# Overfitting And Underfitting



Bill Howe, UW
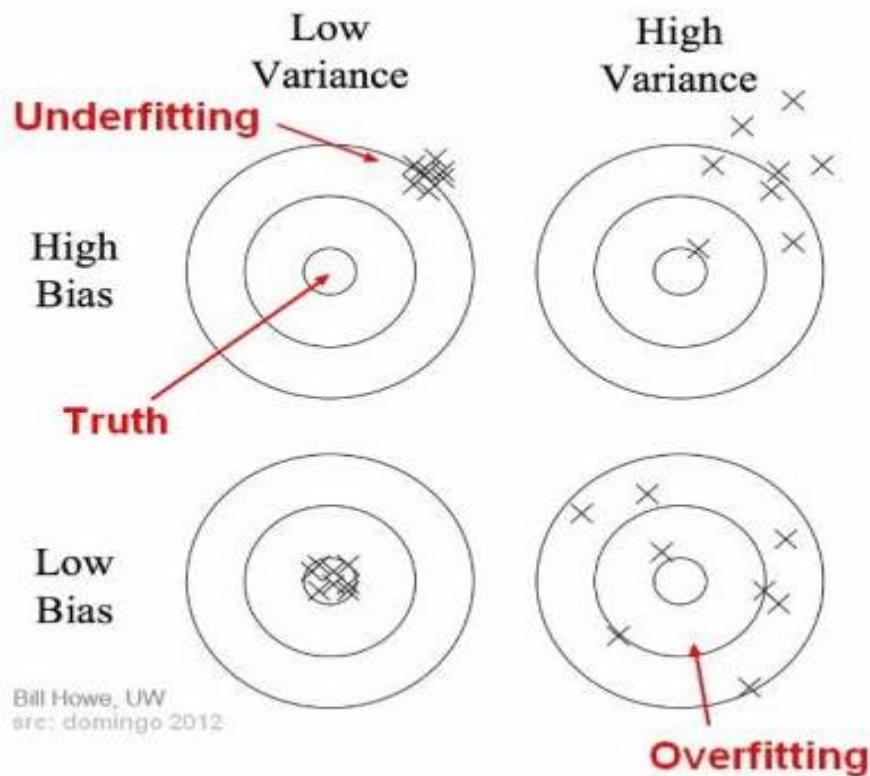src: domingo 2012

# Bias:

In our prediction model when we predict values something several times and all values are close, they may all be wrong if there is a "Bias". Bias is a systematic (built-in) error that makes all measurements wrong by a certain amount.

The term bias is the difference between the average predicted value of the model and the actual value of our model which we are trying to predict. A model with high bias pays very little attention to the training data and oversimplifies the model. Suppose in case the model with high error on training and testing data model with high Bias.
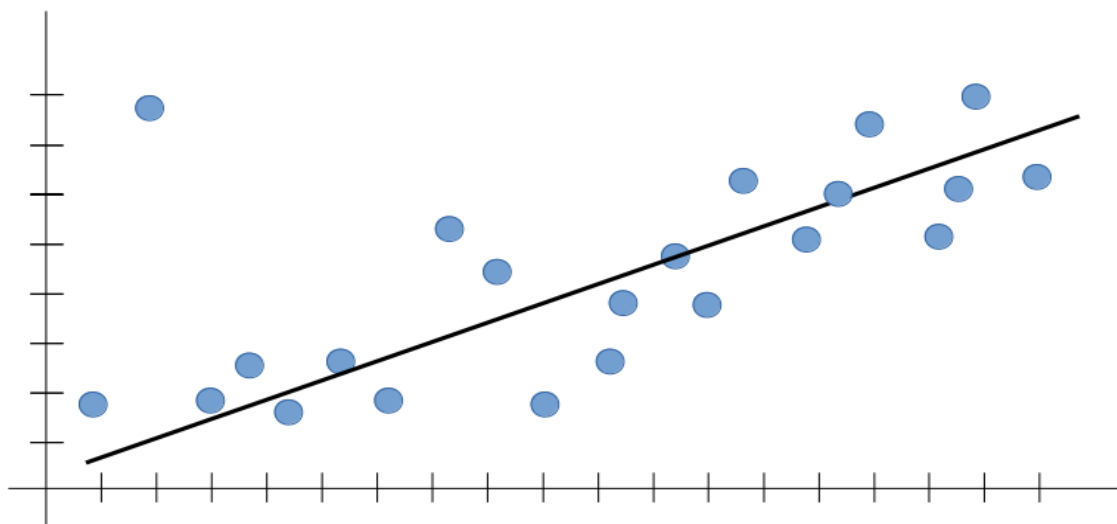
# Variance:

Variance is the variability of model prediction for a given data point or a value that tells us the spread of our data. Model with high variance pays a lot of attention to training data and does not generalize on the data which it hasn't seen before. As a result, such models perform very well on training data but have high error rates on test data.

# Underfitting

The model contains high bias is simpler than it should be and hence tends to underfit. In other words, model fails to learn on the given data and acquire the intricate pattern of the dataset.

It Means a biased model will not fit on the Training Dataset properly and hence will have low training accuracy (or high training loss).

Inability to solve complex problems: A Biased model is too simple and hence is often incapable of learning complex features and solving relatively complex problems.

Suppose we train our machine learning model on training data and it gives the high error this is called underfitting in the above graph their base fit line which contains a high error on training data.

# Overfitting

A model with high Variance will have a tendency to be overly complex. This causes the overfitting of the model.

Suppose the model with high Variance will have very high training accuracy (or very low training loss), but it will have a low testing accuracy (or a low testing loss).

Overcomplicating simpler problems: A model with a high variance means the model is overly complex and tries to fit a much more complex curve to a relatively simpler data. The model is thus capable of solving complex problems but incapable of solving simple problems efficiently.
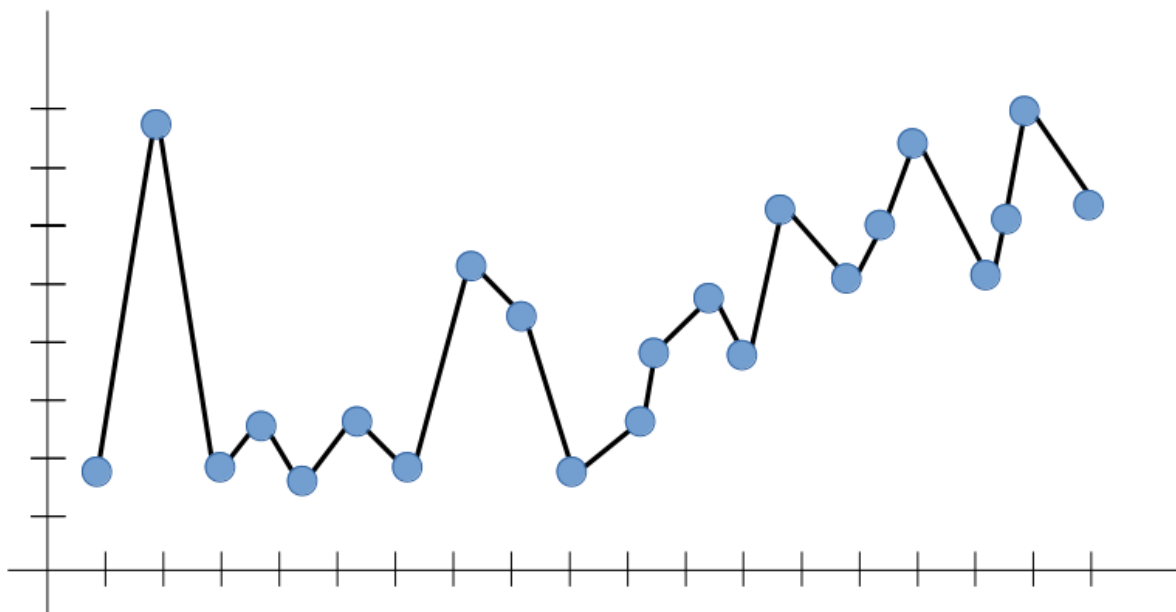


Like Above graph shows that the model fitted in the data according to the training data point so it contains high accuracy on training data but when we pass the testing data into it gives us low accuracy so the model in Overfitting.

# Bias Variance

Underfitting happens when a model unable to capture the underlying pattern of the data. These models usually have high bias and low variance. It happens when we have a very less amount of data to build an accurate model or when we try to build a linear model with nonlinear data. Also, these

kinds of models are very simple to capture the complex patterns in data like Linear and logistic regression.
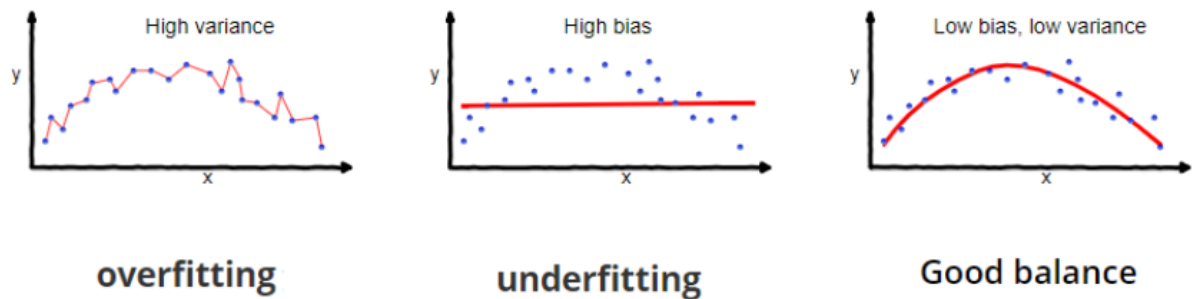


overfitting happens when our model captures the noise along with the underlying pattern in data. It happens when we train our model a lot over noisy datasets. These models have low bias and high variance. These models are very complex like Decision trees which are prone to overfitting.

# Bias Variance Trade-off

When we are facing this underfitting and Overfitting into our predictive model we need to fix this problem generally used bias-variance trade-off.

# Why is Bias Variance Trade-off?

If our model is not so complex is to be too simple and has very few parameters in data then it may have high bias and low variance. On the other hand, if our model has a large number of parameters in data then it's going to have high variance and low bias. In this case, we need to find the optimized balance without overfitting and underfitting the data.

A machine learning algorithm cant be more complex and less complex at the same time so we need to find this trade-off between bias and variance.

So we want an optimized model that contains low bias and low variance such that is minimize the total error.
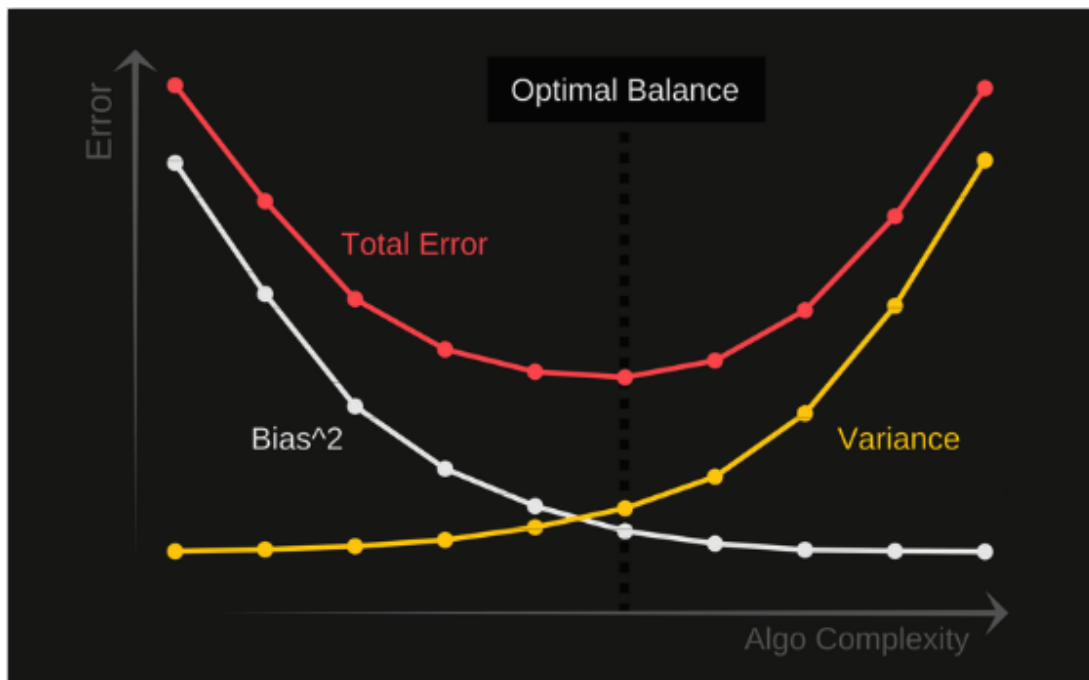
# How to detection of Bias and Variance of a model

In model building, it needs to understand the model contain high bias or high variance following are the methods to detect this high bias and high variance.

# Detection of High Bias

The model which is suffers from a very low Training Accuracy.

The Validation error is also low accuracy similar in magnitude to the training error.

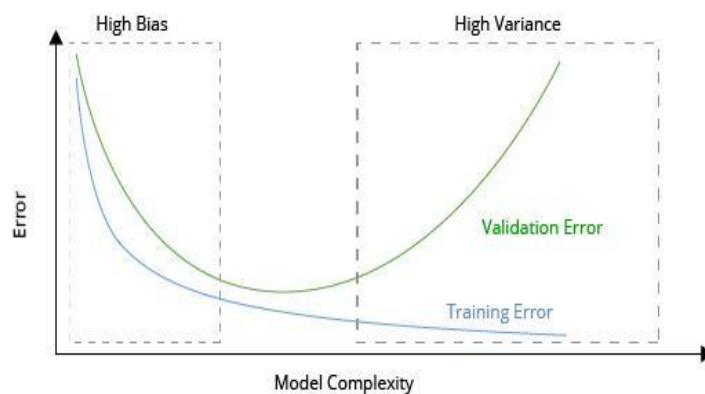The model with high bias contains model is underfitting.

# Detection of High Variance

The model which is suffers from a very Low Training Error.

The Validation error is very high accuracy as compared to training error.

The model with a high variance contains model is overfitting.

Example:- To Detect a model suffering from High Bias and Variance is shown below figure:



| Techniques to fight underfitting and overfitting | | |
|---|---|---|
| | **Underfitting** | **Overfitting** |
| **Complexity of the model** | More complex model<br><br>Try a more powerful model with a larger number of parameters<br>Ensemble learning<br>More layers / number of neurons per layer | More simple model<br><br>Try a less powerful model with a fewer number of parameters<br><br>Less layers / number of neurons per layer |
| **Regularization** | Less regularization<br><br>Decrease regularization | More regularization<br><br>Increase regularization impact<br>Early stopping, L1 / L2 regularization, dropout |
| **Quantity of features** | A larger quantity of features<br><br>Get additional features, feature engineering, polynomial features, etc. | A smaller quantity of features<br><br>Remove all additional features, feature selection |
| **Data** | Data cleaning, hold-out validation or cross validation.<br><br>Getting more data most likely will not help | Data cleaning, hold-out validation or cross validation.<br><br>Getting more data most likely will help<br>(data augmentation) |

# Conclusion

Widget not in any sidebars

In this blog, you will get an understanding of the overfitting and Underfitting of the predictive modelling technique. These methods improve your knowledge of the machine learning model also in this blog give you an understanding of how to solve underfitting and overfitting problems using Bias Variance Trade-off.