**Student: Seif Kungulio**

**Date: 05/04/2025**

**Subject: Project 8**

**Class: DSCI 512**

**Section: 01W**

**Instructor: Dr. Nengbing Tao**

**File Name: Project8\_Kungulio\_Seif.docx**

1. Load the dataset CreditCards.csv into memory.

A computer screen shot of a computer code

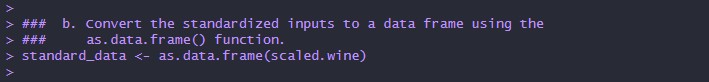
AI-generated content may be incorrect.

1. Preprocess the inputs
   1. Standardize the inputs using the scale() function.

A computer screen with white text

AI-generated content may be incorrect.

* 1. Convert the standardized inputs to a data frame using the as.data.frame() function.



* 1. Split the data into a training set containing 3/4 of the original data (test set containing the remaining 1/4 of the original data).

A computer screen shot of a program

AI-generated content may be incorrect.

The training dataset includes 1,199 rows (observations) and 12 columns (features), and it is used to help the model identify and learn patterns in the data. Meanwhile, the test dataset contains 400 rows (observations) and 12 columns (features), serving to assess the model’s ability to perform on new, unseen data, offering an objective evaluation of its predictive capability.

This division of the data ensures the model is properly trained while also being evaluated for its ability to generalize. The data is split with roughly 75% allocated to training and 25% to testing.

1. Build a neural networks model
   1. The response is quality and the inputs are: volatile.acidity, density, pH, and alcohol. Please use 1 hidden layer with 1 neuron.

A computer screen with text

AI-generated content may be incorrect.

* 1. Plot the neural networks.

A diagram of a network

AI-generated content may be incorrect.

The diagram illustrates a basic neural network featuring a single hidden layer designed to predict “quality” using four input variables: “volatile.acidity,” “density,” “pH,” and “alcohol.” Each input is linked to the hidden neuron through weighted connections, and these weighted inputs are combined and adjusted with a bias before being passed to the output neuron. The model achieved an error value of “258.92” after completing “12,282” training steps. This network appears to be set up for a regression task, aiming to estimate “quality” based on the given features.

* 1. Forecast the wine quality in the test dataset.

A screenshot of a computer

AI-generated content may be incorrect.

The predicted wine quality values range from 4.90 to 6.56. Notable statistics include a median of 5.47, a mean of 5.61, and the first and third quartiles at 5.21 and 6.01, respectively. Most predictions fall within this interquartile range, showing that the values are closely grouped. Since the median is slightly below the mean, there is a mild skew toward lower quality scores. Overall, the predictions suggest wine quality levels generally lie between 5 and 6, reflecting a narrow and reasonable range.

* 1. Get the observed wine quality of the test dataset.

A screenshot of a computer

AI-generated content may be incorrect.

The actual wine quality scores range from a minimum of 3.00 to a maximum of 8.00. Most values are concentrated around 6.00, with the first quartile at 5.00 and the third quartile also at 6.00, indicating that the bulk of the scores lie between these two points. Both the median and mean are near 6.00, reflecting an even distribution, though the mean is slightly lower, hinting at a slight skew toward lower scores. While a few wines are rated as low as 3.0, the majority are clustered around a quality rating of 6.0, with 8.0 being the highest observed score.

* 1. Compute test error (MSE).

A computer screen shot of a computer screen

AI-generated content may be incorrect.

The Mean Squared Error (MSE) of 0.4573465 reflects the average squared difference between the predicted and actual wine quality values. As a standard metric for assessing regression models, a lower MSE means the model's predictions are more accurate, while a higher value indicates less precision.

In this scenario, an MSE of 0.457 suggests that the model's predictions are generally close to the true wine quality scores. Although MSE is based on squared differences and doesn't represent error in the original scale of the data, the relatively low value points to strong predictive performance on this dataset.