CS 4320: Machine Learning

Assignment : Regression with a Neural Network

In this assignment, you will use linear regression to fit a model to a collection of data. Your goal is to minimize the MSE on a set of test data. The data is of similar complexity to the last assignment. You will need to use a data processing pipeline, and use a neural network model.

Use your personal data set available on Canvas in the `regression-3` folder.

Explore and analyze this data as you did in previous assignments. Include the plots and analysis in your report.

Design and use a data processing pipeline that will scale the data into a better range, and will add derived data features.

Fit two neural network models to the data. One model must have only a single layer. The other model must have at least two layers.

It is expected that you will use TensorFlow and Keras to build and fit the model. It is also expected that you will use a SciKitLearn pipeline to pre-process the data.

You will need to record the MSE found on the training data, and the MSE found on the testing data for both models.

Include a comparison between the two models.

Required Steps

- Download your data.
- Explore and analyze your data.
- Split the data 80%/20%, for training/testing.
- Write (or modify) a Python program using sklearn, tensor flow, and keras to process and fit the training data to neural network models.
- Report the MSE loss obtained for your best models on the training data.
- Report the MSE loss obtained for your best models on the testing data.
- Report the linear model coefficients found. (If appropriate.)
- Report *your* model function. (If appropriate.)
- Include your analysis of the quality of your models' fits to the data.
- Include your comparison of the two models.
- Commit and push your code in the git repository.
- Submit the report (as PDF) to Canvas.