Neural Networks for
Time Series Forecasting

In this assignment you are asked to build a multilayer perceptron to forecast future natural gas prices. The original dataset is available via https://www.eia.gov/dnav/ng/hist/rngwhhdW.htm . The dataset is also included in the attached file "ngp.csv". The dataset shows the weekly gas prices between 29/9/2017 and 1/10/1997. Please note that the original dataset is arranged in reverse chronological order.

The goal of this exercise is to forecast the gas price for two weeks into the future based on the prices of the past 8 weeks. To facilitate your task, the original file has been processed to create a time series file "time_series_8_2.csv". In the time series file, the prices have been re-formatted so that each line includes the prices of 10 consecutive weeks. You should use the first 8 prices as input and the 9th and 10th prices as targets.

Requirements and Grading Criteria

1. Use scikit MLPRegressor to train over the dataset.   2%

2. Attempt different configurations for the regressor: 4.5%

   a) Attempt different numbers of hidden layers and different numbers of nodes in each hidden layer. Show 3 different configurations comparing their respective performance, and reasoning for the observed performance.

   b) Attempt 2 different learning rates, reasoning for the observed effect of varying the learning rate on the time before convergence.   1.5%

Having attempted different configurations for the regressor, use the optimal configurations you reached for the following experiments:

3. Observe and report on the progress of both training and validation errors until the optimal performance is reached.   1%

4. Take and report on adequate measures to guarantee that the regressor is not overfitted to the training data.   1%

**A 1% bonus will be awarded to
the highest performing regressor.**

Bonus Exercise: **(2%)**

Use the original file "ngp.csv" to build a time series of 4 inputs to forecast the price of one week into the future.

Is it possible to achieve a higher performance on this time series, with a lower number of hidden layers than the previous time series?

Justify the difference in number and size of hidden layers required to achieve high performance in timeseries_4_1 versus timeseries_8_2- if a difference was observed.