Subject: 22AIE213

Lab Session: 03

Notes:

1. Please read the assignment notes carefully and comply to the guidelines provided.
2. Code should be checked into the GitHub. These details shall be provided in the Lab.
3. If you have not completed the prerequisite assignments, please complete them before the next lab session.

Coding Instructions:

1. The code should be modularized; The asked functionality should be available as a function. Please create multiple functions if needed. However, all functions should be present within a single code block, if you are using Jupyter or Colab notebooks.
2. There should be no print statement within the function. All print statements should be in the main program.
3. Please use proper naming of variables.
4. For lists, strings and matrices, you may use your input values as appropriate.
5. Please make inline documentation / comments as needed within the code blocks.

Mandatory Section

A1. Please refer to the "***Purchase Data***" worksheet of ***Lab Session1 Data.xlsx***. Please load the data and segregate them into 2 matrices A & C (following the nomenclature of AX = C). Do the following activities.

- What is the dimensionality of the vector space for this data?
- How many vectors exist in this vector space?
- What is the rank of Matrix A?
- Using Pseudo-Inverse find the cost of each product available for sale.
  (Suggestion: If you use Python, you can use ***`numpy.linalg.pinv()`*** *function to get a pseudo-inverse.)*

A2. Use the Pseudo-inverse to calculate the model vector X for predicting the cost of the products available with the vendor.

A3. Mark all customers (in "***Purchase Data***" table) with payments above Rs. 200 as RICH and others as POOR. Develop a classifier model to categorize customers into RICH or POOR class based on purchase behavior.

A4. Please refer to the data present in "***IRCTC Stock Price***" data sheet of the above excel file. Do the following after loading the data to your programming platform.

- Calculate the mean and variance of the Price data present in column D.
  (Suggestion: if you use Python, you may use ***`statistics.mean()`*** *&* ***`statistics.variance()`*** *methods).*
- Select the price data for all Wednesdays and calculate the sample mean. Compare the mean with the population mean and note your observations.
- Select the price data for the month of Apr and calculate the sample mean. Compare the mean with the population mean and note your observations.

- From the Chg% (available in column I) find the probability of making a loss over the stock. *(Suggestion: use lambda function to find negative values)*
- Calculate the probability of making a profit on Wednesday.
- Calculate the conditional probability of making profit, given that today is Wednesday.
- Make a scatter plot of Chg% data against the day of the week

Optional Section:

O1. Create 2 separate square matrices from the purchase data matrix. Repeat experiments A2 & A3 with both these matrices. Do the X values obtained from the square matrices match to the one obtained from the whole purchase data matrix?

Report Assignment:

1. Write your understanding of your project in the introduction section of the report.
2. Download at least 10 published papers (from IEEE Xplore, Springer, Elsevier or Science Direct) for your project. Study these papers use them for literature survey section of your report.
3. Using the learnings so far, design a system that could be used for customer / patient segmentation. Enrich your answer with:
    a. Flow diagram to depict the data flow. Example: input handling, preprocessing, similarity scoring, output.
    b. Architecture diagram for the system should be in methodology or system description section. Detail what happens in each block.
    c. define parameters to be used in the system; assign values for these parameters and justify them.

    Since this is a Design work, the solution should be provided in the methodology section of the IEEE format of report. The results may be taken from above experiments and discussed to conclude the paper.