1 Objective

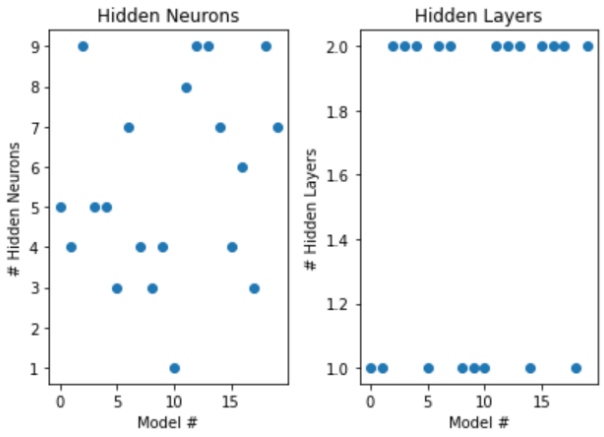
The given data file (data.mat or data.csv) contains two predictors x1 and x2 and target data y, each with 80 observations in time. Following the guidelines below, develop a MLP NN model of y as a function of x1 and x2. Briefly describe the approach taken. Also given are the forecast predictors: x1test and x2test (each with 40 observations in time), i.e. as a separate testing input dataset. Your task is to predict ytest using these test predictors and submit the results for ytest as a spreadsheet (e.g. csv file) together with your assignment. The TA will calculate the root mean squared error (RMSE) between your modelled ytest and true ytest which is not given to you. The smaller the RMSE is between your modelled ytest and true ytest, the higher will be your mark for this question (see point 5 below). Note: no need to standardize the data. Consider that this data is already given in the standardized form.

2 Guidelines

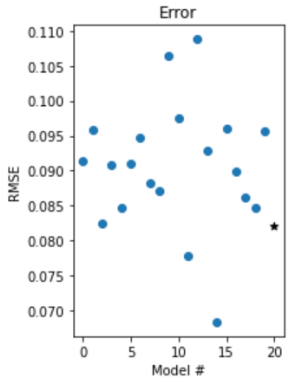
1. Set the model design, i.e. divide the data into training/validation and testing, select the model parameters etc. Provide some rationale for your choice. [2 points for the answers and rationale.]

   I chose to use 70% of the data for training, and the remaining 30% for validation. After several iterations, this seemed to give me fairly consistent results as compared with 90/10 or 60/40 splits

2. Provide several runs of the model (try $\geq$ 10 runs). For each run, use different NN model design, i.e. different number of hidden neurons per one or two hidden layers. Plot the results of the best model for each run, i.e. how many hidden layers and how many hidden neurons per layer [2 points for the plots].

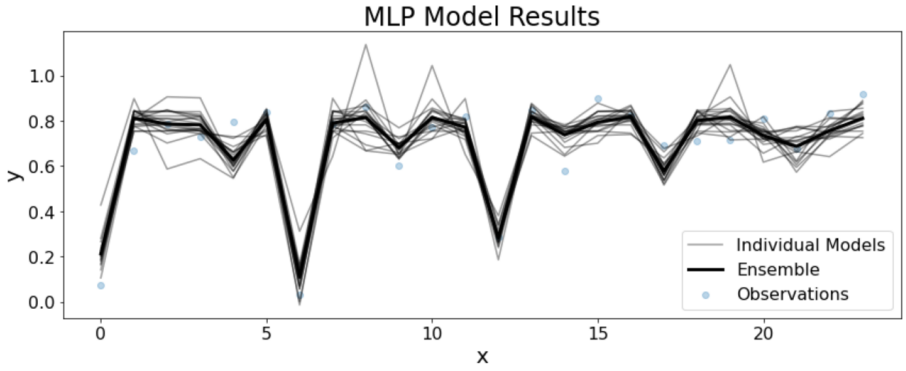


3. Plot the RMSE, calculated over your testing sample, for each model run as well as RMSE from the model ensemble mean [1 point for the plot(s) with RMSE of model runs and 1 point for the ensemble mean RMSE]

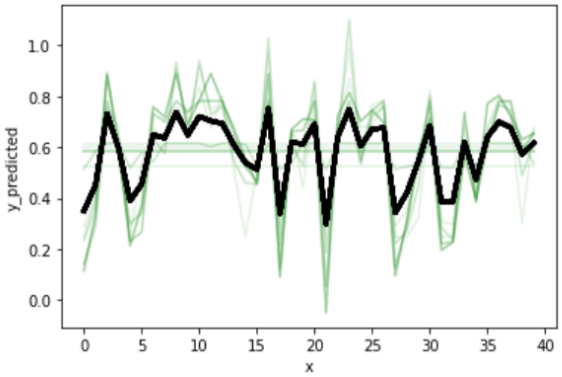


the $\star$ represents the ensemble mean RMSE

4. Plot the results of the ensemble runs and ensemble mean for your testing sample (ymodel vs time), as well as the target (true) y for your testing sample. [3 points for the plot].



5. Predict ytest as an ensemble mean of your treined model runs with x1test and x2test as an input and submit ytest values as a spreadsheet. [1 point for a reasonable estimate; additional 2 points according to how small your RMSE is. These points will be linearly scaled relative to the students' RMSE between their modelled and true ytest. E.g. student with the smallest RSME will get 2 points, and student with the largest RMSE will get 1 point, while the others will get their points scaled accordingly between min=1 and max=2].



see attached spreadsheet for x1, x2 and y values