1.Variance and Bias (Diagram, overfit, underfit) - For best fit model should we have low bias or high variance, low bias or low variance, high bias or high variance, low bias or high variance.

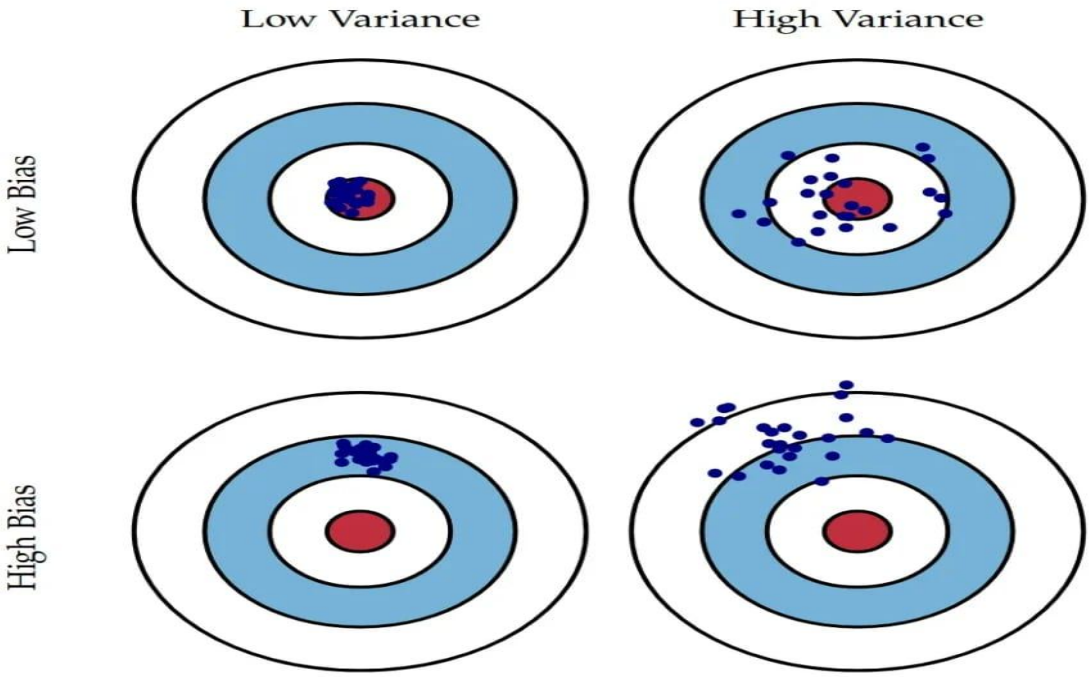


Fig. 1 Graphical illustration of bias and variance.

What is Bias?

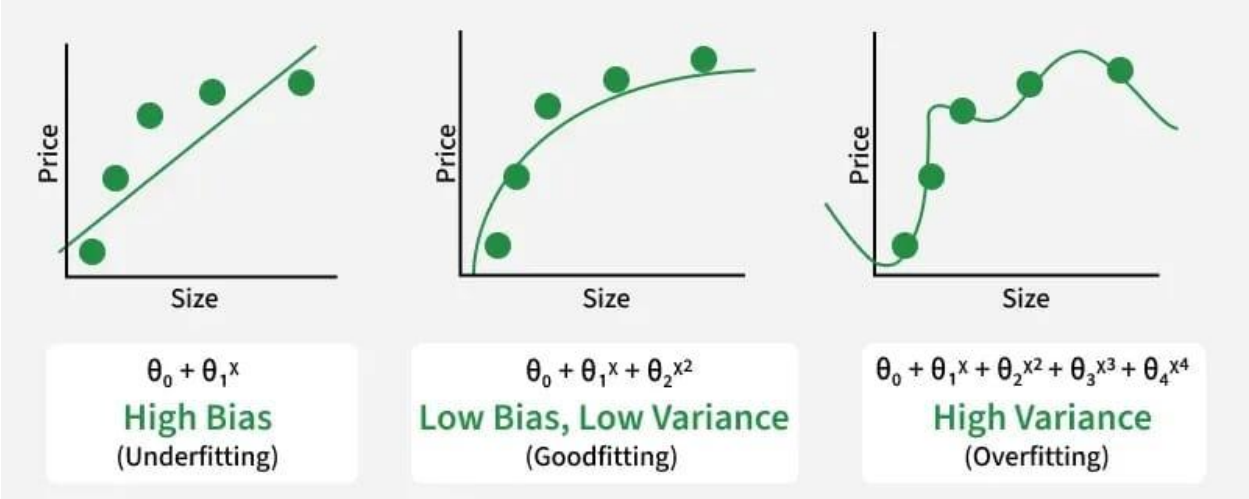
Bias is a disproportionate weight in favor of or against an idea or thing

or in machine learning, we can say bias is a disproportionate weight in favor of or against a feature. THE SAME THING WE HEAR EVERYWHERE.

What is variance?

Variance measures how far a set of numbers is spread out from their average value . AGAIN, THE SAME THING WE HEAR EVERYWHERE.

In easy words, the bias corresponds to training set error and variance corresponds to the test set error.



Overfitting and Underfitting.

What is Overfitted Model?

A model which performs really well on the training set or has high accuracy on training set but it does not perform well on the test set. So, as I told you bias is the training set error and variance is the test set error. Suppose, if we plot a point with new coordinates. Overfitted model might not perform better.

In case of Overfitting, our training set error is less, so it will have low bias and our test set error is high, so it will have high variance.

Overfitted Model — Low Bias and High Variance

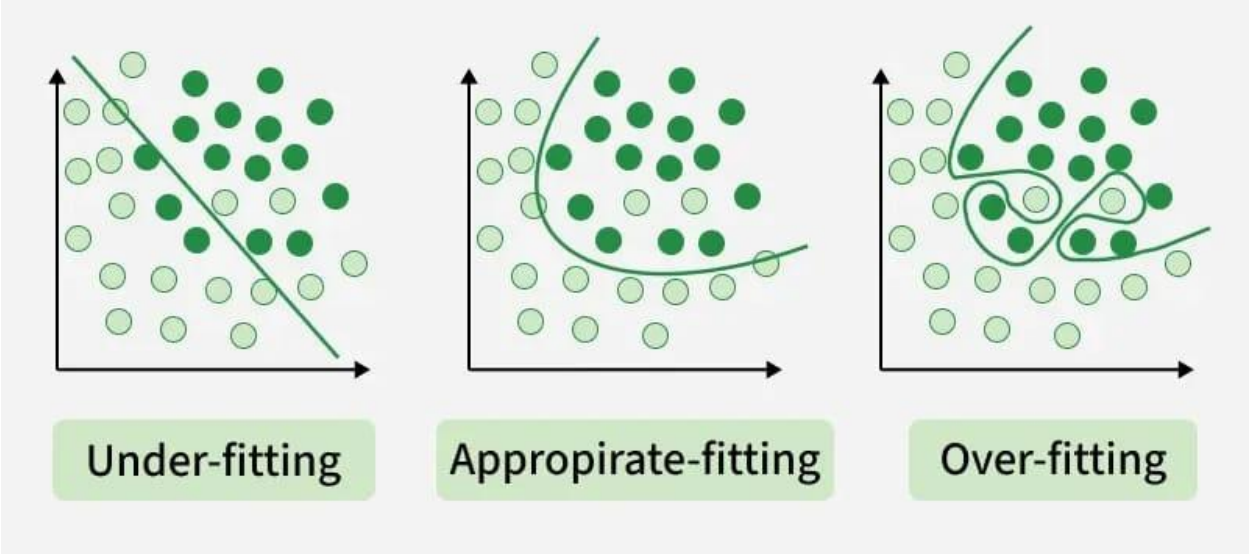
A decision is very prone to Overfitting. If we have a tree which is particularly deep. One way to solve this problem is pruning. But we will not discuss it here, we will only stick to the given topic :)

What is Underfitted Model?

A model which does not perform well on both training and test set. So, it's training error as well as test error is high, So it will have high bias and high variance.

In case of Underfitting, our training set error is high, so it will have high bias and our test set error is also high, so it will have high variance.

So, now we know what is Underfitted model and Overfitted model. We will now see what is Balanced model.



What is a Balanced Model?

A Balanced model is a model which performs well both on training and test set. This may not have as high accuracy as an overfitted model on a training set but a balanced model will perform well on test set as well.

A Balanced model will have low bias and low variance.

Let's take 3 examples to understand Overfitting, Underfitting and Balanced Model.

A model with training error : 2% and test error : 20%

Less Training Error — Low Bias

High Test Error — High Variance

This is an Overfitted Model.

A model with training error : 30% and test error : 30%

High Training Error — High Bias

High Test Error — High Variance

This is an Underfitted model.

A model with training error : 4% and test error : 3%

Less Training Error — Low Bias

Less Test Error — Low Variance

This is a Balanced model.