Time Series Prediction with Multilayer Perceptron

Joseph Nguyen

November 15, 2018

I. Background

Multilayer perceptron is a fundamental class of neural networks. They are flexible because they can be used to solve both classification and regression problems depending on the activation function used. In class, we have already explored the effectiveness of using MLP to solve classification problems, however, we have not studied MLP performance in tackling regression problems. As a result, the goal of this study is to explore whether MLP is an effective tool for solving regression problems. More specifically, the problem of interest is a time-series prediction problem which is a subset of regression problems. As a bonus, the dataset used in testing the predictive ability contains important data. Therefore, if the MLP performs well, the results may shed some light on trends within the data.

II. Dataset

The dataset chosen for this study contains data about the air quality in Harris County – the third most populous county in the United States. The samples within the dataset are measurements of the concentration of carbon monoxide (CO) in the air and the date and time that the measurement was taken. The dataset is fairly large with approximately 118,000 samples.

III. Methods

The data will need to be extracted in a way that will be conducive to effective training for the MLP. As discussed in class, the time-series prediction problem can be formulated as a supervised learning problem where the features are the value of interest at various, equally spaced, points in time.

After extracting the data, a small amount of data should be reserved for use as the test set which will ultimately be used to evaluate the predictive ability of the trained MLP. The rest of the data should be split into N (value of N to be determined) partitions in order to train the MLP with N-fold cross validation where N-1 partitions will be used as the training data and 1 partition will be used as the validation set. The validation set will be used in MLP training for early stopping to prevent overfitting.

IV. Expectations

The performance of the neural network will be evaluated in terms of how accurately it can predict the air quality on a given day and if it can provide some insight on the trend of air quality over time. I believe that MLP will perform well in a time-series prediction problem because of its complexity, especially within the hidden layers. Parameters can be tweaked through experimentation to alter the complexity and achieve the best performance.

V. Timeline

November 19	Solidify Project Planning
November 22	Complete Core Implementation
November 29	Experimentation and Data Collection
December 6	Complete Report