DATA 527 – Predictive Modeling
Assignment 4
40 points

The problem assignment (70%)

The purpose of this exercise is to implement a feed forward neural network to predict the car market prices. The main goal is to work on data exploration and preparation besides the network model implementation. The dataset to use contains 9 columns for 5512 entries. The dataset for this assignment can be found here.

The lecture note provides the algorithm and the equations needed for the program implementation.

Implement the algorithm following the requirements described below and report your results in the writeup.

Requirements:

- Analyze the dataset. The purpose is to do a simple walk through of each column:
  - Define any columns that has no impact on the car price prediction and remove them
  - Define any missing values and clean them up using a chosen method
  - Check is there is any column that needs to be split into more features
  - The car price is in rupee, it might be worth it to convert it to dollars
  - Check for any duplicate data
  - Check for any data that needs to be converted into numerical values
- Split the data accordingly
- Define the neural network structure that fits the problem to solve
- Save the model parameters obtained from the training process in a log file *NNCarPriceParameters*. Save the following parameters in the file: The Learning rate, the number of iterations, the final error, the structure of the neural network (layers, nodes, links), and the final weights.
- Plot the cost error per iteration and save the plot in the current directory.
- Add a function that takes an input and make predictions using the implemented neural network.

The problem report (30%)

Possible outline for the report might be:

- Overview/ Description
- Methodology/Strategy
- Implementation
- Results
- Discussion
- Conclusion
- References

Instructions

You must complete this assignment individually; you are not allowed to collaborate with anyone else. You may discuss the homework to understand the problems and the mathematics behind the various learning algorithms, but you are not allowed to share problem solutions or your code with any other students. Using the discussion board on Blackboard for this topic is encouraged to post any questions you may have while working on the assignment and get back to any of your colleagues' questions that you might know the answer of.

Any sources of help that you use while completing this assignment (other students, textbooks, websites, etc.) must be cited in your PDF report.

This assignment consists of two parts: a problem report and a program exercise.

- For the problem report, you write up your solutions electronically and submit it as a single PDF document.
- Your solution to the programming exercise must be implemented in python. Any plots created in the program or results generated from the program should be reported in the PDF report.

You will be submitting the following files, including the log files:

- hw4-yourFullName.pdf (a PDF writeup)
- NNCarPricePrediction.py

Please follow the naming conventions exactly, and do not submit additional files including data sets. The files should be submitted as a zipped folder. Your name, class name, deadline, and term should be included in each file you submit (except the log files). In the program as a comment section on the top and in the report before your start the document.