NAME: Raman Kumar

REG NO: BL.EN.P2DSC22009

PROGRAM: MTech Data Science

COURSE: Machine Learning

Practical Session: 01 (Report)

Answers

Q.1. Discuss the importance of the Rank of an observation matrix in model building for classification.
Dataset: Purchase data

The rank of the observation matrix can tell you if it is possible to find a unique solution for your model. The rank of the matrix tells us the total linearly independent rows or columns in the observation matrix.

In the dataset, the rank of the observable matrix (A) was 3 equivalent to the number of unknowns. By finding the linearly independent rows (columns) you can observe which rows or columns in the matrix are removable or not. In long term having unnecessary columns or features are not very efficient.

In the given dataset of 'Purchase Data,' I took a 3*3 matrix initially. After finding the rank I observed that all three columns i.e features were equally needed. So, If the operation applied had given me rank two for the observable matrix then it has to be me who decides whether to delete that row(column) or not.

In another case, we need the rank of an observation matrix equally likely to the number of unknowns as we do not want only a significantly less number of features in our dataset. The rank of the matrix should always be near the number of unknowns i.e the initial condition for any model building.

Q.2 Discuss on regression (Ex: A2) and classification (Ex: A3) tasks. How would you differentiate between them.

In this given dataset of Purchase data, I used *a classification model i.e* task *A.3*, where the model is predicting a label that belongs to one of two classes i.e known as binary classification. In this exercise, l trained a binary classifier to predict whether or not a person belongs to the 'RICH' or 'POOR' category (labels) based on some payment done to the purchase. I distinguished or mapped the Category of rich or poor based on payment to binary values 0 and 1.
So, based on the results of trained data I found the test outputs and found fascinating questions:

- Of all the customers the model predicted are RICH i.e 1, how many are actually RICH?
- Of all the customers that are actually Rich or Poor, how many did the model identify?

I also applied the regression model for the given dataset, in which we use the mapping functions
to assign the values to continuous outputs. Regression algorithm used to predict continuously.
the variable is given a set of independent variables because it makes certain assumptions based on the already trained data set.

Q.3 Observing the stock data provided, record your suggestions to build a system that
may be able to predict the price and Change % in the future.
Dataset: IRCTC Stock Price

Stock Price data have much uncertainty in their behavior and are known for non-linear structures. On the given dataset we can not use the classification model, because this data needs to learn but the classification model cannot give that accurate learning inputs to the stock data. We need to find the evident pattern in the data, and then the algorithm to learn from it and derive or predict the useful values from seen data. For that exact purpose regression model will be helpful. I used the linear regression model to build a system that predicts the price of stock data and the respective change to the previous input.