**Bias Vs Variance**

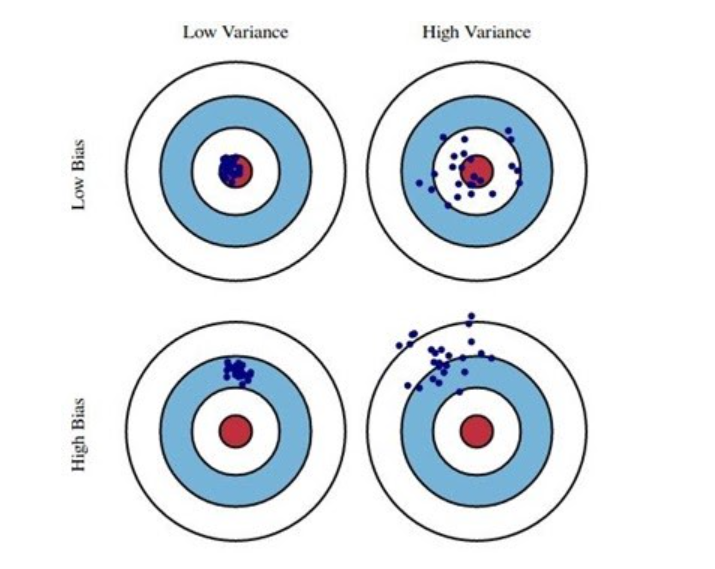
Bias:-Difference between predicted and actual value(Sum of squares),occurs due to assumption in models

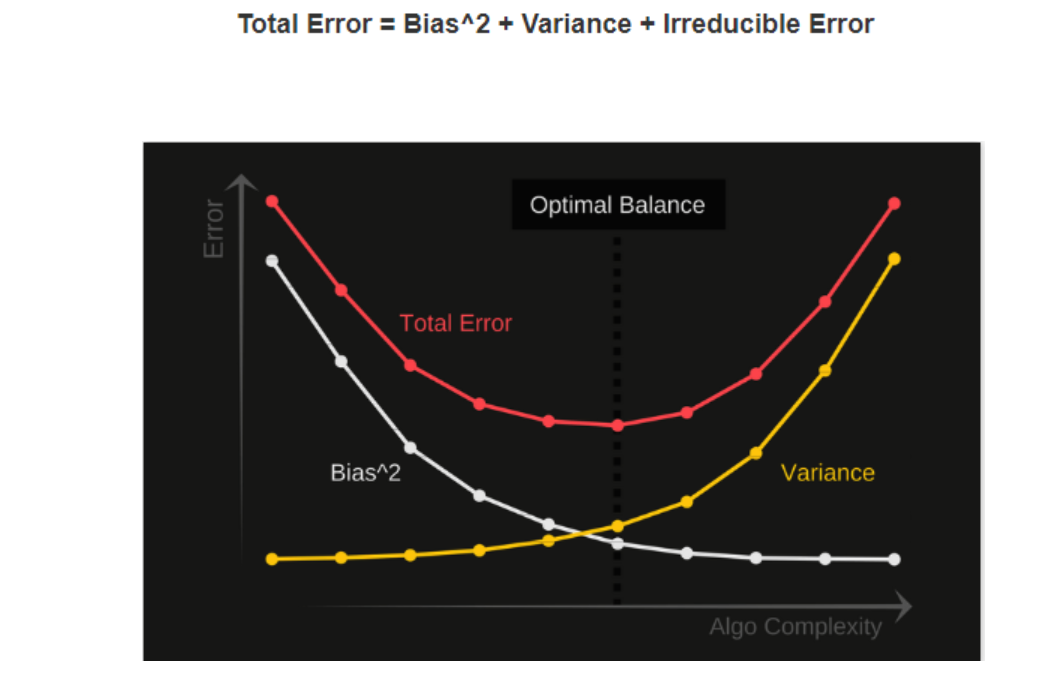
Variance:-Difference between fits across various data sets

High bias:-underfit

High Variance:-overfit

Good model has low bias and low variance





How to find balance between Bias and Variance:-

1. Dimensionality reduction
2. Regularization in linear model/ANN
3. Using Mixture model and ensemble learning
4. Optimal value of K in case of K NN.

**Cross Validation**

In **K Fold cross validation**, the data is divided into k subsets. Now the holdout method is repeated k times, such that each time, one of the k subsets is used as the test set/ validation set and the other k-1 subsets are put together to form a training set.

This significantly reduces bias as we are using most of the data for fitting, and also significantly reduces variance as most of the data is also being used in validation set

a slight variation in the K Fold cross validation technique is made, such that each fold contains approximately the same percentage of samples of each target class as the complete set, or in case of prediction problems

After that we apply cross validation to different model and check which one is more accurate.

**Bootstrapping**

Training records are sampled with replacement

The bootstrap method is a resampling technique used to estimate statistics on a population by sampling a dataset with replacement.

samples are constructed by drawing observations from a large data sample one at a time and returning them to the data sample after they have been chosen. This allows a given observation to be included in a given small sample more than once. This approach to sampling is called sampling with replacement.

* Choose the size of the sample.
* While the size of the sample is less than the chosen size
  + Randomly select an observation from the dataset
  + Add it to the sample

This is done by training the model on the sample and evaluating the skill of the model on those samples not included in the sample. These samples not included in a given sample are called the out-of-bag samples, or OOB for short.

The samples not selected are usually referred to as the “out-of-bag” samples. For a given iteration of bootstrap resampling, a model is built on the selected samples and is used to predict the out-of-bag samples.

A useful feature of the bootstrap method is that the resulting sample of estimations often forms a Gaussian distribution(bell shaped curve)

Parameters: -

* + Sample size: -Usually same as dataset size, if data too large then sample size 50%,60%
  + Repetition: - The number of repetitions must be large enough to ensure that meaningful statistics, such as the mean, standard deviation, and standard error can be calculated on the sample.

**Ensemble Learning**

Combine individual models together to improve stability and predictive power of model. Each model

Is better in some aspects of data.

Decrease variance using bagging

Decrease bias using boosting

Improve prediction using stacking

Two Types

* + Sequential learner: -Base learners are generated consecutively. Basic motivation is to use dependence between base learners
  + Parallel learner: - Basic motivation is to use

In case of ensemble prediction, the final prediction is carried out by averaging the all model prediction

In case of classification, it is calculated using the mode of model classifications.

The class probability is calculated using argmax of the summed probabilities for each class label

**Bagging**

* Create random sampled datasets of original training dataset(bootstrapping)
* Build and fit several classifiers to each of these diverse copies
* Take average of all predictions to make final overall predictions

Random forest combines various decision trees to produce a more generalized machine learning model. It creates random subsets of features. It overcomes overfitting.

**Boosting**

Boosting reduces bias by training weak learners sequentially, each trying to correct its predecessor. Changing weak learners to strong learners.

Example:-

**Adaboost**

* It helps mixing multiple weak classifiers to one strong classifier. Assign equal weights to each data points and apply a decision stump to classify as + or –.
* Only binary classification problems are supported.
* Firstly, it selects a training data randomly. It iteratively trains Adaboost machine model. It assigns higher weight to a wrongly classified observation.
* It assigns weight to the trained classifier in each iteration according to the accuracy of the classifier.

**Gradient Boosting**

* GBM minimizes the loss function of a model by adding weak learners using gradient descent procedure
* Consists of loss function, Weak learner to make prediction , additive model to add weak learners to minimize loss functions
* Each new model gradually minimizes the loss function of the whole system
* GBM uses the boosting technique, combining a number of weak learners to form a strong learner.
* Fit the model, calculate error residuals.
* Fit the new model on error residuals as target variables with the same input
* Add the predicted residuals to previous predictions, continue to fit until sum of residuals become constants or model overfits