TECHNO MAIN SALT LAKE

(FORMERLY TECHNO INDIA, SALT LAKE)

Name.....Karan Mondal.....

Roll No. 13030822016 .....Stream.....CSE-AIML.....

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

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

1. The two most common supervised tasks are —
Image Recognition and customer-churn Analysis.

2. The purpose of a validation set —
After analysing the data through trainset and testing it through test set for multiple times to find the final accuracy validation set is needed.

3. Model parameters for linear regression model —
(i) coefficients of the variable.
(ii) Loss function i.e. mean squared error (MSE)
(iii) Gradient Decent.

4. AUC value of a perfect classifier - town ≥1

5. out of precision and recall, recall is more important - for a spam email detection system.

6. Train-test-split :-
when we get a new dataset we devide the dataset initially into train-test split. i.e. 70-80% of data are kept initially for training and after training the model get tested on the rest 20-30%. data. i.e. known as test set. so if we sum up the whole train test theory we get the train test split.
overfeeting of ~~training~~ training dataset takes place when the train data set executes the number of test dataset underfitting occours ~~eee~~ overfitting where the test data sets is fitting so overfitting and underfitting of training data.

8. General algorithms that are available to minimize the cost function —
Naives Bias algorithm
Random forest algorithm.
Decision tree.

9. confusion matrix is said to be the generation of matrix data inorder to analyse the statistical regression of a data set.
In classification problem, true negative = 82, false positive = 3,
false negative = 5
true positive = 10,
precision, recall, false negative rate.

10. ROC is a type of curve that is use for perfect classifier, practical classifier and random classifier.