# Quiz # 7. AMS 580

**Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_SBU ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Dear all, this is a take-home quiz. It is open book however you must do the work entirely on your own. It is due by 8:30am on Monday, March 27.**

**Please submit your Quiz in two documents: 1. The Rmd file, and 2. The output file (word or pdf) – to the SBU Brightspace (BS). If you have difficulty submitting to BS, you may then email the solutions (with the subject of “Quiz 7, AMS 580”, and in one email with the two attachments), to your TA Ian at:** [**weihao.wang@stonybrook.edu**](mailto:weihao.wang@stonybrook.edu)

#### Neural Network with the QuestionMark Data – Regression Task

The **QuestionMark.csv** data contain 14 predictors and one continuous response variable **y**.

Our goal is to establish a regression model predicting the response variable using the predictors – and in particular, using various modern and classical techniques to perform variable selection.

1. Please perform data cleaning by checking whether there are any missing values and if so, please delete observations with missing values. Please report how many observations with missing values we have in our dataset. Please use the random seed 123 to divide the data into 95% training and 5% testing. Please normalize the data using the R scale() function.
2. Please first build the predictive model to predict **y** using the training data and the NN model with (i) no hidden layer, (ii) the default loss function of “sse”, and (iii) the default activation function of “identity” (namely, the linear activation function). Please plot the model obtained using the training data. Please compute the Test MSE using the testing data.
3. Please first build the predictive model to predict **y** using the training data and the NN model with (i) one hidden layer with 3 neurons, (ii) the default loss function of “sse”, and (iii) the default activation function of “identity”. Please plot the model obtained using the training data. Please compute the Test MSE using the testing data.
4. Now we shall build the predictive model to predict **y** using the training data and the Multiple Linear Regression model. Please compute the Test MSE using the testing data. Please report the fitted linear regression model obtained using the training data – and compare to the NN model in Question 2. Please plot the predicted **y** using the method in Question 2 (NN with no hidden layer), Question 3 (NN with one hidden layer, linear activation function), and Question 4 (MLR: multiple linear regression) (all three on the y-axis), against the true values of **y**, using the testing data. Please also plot the predicted **y** using the neural network method in Question 2 (NN with no hidden layer) (on the y-axis) versus the predicted house price using the multiple linear regression method in Question 4 (MLR) (on the x-axis), using the testing data.
5. Please first build the predictive model to predict **y** using the training data and the NN model with (i) one hidden layer with 3 neurons, (ii) the default loss function of “sse”, and (iii) the output layer with the default activation function of 'identity', but the hidden layer with the activation function of 'relu'. Please compute the Test MSE using the testing data. Please plot the predicted **y** using the method in Question 5, and the predicted **y** using the multiple linear regression in Question 4 (both on the y-axis), against the true values of **y**, using the testing data.

A picture containing text, outdoor, water

Description automatically generated