Name: AYUSH KUMAR DUBEY

Roll No.: 13030822014

**Part A**

1. Two most common supervised tasks are -

* Regression
* Classification

3. There are 2 model parameters in a linear regression problem with a single feature variable.  
They are intercept and co-efficient.

2. The main purpose of validation set is that it is used to evaluate the performance of a model.

5. Precision is more important for a spam e-mail detection.

4. The AUC value of a perfect classifier is: 1.

**Part B**

6). Train-test-split is a model that is used to divide the dataset into training dataset and testing dataset

Implementation: From sklearn.model\_selection, import train\_test\_split, x\_train, x\_test, y\_train, y\_test = train\_test\_split (x, y, test\_size = 0.2, random\_state = 42)

Overfitting is a problem that occurs in a training model where it can predict accurately for data in a dataset but fail on real world data.

Underfitting makes the training model such way that it can predict accurately for real world data

Prevention:

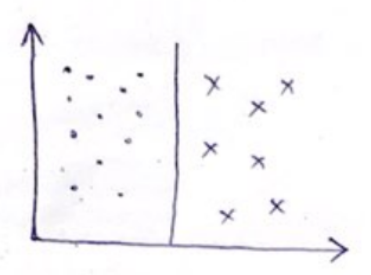
* Prunning,
* Using right machine learning algorithm,
* Scaling of data.

8). Linear Regression is supervised algorithm where output a dependent variable is found from one or more independent variable

a → intercept, b → slope  
 y → dependent variable  
 x → independent variable

The cost function / Loss function for linear regression is :

Logistic regression is a supervised algorithm where it is mainly used for the classification, i.e. classifying the data into their respective classes.



The cost function / loss function of Logistic regression is:

The General Algorithm by which lort function are minimized are Lasss Regression & Ridge regression

Here is absolute value of .

Here is a squared value of .