

## What Is Bias ?

- It is difference between Predicted Value and Actual Value.
- Bias refers to the error introduced due to assumptions made to model a relatively complex real-life problem by an approximate simple model.
- Low Bias: Predicting less assumption about Target Function(Predict Value >> Actual Value)
  - difference is Very Small Between predicted and Actual Value
- High Bias: Predicting more assumption about Target Function (Predict Value >>>>> Actual Value)
  - difference is very much large between predicted and Actual Value

Examples of low-bias machine learning algorithms include Decision Trees, k-Nearest Neighbors and Support Vector Machines.

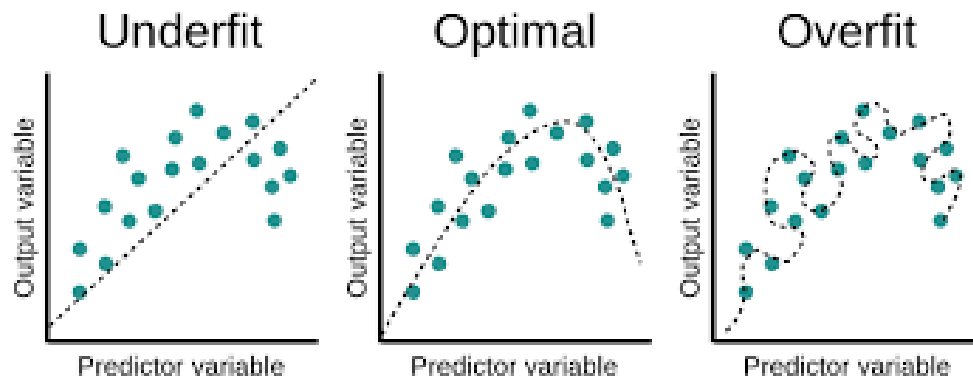
Examples of high-bias machine learning algorithms include Linear Regression, Linear Discriminant Analysis, and Logistic Regression.

## What Is Variance ?

- Variance is the amount that the estimate of the target function will change if different training data was used.
- It determine how spread of predicted value each other
  - Low Variance: Predicting small changes to the estimate of the target function with changes to the training dataset.
  - High Variance: Predicting large changes to the estimate of the target function with changes to the training dataset.
- Examples of low-variance machine learning algorithms include Linear Regression, Linear Discriminant Analysis, and Logistic Regression.
- Examples of high-variance machine learning algorithms include Decision Trees, k-Nearest Neighbors and Support Vector Machines.

**Till now we have some basic knowledge about Bias and Variance, now let's know some other important terminology used in Bias and Variance that are Underfitting and Overfitting.**

## Overfit And Underfit :



## Underfit :

- These models usually have high bias and low variance.
- underfitting happens when a model unable to capture the underlying pattern of the data.
- It happens when we have very less amount of data to build an accurate model or when we try to build a linear model with a nonlinear data. Also, these kind of models are very simple to capture the complex patterns in data like Linear and logistic regression.
- Here Training Accuracy is 50% But Testing Accuracy is around 49%. See there is less accuracy this tends to underfit.
- Underfitting is when the model's error on both the training and test sets (i.e. during training and testing) is very high.

### How to Oercome Underfitting :

- In that case, there are 2 gold standard approaches:
  1. Try another model
  2. Increase the complexity of the current model

Solution 1 is trivial. Concerning solution 2, an example an be the following: if someone is fitting a linear regression to some data, then increasing the complexity would mean to fit a polynomial model.

## Overfit :

- These models have low bias and high variance.
- 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 the noisy dataset.
- These models are very complex like Decision trees which are prone to overfitting.
- ■ Here Training Accuracy is 99% But Testing Accuracy is around 50%. there is huge difference between training and testing score this tends to overfit
- Overfitting is when the model's error on the training set (i.e. during training) is very low but then, the model's error on the test set (i.e. unseen samples) is large!

### How to overcome Overfitting :

- The most common problem in the ML learning filed is overfitting.
- Action that could (potentially) limit overfitting:
  1. We can use a Cross-validation (CV) scheme.
  2. Reduce the complexity of the model (make the model less complex).

- When it comes to solution 1 i.e. the use of cross-validation, the most famous CV scheme is the KFold cross-validation. Using a KFold scheme, we train and test your model k-times on different subsets of the training data and estimate a performance metric using the test (unseen) data.
- When it comes to solution 2 i.e. reducing the complexity of the model can help reduce the overfitting. For example, if someone is using an SVM model with RBF kernel then reducing the complexity would mean to use a linear kernel. In another case, if someone is fitting a polynomial to some data, then reducing the complexity would mean to fit a linear model instead (linear regression).

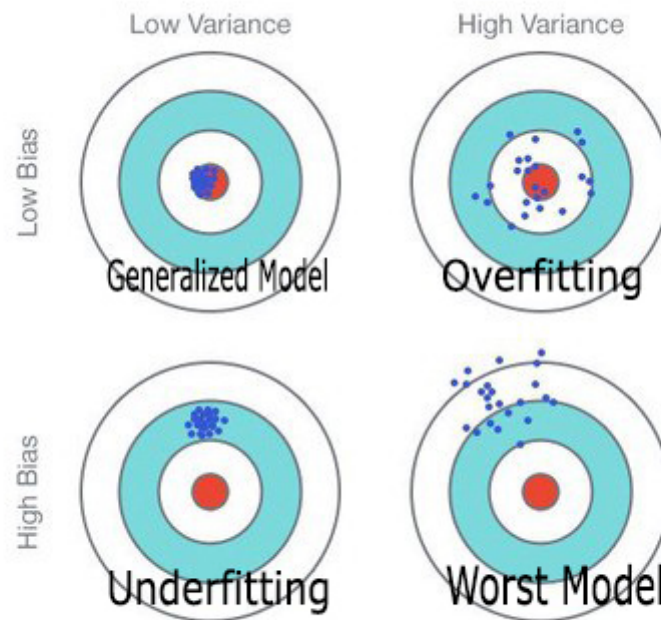
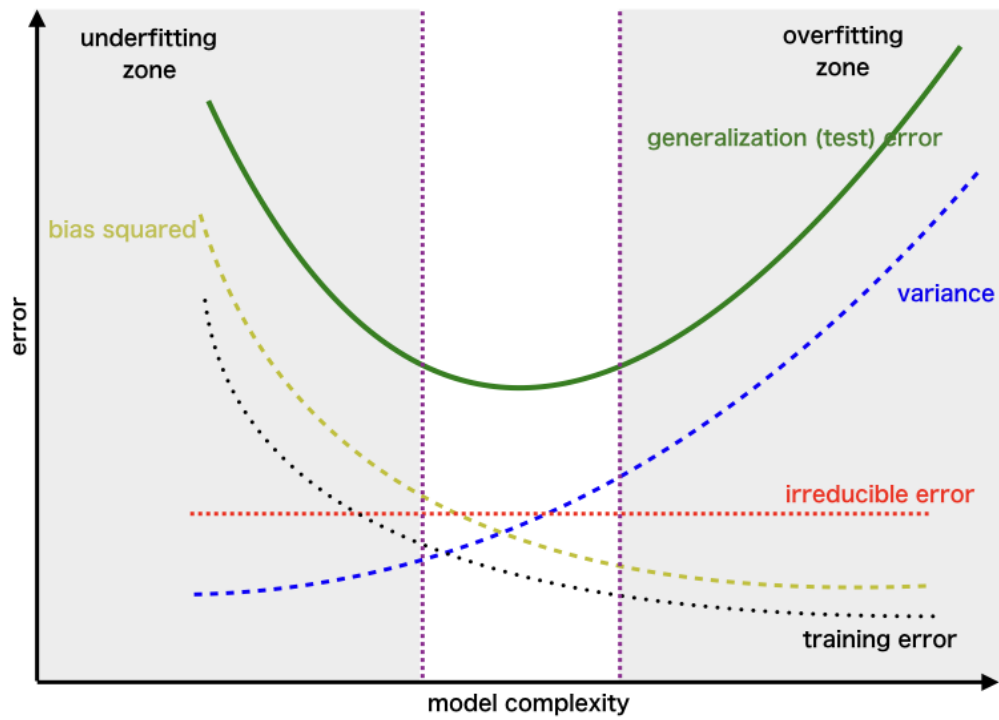


Fig. 1: Graphical Illustration of bias-variance trade-off , Source: Scott Fortmann-Roe., Understanding Bias-Variance Trade-off

## What is Bias-Variance Tradeoff ?

- If our model is too simple and has very few parameters then it may have high bias and low variance.
- On the other hand, if our model has a large number of parameters then it's going to have high variance and low bias.
- So we need to find the right/good balance without overfitting and underfitting the data.
- There is no escaping the relationship between bias and variance in machine learning.
  - \* Increasing the bias will decrease the variance.
  - \* Increasing the variance will decrease the bias.
- If the algorithm is too simple then it may be on high bias and low variance condition and thus is error-prone. If algorithms fit too complex then it may be on high variance and low bias. In the latter condition, the new entries will not perform well. Well, there is something between both of these conditions, known as Trade-off or Bias Variance Trade-off.
- The error to complexity graph to show trade-off is given as –



The above picture shows that :

- Bias initially decreases faster than variance.
- After a point, variance increases significantly with an increase in flexibility with little impact on bias.

This is referred to as the best point chosen for the training of the algorithm which gives low error in training as well as testing data.