Name: Shohan Nayak

Roll No.: 13030823138

**Part A**

1. Classification in Regression are the two most common supervised tasks.
2. To check the accuracy of the prediction of a validation set.
3. There are two model parameters are there in a linear regression problem with a single feature variable.
4. AUC value of a perfect classifier is 90-100%.
5. Precision is more important for a spam email detection system.

**Part B**

1. train-test-split is a method available in sklean that we used to split the date into training set and testing set. We use 30%. of the total data.

Underfitting and Overfitting:

Overfitting is a phenomenon where the machine learning model predicts two well on the training data that is seen data that does not work well on unseen data that is testing data.

Where underfitting happens when the made performs tradly on the training data that is the seen data that predicts too well on the unseen date that is testing dataset.

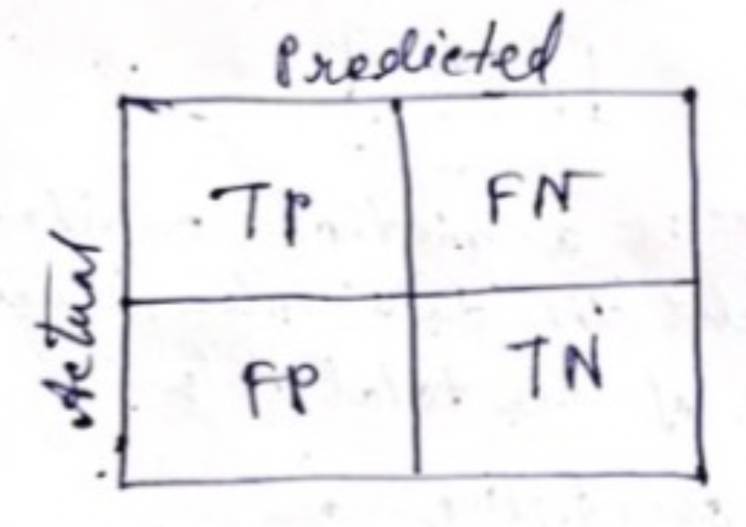
Prevention technique:

1. Use of suitable model
2. Scale the data
3. Reduce the noice from the data
4. Cross-validation
5. bias – Bias is a terminology in machine learning. It signifies how a machine learning model predicting poorly in the training phase.

variance – Variance is where the model gives error in testing phase.

* To reduce it we take date that is properly encoded, scalled and the data is moderate.
* Bias-Variance tradeoff is a scenario where if the model perform poorly in the training phase.

1. Confusion matrix is the insight of model performance. It include correctly predicted and incorrectly predicted data value.



TN = 82, FP = 3, FN = 5, TP = 10

precision =

recall =

false negative rate =

false positive rate =