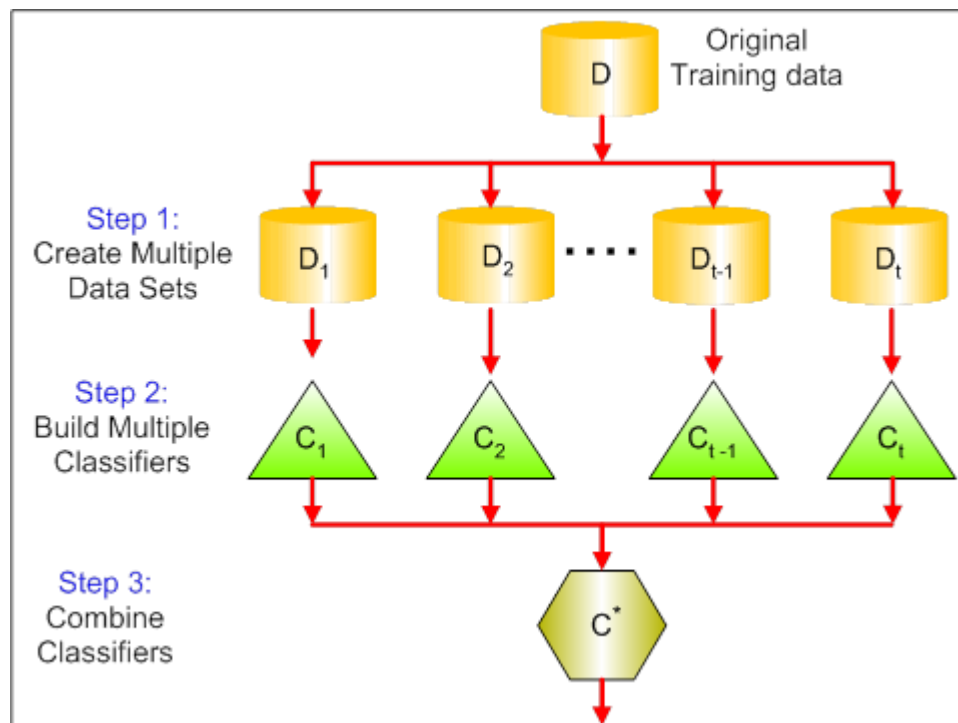
Normally, as you increase the complexity of your model, you will see a reduction in error due to lower bias in the model. However, this only happens till a particular point. As you continue to make your model more complex, you end up over-fitting your model and hence your model will start suffering from high variance.

A champion model should maintain a balance between these two types of errors. This is known as the **trade-off management** of bias-variance errors. Ensemble learning is one way to execute this trade off analysis.

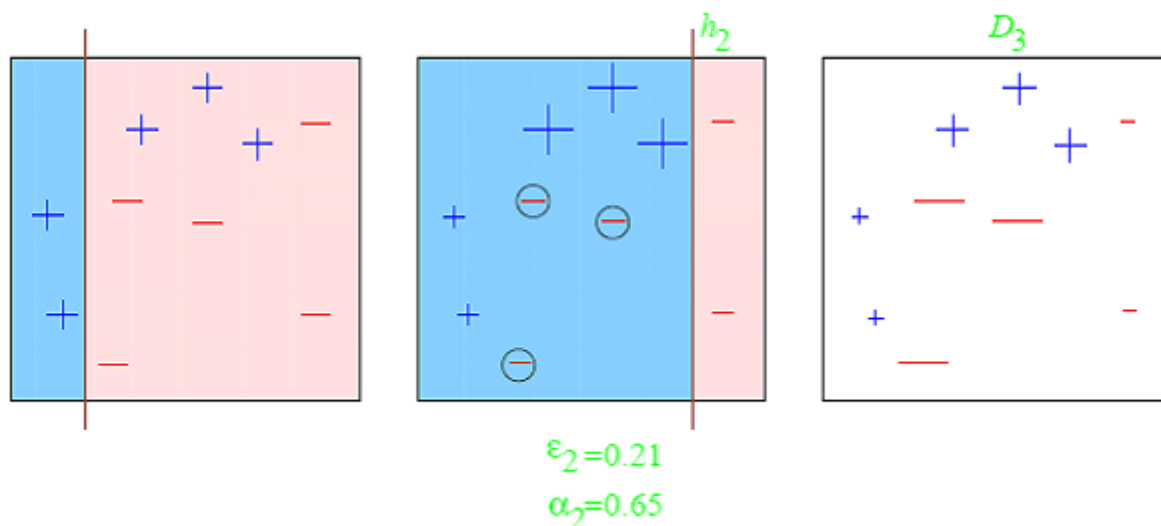


Some Commonly used Ensemble learning techniques

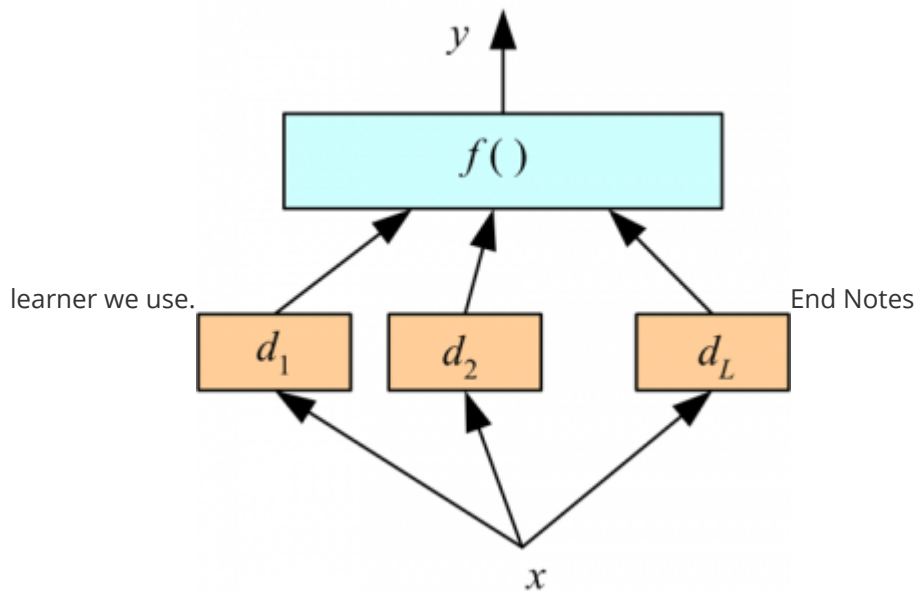
1. Bagging : Bagging tries to implement similar learners on small sample populations and then takes a mean of all the predictions. In generalized bagging, you can use different learners on different population. As you can expect this helps us to reduce the variance error.



2. Boosting : Boosting is an iterative technique which adjust the weight of an observation based on the last classification. If an observation was classified incorrectly, it tries to increase the weight of this observation and vice versa. Boosting in general decreases the bias error and builds strong predictive models. However, they may sometimes over fit on the training data.



3. Stacking : This is a very interesting way of combining models. Here we use a learner to combine output from different learners. This can lead to decrease in either bias or variance error depending on the combining



Ensemble techniques are being used in every [Kaggle Problem](#). Choosing the right ensembles is more of an art than straight forward science. With experience, you will develop a knack of which ensemble learner to use in different kinds of scenario and base learners.

Did you enjoy reading this article? Have you built an Ensemble learner before? How did you go about choosing the right ensemble technique?