Assignment 1

Tuesday, 6 December 2022 12:42 AM

Submitted by → Sarthak Chauhan
Enrollment no. → MTE4220003

Questions solved in this Assignment.

→ T1 Question Paper Solution

Q1 a) Linear Regression can be used to predict the stock prices of NIFTY. LR is a type of supervised learning technique.

b) Parameters used to predict the prices can be

① opening price of a stock
② highest price possible at any time.
③ lowest " " " " ".
④ Closing " of stock.

c) Yes, there is a limitation with Linear Regression. It is limited to linear relationships only. By its nature linear Regression only looks at linear relationship between dependent & independent variable & ignores any non-linear relationship.

Q2 There are 5 stages in machine learning project life cycle.

① Project Initiation

- this step involves 3 major things
  - Idea
  - Requirements
  - Data Aquisition.
- Idea is developed & analysed so that the project requirements are clear & data can be aquired which basically is the back-bone of any ML project.

② Data Exploration: here in this phase the data aquired is explored in depth to understand the nature of the data, the kind of relationship its attributes hold etc.

③ data preprocessing & feature selection

- this phase involves
  - Removing irrelevant features from dataset.
  - handling NULL or missing values.
  - Reducing the size of dataset.
  - transform categorical variable to numerical.
  - Normalising data points.

④ Model Development

now the actual model is developed while choosing the algo we consider:

- Data size
- Type of problem
- Availability of packages.

⑤ Model Evaluation

- In this phase the model is evaluated based on its prediction accuracy by multiple metrics, such as precision, recall, F1-score. etc.

Q3

X	Y	distance from 6.5
1	23	5.5
1.2	17	5.3
3.2	12	3.3
4	27	2.5
5.1	8	1.4
6.5	?	0

If $K = 5$ then all are selected. as NN.

then prediction is $\frac{23+17+12+27+8}{5}$

$= 17.4$

Q4 Advantages → ① No training period.
② & also due to that data can be added to the dataset at any time
③ Very easy to implement.

Dis-Advantages → ① Does not work well with large datasets.
② Does not work well with v. large datasets.
③ Need feature scaling.
④ Sensitive to missing noisy data & outliers.

→ best way to use KNN so it works best is by choosing the right value of k.