Complex Engineering Activity (CEA-2)

Computer Application in Engineering Design Lab
Semester: Fall 2021
Instructor: M. Farooq Khan

Objective:

The objective of this complex engineering activity is to carry out research, analysis, design, investigation, and implementation of a real-world complex programming project that has the following attributes:

1. The activity requires abstract thinking, originality in analysis to formulate suitable programming models of the activity;
2. The activity involves creative use of programming principles and research-based knowledge in novel ways;
3. The activity can extend beyond previous experiences by applying principles-based approaches.

Complex Engineering Activity Outcomes (CEAO):

The outcome of this activity is that the students may be able to partially attain some or all the following graduate attributes:

1. GA1: Engineering Knowledge
2. GA2: Problem Analysis
3. GA3: Design/Development of Solutions
4. GA4: Investigation
5. GA5: Modern Tool Usage
6. GA6: The Engineer and Society
7. GA7: Environment and Sustainability
8. GA8: Professional Ethics
9. GA9: Individual and Teamwork
10. GA10: Communication
11. GA11: Project Management
12. GA12: Lifelong Learning

This activity is mapped on CLO3 and your performance in this project will play an important role in their attainment.

CLO3	Students should be able to work effectively as an individual or member/leader of a technical team to **improve efficiency** & **pace** of their work.	P2

Task Instructions

Problem 1:

We are approached by a client from a food truck business owner to develop a model to predict profits. The CEO of the food truck dispatch fleet is planning to expand the business in different cities. The company already has trucks in various cities and also have the data for population and profits from those cities. You are required to analyze the data and propose a framework to predict profits, and facilitate in selecting the city for expansion.

1. The file 'ex1data1.txt' contains the dataset for our problem. The first column is the population of a city (which is the input X) and the second column is the profit of a food truck (which is the output Y) in that city. A negative value for profit indicates a loss. The population is in 10,000s and the profits are in millions of Rupees. Load the data in your framework.
2. Before starting on any task, it is often useful to understand the data by visualizing it. For this dataset, you can use a scatter plot to visualize the data, since it has only two properties to plot (profit Y and population X). Plot the data points in red.
3. To propose the linear framework to predict profits we can use the linear equation given below to solve for unknown coefficients for our model

$$Y = X\beta$$

4. For correct linear model we need to stack a column of ones in the X vector horizontally and solve the linear equation to obtain the unknown coefficients vector β.
5. The inverse of X is not possible so you can solve the equation for β using pseudo inverse.
6. Plot the predicted linear approximation line on top of the data points in scatter plot.
7. Perform predictions on 2 new inputs

Problem 2:

In another problem, we are approached by a CEO of a real estate company to design a housing price prediction engine. Suppose, the client wants to sell a house and he wants to know what a good market price would be. One way to do this is to first collect information on recent houses sold and make a model of housing prices. You are required to analyze the data and propose a 'Linear prediction' framework to predict housing prices.

8. The file 'ex1data2.txt' contains a training set of housing prices in Sector F-10, Islamabad. The first column is the size of the house (in square feet), the second column is the number of bedrooms, and the third column is the price of the house (in hundreds, 100s).
9. Before starting the task, perform mean normalization of features.
10. For correct linear approximation we need to stack a column of ones in the X vector horizontally and solve the linear equation for the unknown vector of coefficients β.
11. Solve the equation for β using pseudo inverse.
12. Perform predictions on 2 new inputs

Policy on Professional Ethics & Plagiarism

You are free to consult any book and online resources during the design and analysis phase, but you cannot copy them. Your design and implementation must be your original effort. Remember that if anyone is found to copy the code from the Internet or another individual, the individual shall face a severe penalty. You are not allowed to copy any material or code directly from the web or elsewhere. Note that if you are found to violate this policy or it becomes obvious that the work you have submitted is not your own or has been taken from some other source you shall be facing severe consequences.

The Formula that will be generously applied in all cases is:

Final Score = Raw Score – 2*CF*Total

Where CF is the Copying Factor. This means that even if you are found to copy only in 50% of the activity deliverables, and you end up scoring 100% raw score, but your final score will be 0.

Deliverables

1. The project code is to be submitted on GCR before **Saturday 25th December, 11:59 pm.**
2. The project report is to be submitted on GCR before **Sunday 26th December, 11:59 pm.**
3. Project viva on **Monday 27th December, between 8:40 to 4:30 in CSD Lab.**
4. Every student must have a hard copy of the project report during the viva.

Project Report Guidelines

The Report should contain the:

- Title (Complex Engineering Activity 2)
- Project name (Linear Regression)
- Introduction
- Implementation (include code snippets and explain how you solved each step)
- Learning outcomes
- GitHub link of your code in the appendix