Gage Halverson

Computational Project Proposal

EIND 558

11/1/2022

For the computational project I will be doing option two and using multivariate time series analysis. The dataset used will be Hourly Bike Demand dataset from Kaggle.com which includes other variables such as weather, user type, type of day (holiday, weekend, etc.). The dependent variable is the number of total rentals per hour. The author of the dataset has already split the training and testing data and provided an error computation notebook. I hope to be able to use a combination of the time series forecasting and other independent variables to make better predictions.

In the beginning stages of this project, I will be following the techniques described in the Deep Forecasting article, this will allow me to get PyTorch up and running and learn the basics of exploring a multivariate time series data set. I will then use the examples in the paper and other resources to do feature selection, which I hope to incorporate PCA for dimension reduction. Once feature engineering has been completed, I will be training the model using the training data and finding the error levels between the testing data and model predictions. These predictions will then be compared to the predictions of a model that is created using a technique from course work.

My personal goal for this computational project is to expand my knowledge of using predictive and machine learning tools in Python and explore PyTorch. I will be publishing this project on GitHub with appropriate licensing so that this project can be referenced in future EIND 558 course material.

Citations:

Method Paper: <https://towardsdatascience.com/transformer-unleashed-deep-forecasting-of-multivariate-time-series-in-python-9ca729dac019>

Data: <https://www.kaggle.com/datasets/brajeshmohapatra/bike-count-prediction-data-set?select=train.csv>