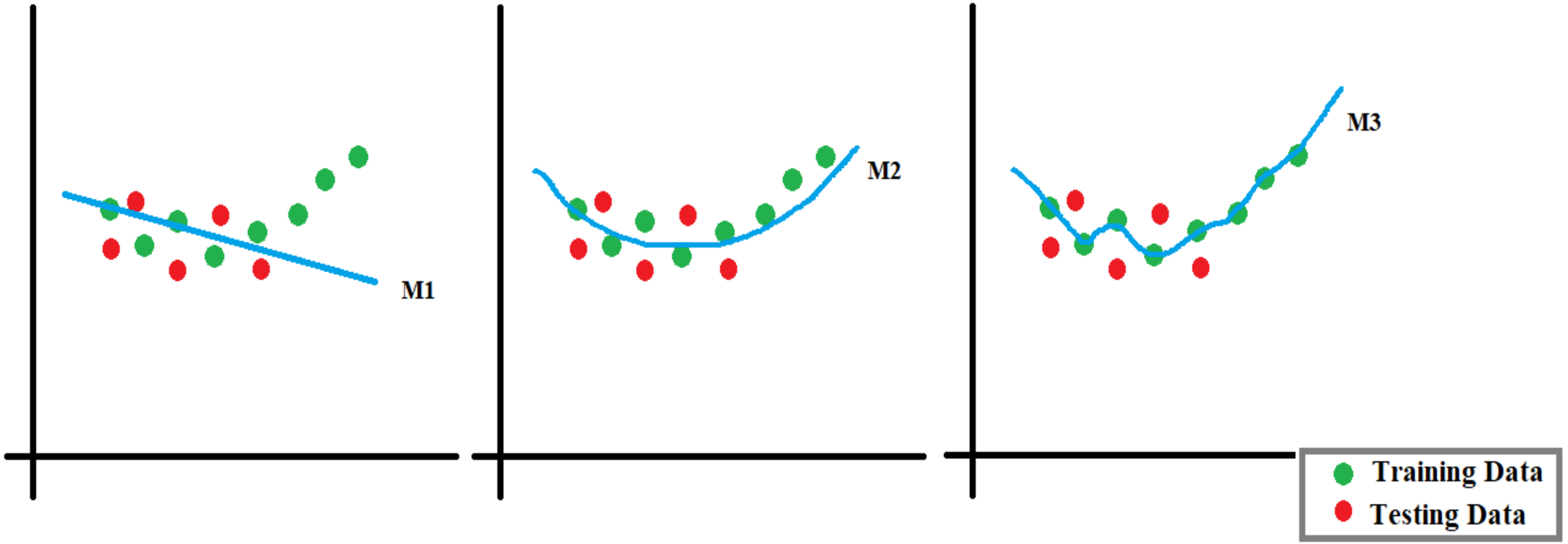
Bias and Variance in ML

- Machine learning allows machines to perform data analysis and make predictions. However, if the machine learning model is not accurate, it can make predictions errors, and these prediction errors are usually known as Bias and Variance. In machine learning, these errors will always be present as there is always a slight difference between the model predictions and actual predictions.
- The main aim of ML/data science analysts is to reduce these errors in order to get more accurate results.

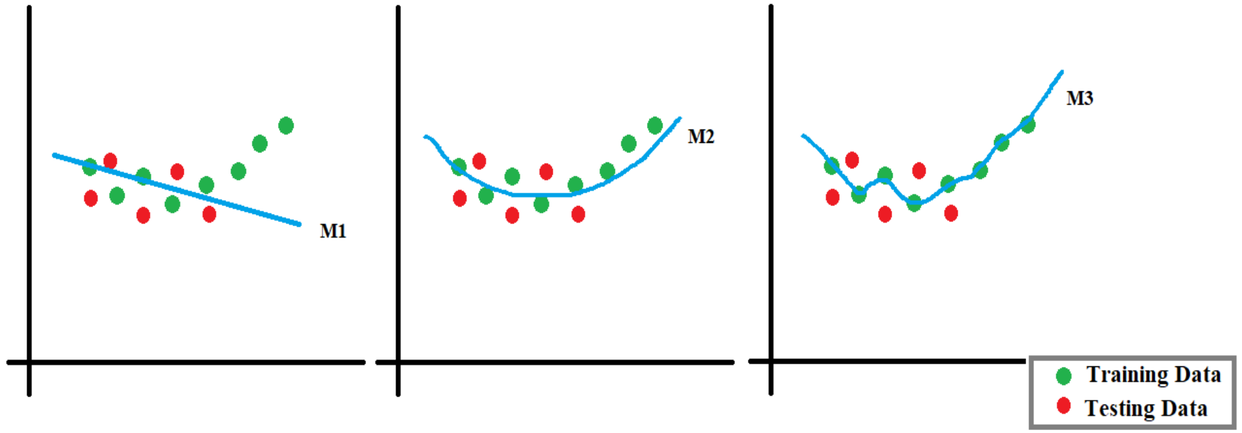
By Shraddha Dalvi

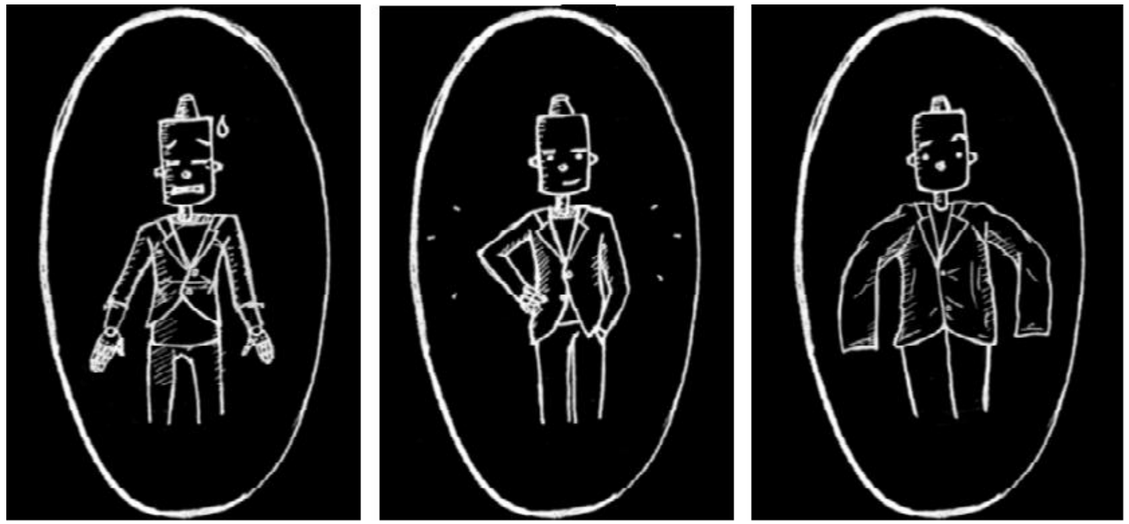
Bias and Variance in ML



By Shraddha Dalvi

Bias and Variance in ML





Underfitting | Best Fit | Overfitting

By Shraddha Dalvi

Bias and Variance in ML

What is Bias?

- While making predictions, a difference occurs between prediction values made by the model and actual values/expected values, and this difference is known as bias errors or Errors due to bias.

Low Bias:

- A low bias model will make fewer assumptions about the form of the target function.

High Bias:

- A model with a high bias makes more assumptions, and the model becomes unable to capture the important features of our dataset. A high bias model also cannot perform well on new data.

By Shraddha Dalvi

Bias and Variance in ML

What is a Variance?

- The variance would specify the amount of variation in the prediction if the different training data was used. In simple words, variance tells that how much a random variable is different from its expected value.

Low variance :

- Low variance means there is a small variation in the prediction of the target function with changes in the training data set.

High variance :

- High variance shows a large variation in the prediction of the target function with changes in the training dataset.

By Shraddha Dalvi

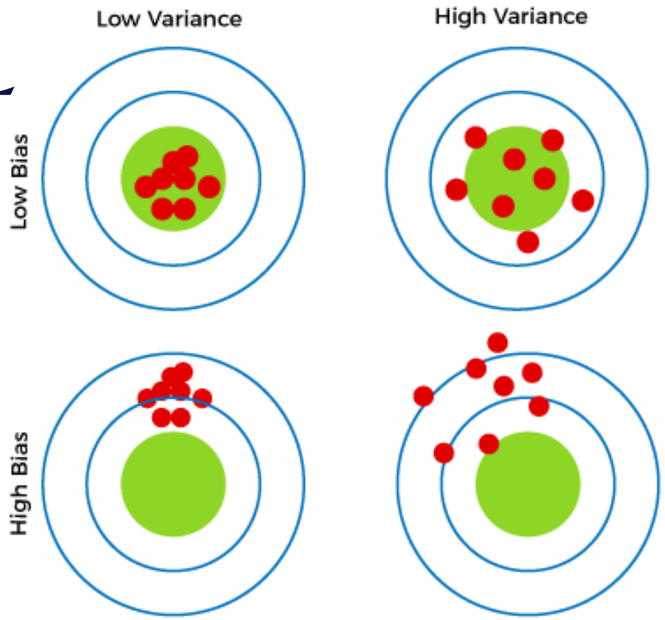
Bias and Variance in ML

Different Combinations of Bias-Variance

- There are four possible combinations of bias and variances, which are represented by the below diagram:

- Low-Bias, Low-Variance
- Low-Bias, High-Variance
- High-Bias, Low-Variance
- High-Bias, High-Variance

Bias and Variance in ML



Different Combinations of Bias-Variance

1. Low-Bias, Low-Variance

- The combination of low bias and low variance shows an ideal machine learning model. However, it is not possible practically.

2. Low-Bias, High-Variance

- With low bias and high variance, model predictions are inconsistent and accurate on average. This case occurs when the model learns with a large number of parameters and hence leads to an overfitting

3. High-Bias, Low-Variance

- With High bias and low variance, predictions are consistent but inaccurate on average. This case occurs when a model does not learn well with the training dataset or uses few numbers of the parameter. It leads to underfitting problems in the model.

4. High-Bias, High-Variance

- With high bias and high variance, predictions are inconsistent and also inaccurate on average.

By Shraddha Dalvi

Overfitting and Underfitting in ML

- Overfitting and Underfitting are the two main problems that occur in machine learning and degrade the performance of the machine learning models.
- Before understanding the overfitting and underfitting, let's understand some basic term that will help to understand this topic well:

Signal: It refers to the true underlying pattern of the data that helps the machine learning model to learn from the data.

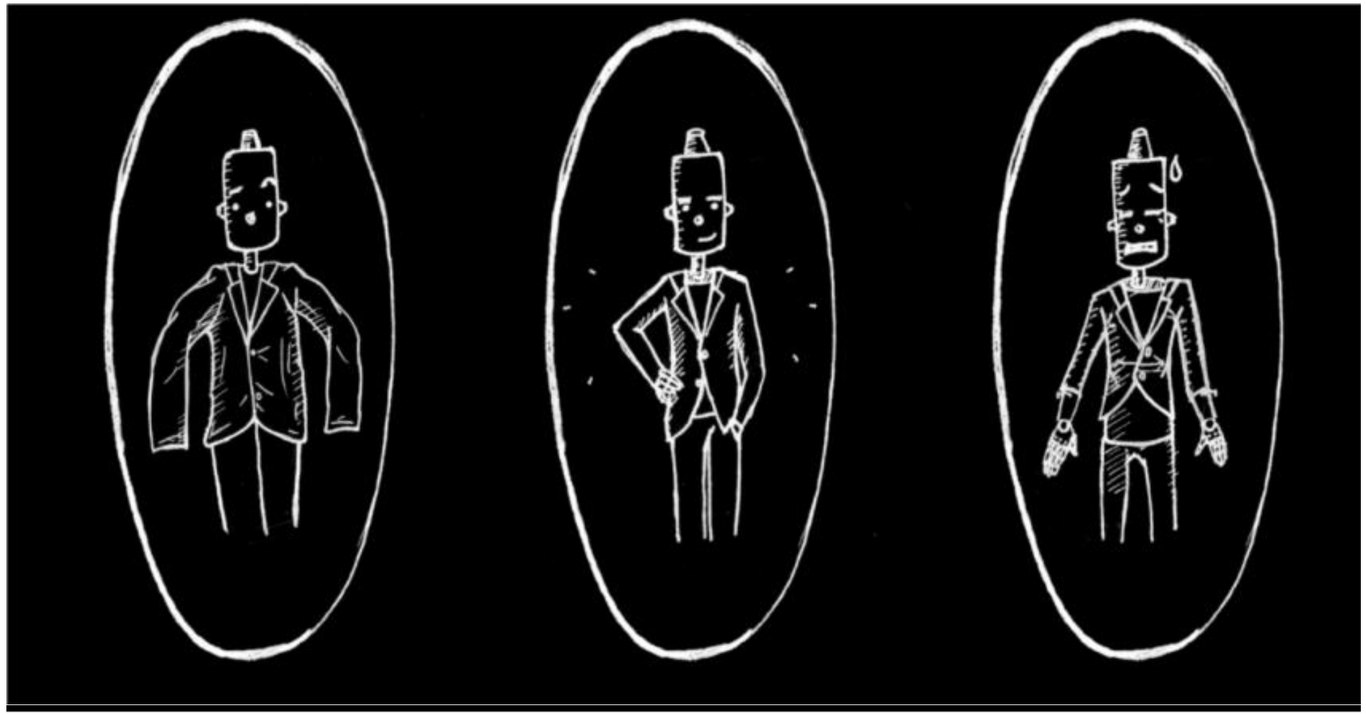
Noise: Noise is unnecessary and irrelevant data that reduces the performance of the model.

Bias: Bias is a prediction error that is introduced in the model due to oversimplifying the machine learning algorithms. Or it is the difference between the predicted values and the actual values.

Variance: If the machine learning model performs well with the training dataset, but does not perform well with the test dataset, then variance occurs.

By Shraddha Dalvi

Overfitting and Underfitting in ML



Overfitting **Best-Fit** **Underfitting**

- https://www.analyticsvidhya.com/blog/2020/02/underfitting-overfitting-best-fitting-machine-learning/

By Shraddha Dalvi

Overfitting and Underfitting in ML

Overfitting:

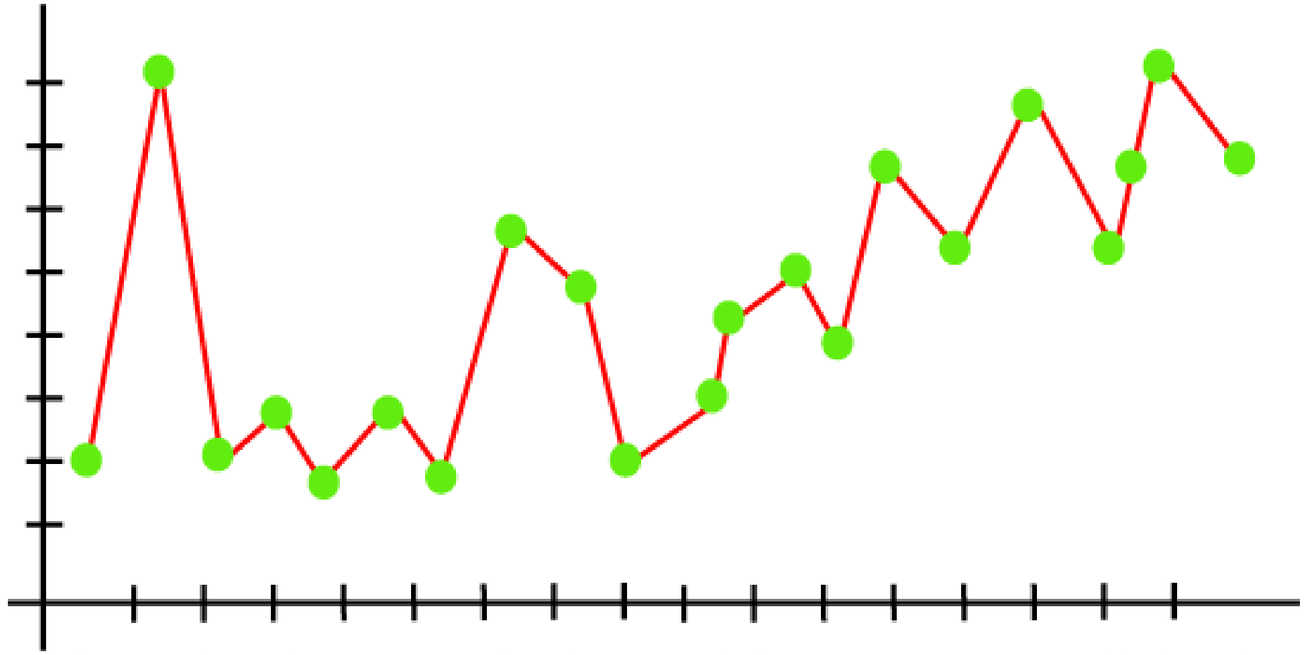
- Overfitting occurs when our machine learning model tries to cover all the data points or more than the required data points present in the given dataset. Because of this, the model starts caching noise and inaccurate values present in the dataset, and all these factors reduce the efficiency and accuracy of the model. The overfitted model has low bias and high variance.
- The chances of occurrence of overfitting increase as much we provide training to our model. It means the more we train our model, the more chances of occurring the overfitted model.

By Shraddha Dalvi

Overfitting and Underfitting in ML

Overfitting:

- Example: The concept of the overfitting can be understood by the below graph of the linear regression output:



- As we can see from the above graph, the model tries to cover all the data points present in the scatter plot. It may look efficient, but in reality, it is not so. Because the goal of the regression model to find the best fit line, but here we have not got any best fit, so, it will generate the prediction errors.

Overfitting and Underfitting in ML

Underfitting

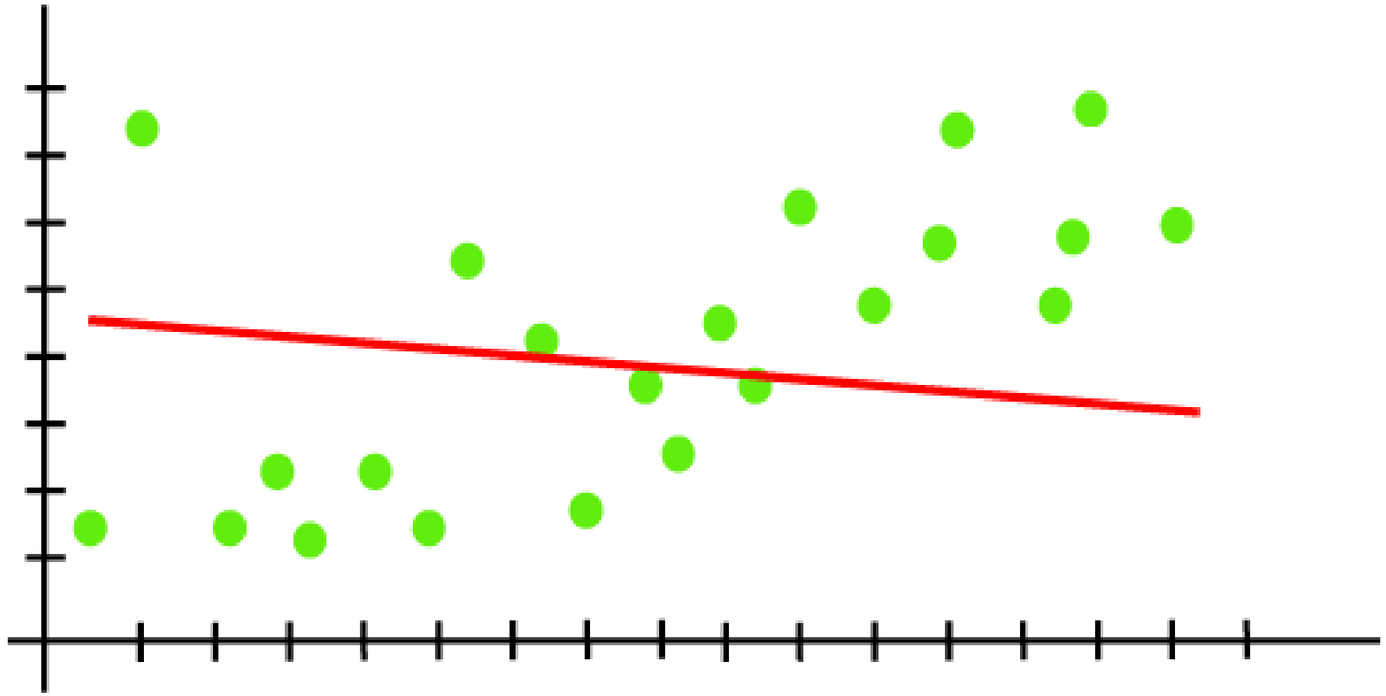
- Underfitting occurs when our machine learning model is not able to capture the underlying trend of the data. To avoid the overfitting in the model, the fed of training data can be stopped at an early stage, due to which the model may not learn enough from the training data. As a result, it may fail to find the best fit of the dominant trend in the data.
- In the case of underfitting, the model is not able to learn enough from the training data, and hence it reduces the accuracy and produces unreliable predictions.

By Shraddha Dalvi

Overfitting and Underfitting in ML

Underfitting

- Example: We can understand the underfitting using below output of the linear regression model:



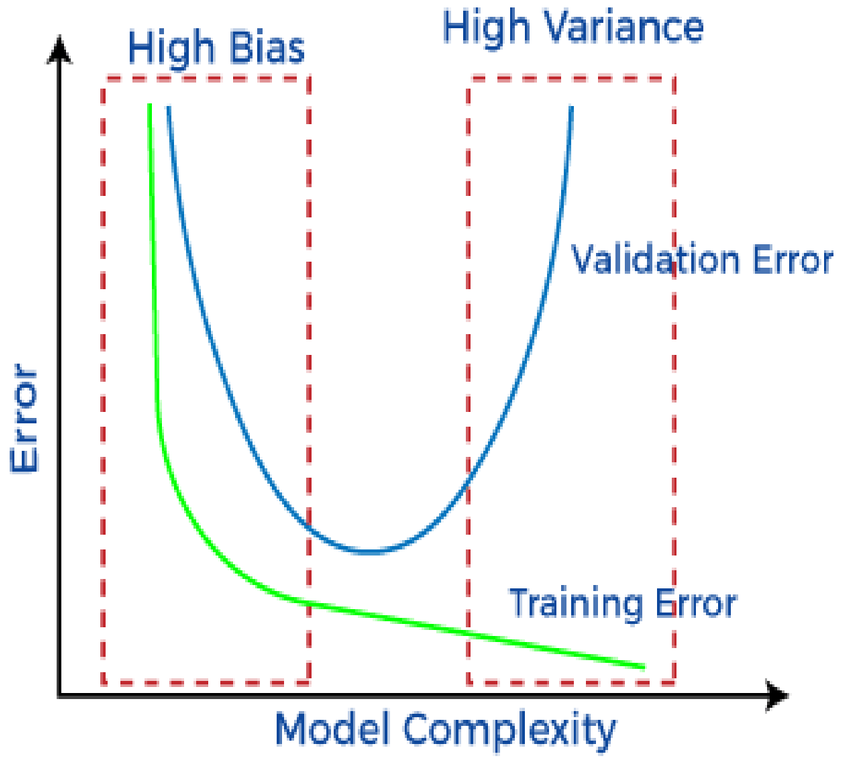
- As we can see from the above diagram, the model is unable to capture the data points present in the plot.

By Shraddha Dalvi

Bias and Variance in ML

Bias-Variance Trade-Off

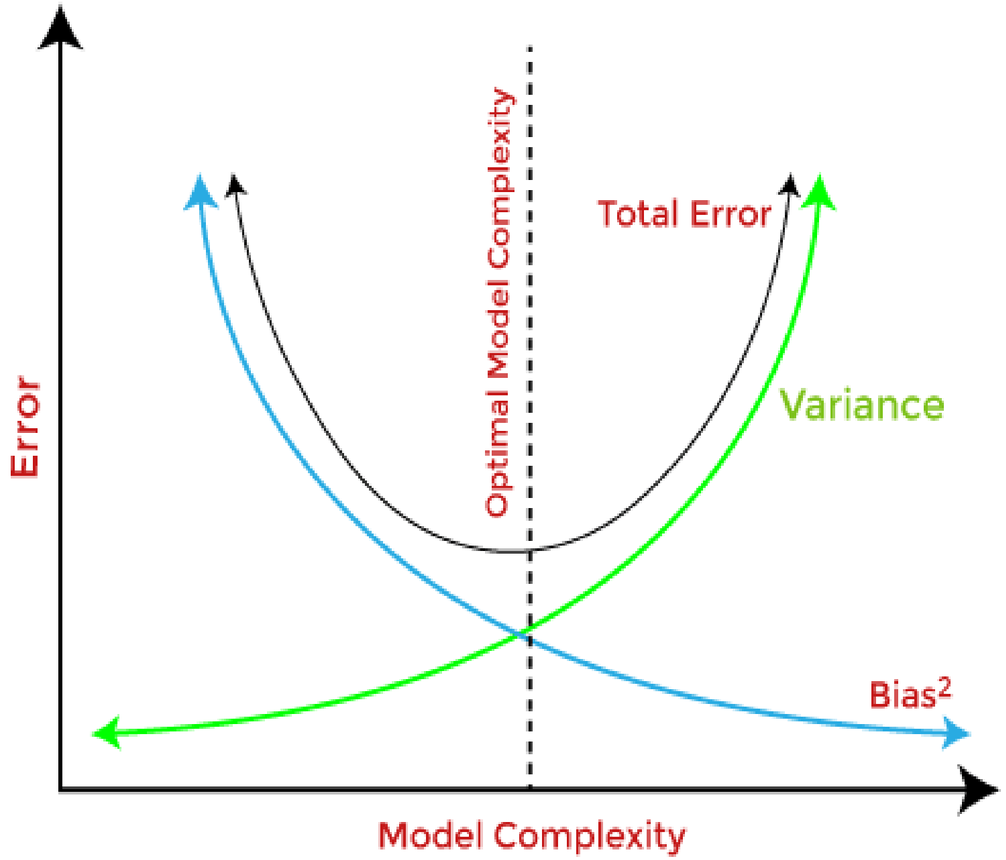
- While building the machine learning model, it is really important to take care of bias and variance in order to avoid overfitting and underfitting in the model.
- If the model is very simple with fewer parameters, it may have low variance and high bias. Whereas, if the model has a large number of parameters, it will have high variance and low bias.
- So, it is required to make a balance between bias and variance errors, and this balance between the bias error and variance error is known as the Bias-Variance trade-off.



Bias and Variance in ML

Bias-Variance Trade-Off

- For an accurate prediction of the model, algorithms need a low variance and low bias. But this is not possible because bias and variance are related to each other:
  - If we decrease the variance, it will increase the bias.
  - If we decrease the bias, it will increase the variance.



Hence, the Bias-Variance trade-off is about finding the sweet spot to make a balance between bias and variance errors. https://youtu.be/EuBBz3bI-aA?feature=shared

By Shraddha Dalvi