Name: Swarnali Pramanik

Roll No.: 13030823136

**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 classifier is 1.
5. Precision is more important for a spam email detection system.

**Part B**

1. 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 data from the training set and the model performed well on the 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 is called underfitting.

Prevent them :-

1. Use of good model
2. Reduce the noise from data
3. Scale the data
4. Cross-validation
5. Confusion Matrix:

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

**It is important:**

1. It can check accuracy from matrix.
2. It can calculate Entropy.
3. It also can calculate precision, recall and F-score.

tn = 82

Precision

fp = 3

fn = 5

tp = 10

Recall

false negative rate =

false positive rate =

1. In machine learning model,

Bias is a 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:

1. To reduce, it take data that it is properly scalled and the size of the data is measured.
2. To reduce variance, it can use dimensionally reduction.

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