TECHNO MAIN SALT LAKE

(FORMERLY TECHNO INDIA, SALT LAKE)

Name AYUSH KUMAR DUBEY

Roll No. 13030822014 Stream B.TECH CSE (AIML-A)

Subject ~~Application of Machine Learning~~ in Industries Semester 6th

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

PART-A

1). Two most common ~~of~~ 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 the Validation set is to make the given datasets valid and to make them useful for the further functions.~~

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 sham 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)

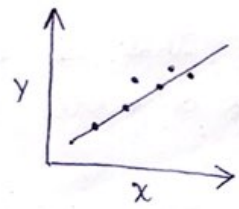
Overfitting is a problem that occurs ~~it~~ in a training model where it can predict accurately for data in a dataset but fail on real world data.

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

Prevention:

- Pruning
- Using right machine learning algorithm
- Scaling of data

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



$Y = a + bx + \epsilon$

$a \rightarrow$ intercept, $b \rightarrow$ slope

$Y \rightarrow$ dependent variable

$x \rightarrow$ independent variable

The cost function / ~~class~~ Loss function for linear regression is: $L = \sum_{i=1}^{n} (y_i - \hat{y}_i)$

$y_i \rightarrow$ Actual output

$\hat{y}_i \rightarrow$ Predicted output

P.T.O

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

The cost function / Loss function of Logistic regression is:

$$L = y_i \log(\hat{y}_i) + (1 - y_i) \log(\hat{y}_i)$$

$y_i \rightarrow$ Actual output

$\hat{y}_i \rightarrow$ Predicted output

The General algorithm by which cost function are minimized are Lasso Regression & Ridge regression

For Lasso : $L(\text{Lasso}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda |m|$

Here is absolute value of m.

For Ridge : $L(\text{Ridge}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda (m^2)$

$\rightarrow \lambda$ (alpha) is a hyperparameter

Here is a squared value of m.