Github link:<https://github.com/sumanth1904/ice_7>

Overfitting occurs when a model learns the detail and noise in the training data. It degrades the model's performance on fresh data. Underfitting occurs when the model does not able to learn properly from the training data. Therefore making it less accurate. A good fit is somewhere between overfitting and underfitting. A good fit eliminates noise in the data set, considering only the important datapoints. By splitting the dataset properly into test and train data help reduce underfitting and overfitting problems. Apart from that increasing training data reduces overfitting.increasing number of features and removing noise from data reduce underfitting. Since the model is trained on training data multiple times and this part of dataset is familiar to it training accuracy is generally greater than testing accuracy. But in case of overfitting training accuracy is always closer to 1 and since the model considers noise in training it negatively impacts its performance new data resulting in lower testing accuracy. Cross validation is used for statistical assessment of performance of machine learning models.Cross-validation is a procedure that is used to avoid overfitting and estimate the skill of the model on new data. There are common tactics that you can use to select the value of k for your dataset.ROC provides graphical representation of a classifier’s performance. It is produced by calcualating true positive rate against false postive rate. It is used find classification threshold or k that suits our problem. By looking at validation curve we can know if our model is underfitting ,over fitting or is a good fit.