TECHNO MAIN SALT LAKE

(FORMERLY TECHNO INDIA, SALT LAKE)

Name Swannali Pramanik

Roll No. 13030823136 Stream CSEAIML

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

Invigilator's Signature ..... Date 24/03/25

Part A

1. The two most common supervised tasks are Regression and classification.
2. The purpose of a validation set is used after training the model. It is used to evaluate the model performance and tuning the model.
3. 2 model parameters are there in a linear regression problem with a single feature variable.
4. The AUC value of a perfect classifiers is 1.
5. precision is more important for a spam email detection system.

Part B

6) train-test-split:

The dataset is splitted into training, testing and validation. The model trained on training set, then the model is tested on unseen test dataset.

Overfitting:

When a model learn noise from data the training set and the model performed well in training set but it does not perform well on unseen data. This indicates overfitting.

Underfitting:

When a model is unable to learn from the training set and the model perform does not well on training set. This called underfitting.

Prevent them:

a) use a good model.
b) Reduce the noise from data.
c) Scale the data.
d) Cross-validation.

2) Confusion matrix:

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

It is important:

a) It can check accuracy from matrix.

b) It can calculate Entropy.

c) It also can calculate precision, recall, f-score.

$tn = 82$
$fp = 3$
$fn = 5$
$tp = 10$

$$\text{precision} = \frac{TP}{TP+fP} = \frac{10}{10+3} = \frac{10}{13} = 0.769$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{10}{10+5} = \frac{10}{15} = 0.666$$

$$\text{false negative rate} = \frac{FN}{TP+FN} = \frac{5}{10+5} = \frac{5}{15} = 0.333$$

$$\text{false positive rate} = \frac{FP}{fp+TN} = \frac{3}{3+82} = \frac{3}{85} = 0.035$$

$= 1 - \text{precision}$
$= 1 - 0.769 = 0.231$

7) In machine learning model,

Bias is a ~~training~~ terminology in machine learning. It signifies a training model performing poorly in training phase.

variance is where the model mostly give some error in testing phase.

It reduces -

a) to reduce, it take data that it is properly scalled and the size of data is measured.

b) to reduce variance, it can use dimensionality reduction.

Bias-variance tradeoff is a scenario where the model perform poorly in the training phase.