# TECHNO MAIN SALT LAKE

( FORMERLY TECHNO INDIA, SALT LAKE )

Name.....Aditya Chakraborty.....

Roll No. .....13020822040.....Stream.....AIML A.....

Subject .....Machine Learning.....Semester.....6th.....
(PCC-AIML601)

Invigilator's Signature .....Date.....

Part A

1. Customer - Churn Analysis and Image Recognition

2. Purpose of a validation set: -
   - After analysing the data through of train set and testing it through test set for multiple times, to find the final accuracy validation set is needed.

3. - Co-efficients of the variable
   - Loss function i.e. Mean Squared Error (MSE).
   - Gradient Decent.

4. The AUC of a perfect classifier is: -

$$\frac{y_2 - y_1}{x_2 - x_1} = \tan s$$

$\therefore \tan s \geq 1$.

5. Out of Precision and Recall, recall is more important in Spam Detection.

Part B.

6. When we get a raw dataset we divide the dataset initially into train-test-split i.e. 70-80%. of data are kept in initially for training and after training the model, it gets tested on the rest 20-30% data i.e. known as the data set. So, if we sum up the train test they we get the train test split.

Overfitting of training data takes place when the ~~best data set~~ train data set exceeds the number of test data sets, overfitting occurs. Overfitting can be prevented by splitting up of data during preprocessing in order to execute each dataset during single processes.

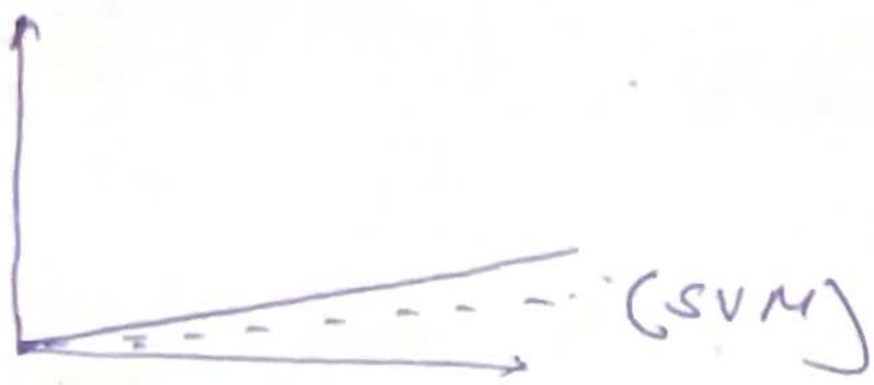
Underfitting occurs when the test data set is less than the train data set. So much is planned in order to reduce this phenomenon.

8. Several Algorithms that are utilized to minimize the cost function. —
   - Naive Bayes Algorithm
   - Random Forest Algorithm
   - Decision Tree

9. Confusion Matrix. — it is used to see the generation of matrix data in order to analyse the statistical regression of a data set so that the model data is plotted and can be compared to that of a trained model data to compare the accuracy between the data sets.

False negative rate — 5%.
False positive rate — 3%.
Precision :- 92%.
Recall — 0.63

10. ROC stands for Radius of a curve and AUC stands for Area under curve, where ROC is used to see the radial point of the curve with respect to its origins and AUC is the total area of the respective curve.


(~~poor~~ practical classifier)


(Random classifier)

Precision Recall trade off - It is the single process that is executed multiple times in order to achieve precision of the whole data set execution.