# TYPES OF ERRORS

There are mainly two types of errors

## UNDERFITTING

* Oversimplifying the problem
* Leads to misclassification in training data
* Error due to bias

## OVERFITTING

* Overcomplicating a problem
* Leads to misclassification in test data
* Error due to variance
* Tends to memorize rather than learning the charactersticks

So when choosing the model we should neither underfit nor overfit ,this is done by plotting the training and testing errors,but since we cant use testing data for training we should split the training data to training and cross validation sets and then compute cross validation errors

The point where both the errors are minimum is the viable model

## K-FOLD CROSS VALIDATION

* Divide training set into ‘k’ buckets
* Train the model ‘k’ times each time using a different test set

In Underfitting models the plotted training errors and cross validation errors converge at a high point while in overfitting models they do not converge at all and in perfect models they converge at a lower point

## HOW TO CHOOSE THE MODEL

* First we train the different models using training set and validate it using cross validation set
* Then we calculate f1 score and choose the model with the highest f1 score
* Then we guess the parameters by a grid search where we’ll validate it using all the parameters and choose the one with the highest f1 score