UNIVERSITY OF THE FREE STATE
DEPARTMENT OF MATHEMATICAL STATISTICS AND ACTUARIAL SCIENCE
STSM 2634

Assignment 5

Full marks: 20

Date: 12 May, 2025

Deadline: 17 May, 2025

FOLLOW THESE INSTRUCTIONS METICULOUSLY, OTHERWISE MARKS WILL BE SUBTRACTED:

- **You must use R-Markdown to complete this assignment**. Assignments in any other format will NOT be accepted. **Please submit the Word file generated by R-Markdown.**

- Save the answer file as **'Assignment5'** as the file name and '**student number'** as the author name. Your **programming (code), and the output must be included in your answers**. Write the explanation after the code and the output as necessary.

- **If the code fails to run, you will get 0 mark.**

- You are allowed to use the class notes, or other helps from internet.

- All computations must be done with the help of suitable R functions. Manual or calculator-based answers will not be accepted.

Q1. Use the following packages: mlbench, neuralnet, NeuralNetTools, dplyr, caret. Use set.seed() function for the reproducibility of the output.

Consider the "BostonHousing" dataset from the mlbench library. This dataset has 506 instances and 14 attributes, and it is considered a relatively large dataset for neural network examples. The goal is to predict the median value of owner-occupied homes (medv) based on the other 13 variables.

The "chas" variable in the BostonHousing dataset is a factor and needs to be converted to numeric before scaling and training the neural network.

Follow the steps below:

1. Call the packages.

2. Load the dataset.

3. Convert the ‘chas’ variable into a numeric variable.

4. Split the dataset into two parts, training data and test data with 0.7 probability of an observation to be in the training data.

5. Scale the dataset by subtracting the minimum and dividing by the range of each column.

4. Create a neural network model using the neuralnet() function and save the output in a variable.

5. Print the summary of the output using the summary() function.

6. Predict the ‘medv’ using the predict() function.

7. Then create a neural network plot using the plotnet() function.

Then calculate the mean square error (MSE) for the above model. To calculate the Mean Squared Error (MSE) for the above neural network model, you can use the predict function to generate predictions and then compare these to the actual values.

Note: The **as.formula()** function in R is used to convert a character string into a formula, which can be used in various statistical and machine learning functions in R.
In the neuralnet() function, instead of writing all the variables in the formula, use the following code:

```
as.formula(paste("medv ~", paste(vars[!vars %in% "medv"], collapse = " + "))).
```

Q2. In the previous model, vary the number of hidden layers in the neuralnet() function and the number of neurons as (10,10); (5,5,5); (10,10,10). Then train the model again, predict the medv values as before, and calculate the MSE.

--- End ---