TECHNO MAIN SALT LAKE

(FORMERLY TECHNO INDIA, SALT LAKE)

Name Harshamlit Raj

Roll No. 13030822007 Stream CSE(AI&ML)

Subject Application of Machine Learning in Industries (PCC-AIML601) Semester 6th

Invigilator's Signature Date

Part A:

1) The two most common supervised tasks are classification and regression.

2) The purpose of validation set is to validate the performance & check the accuracy of the given trained model.

3)

4) For a perfect classifier the AUC value is always 1.

5) Out of precision & recall, recall is more important because for a spam email detection system it gives the ratio of True Positive by True Positive + False negative which gives the actual true value.

Part B:

6) Traintestsplit is a function for splitting a whole dataset into two broad set i.e., training set & testing set. It called to select following data & split those datapoint into the two sets training & testing set.

Overfitting & Under fitting are the two problems faced during training of a model. In overfitting the trained model perform well on training set but fails in performing on testing data. Meanwhile, in underfitting it occurs due model failing to perform in a training set.

For preventing overfitting & underfitting in a training model we have to perform following steps. For overfitting we have to handle the missing values and also supply adequate amount of datapoints. While in underfitting we have to generate synthetic data and provide an adequate amount of data to the training set to perform well during training of model.

7) Bias & variance of a machine learning model are the most crucial parts for the performance of a learning model by which accuracy is totally dependent.

In Bias, when the prediction is bent towards one side of a graph i.e., either positive or negative,

In variance, when the deflection of output is slightly changes with the input.

When the datapoints are near to eachother then the biasness of model can be changed. By manipulation of datapoints i.e., preprocessing we can reduce both bias & variance of a model.

When both bias & variance are closed to eachother we can see a bias-variance trade off.

9) Confusion Matrix is used to display the accuracy of a training model by showing the true positives, true negative, false positive & false negative. Of a model.

It is important because by the confusion matrix we are able to predict the the accuracy of a trained model.

Precision = $\frac{10}{10+3} = \frac{10}{13}$

Recall = $\frac{10}{10+5} = \frac{10}{15}$

False negative rate = $\frac{5}{5+82} = \frac{5}{87}$

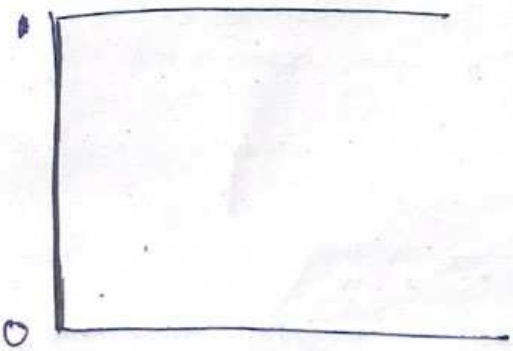
False positive rate = $\frac{3}{3+5} = \frac{3}{8}$

10) ROC & AUC curve are the crucial parameters for determination of a training model.
ROC is inversely proportion to AUC i.e.,

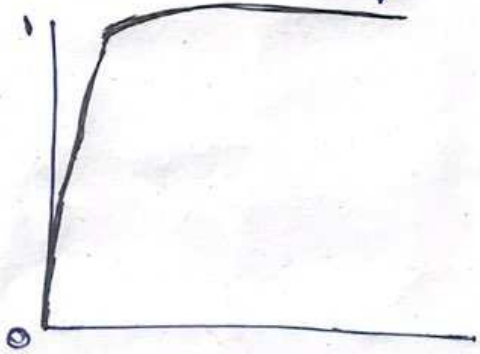
$$ROC \propto \frac{1}{AUC}$$

AUC is area under curve i.e, the graph plotted by ROC & the area which is covered by the ROC is its AUC.
They are used for showcasing the performance of a model.

perfect classifier —



practical classifier



random classifier