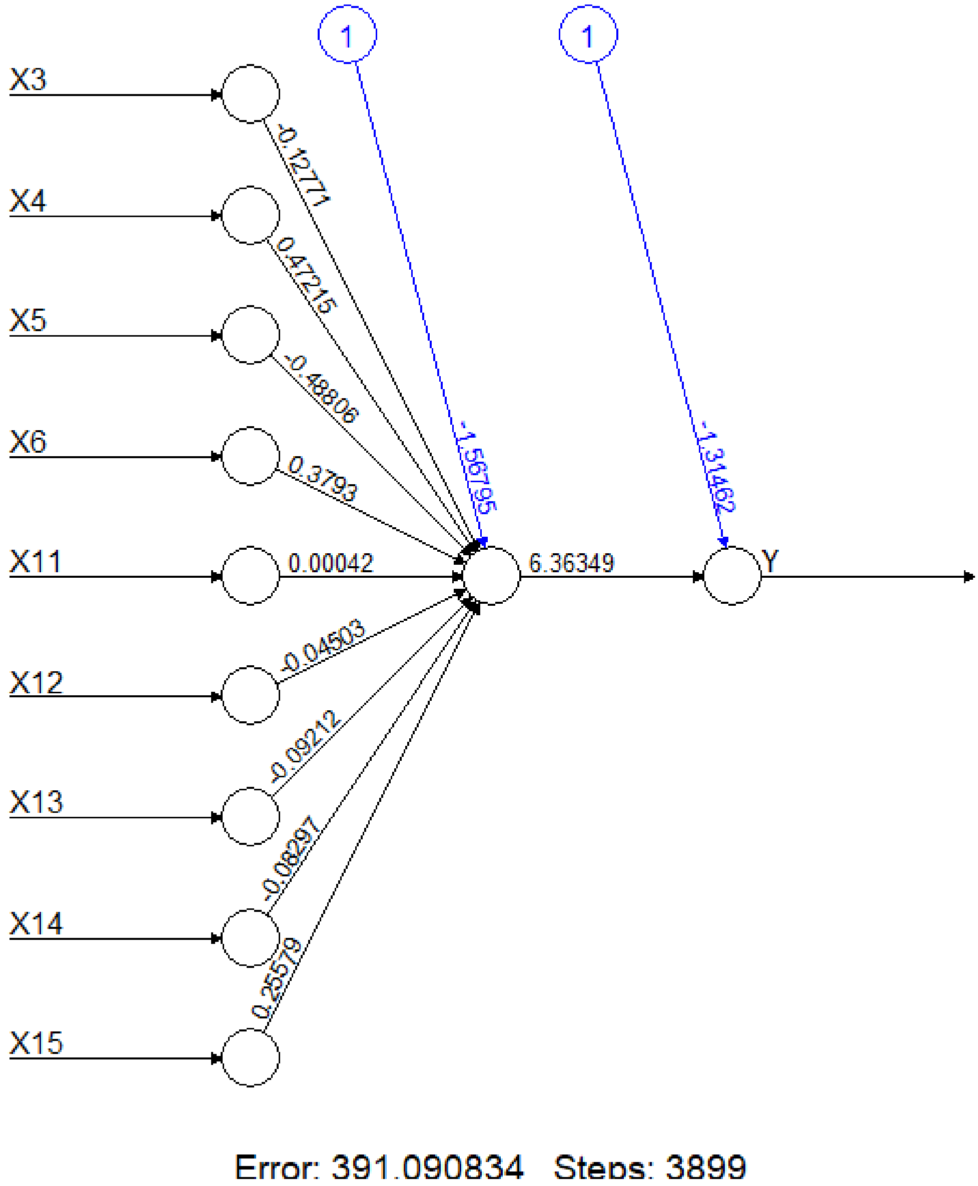
Question 1



The above neural network is the result of using all 9 predictors to find a single output "Y" with just **"1 hidden layer".** The blue lines above give the values of the bias terms and the other black lines represents the respective weights of the variables in getting the next layer or final output.

Assuming the layer before the final output is represented as **"X"**; therefore, the input at Y is given as:

Y = -1.31462 + 6.36349(X)

Where X is calculated from the weights of the predictor variables.

The Neural model is used to predict using the test data as input and the correlation coefficient between the actual and predicted values is given as **0.8675359** which implies that the is almost an 87% similarity in the values predicted by the model and the actual value. And a low mean Squared Error of 0.2578132 shows that the model prediction is close to the actual value as seen in the figure below.

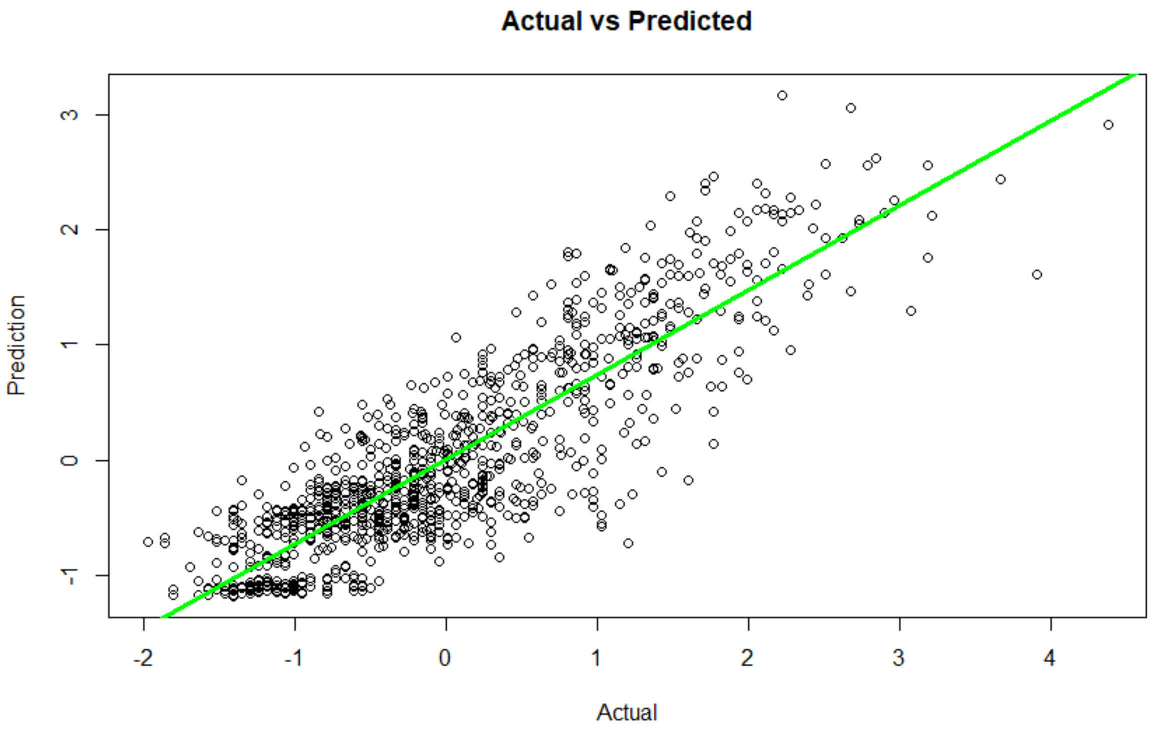


Figure 1: Actual vs Predictions with hidden layer 1

Question 2

	Hidden_unit	Mean_squared_error	Correlation_coeff
1	1	0.2578127	0.8675349
2	2	0.2129778	0.8919064
3	3	0.2001936	0.8988133
4	4	0.1829387	0.9079875
5	5	0.1684967	0.9156705
6	6	0.1479093	0.9262625
7	7	0.1459080	0.9272759

The table above shows the values of the mean squared error and correlation coefficient associated with respective number of hidden layer units. And it would be observed that the MSE (mean squared error) decreases with increasing hidden unit while the correlation coefficient increases with increasing unit. This implies that as the hidden unit increases, the model is able to account for more of the predictor variability which helps reduce error, therefore increasing the model accuracy in predicting with high correlation coefficient between the actual and the predicted values. The figures below show the plot of the MSE and correlation coefficient against the hidden unit and it would be seen that change momentum reduces after the 6th hidden unit; but the 7th unit gives the highest correlation coefficient of about 92.7%

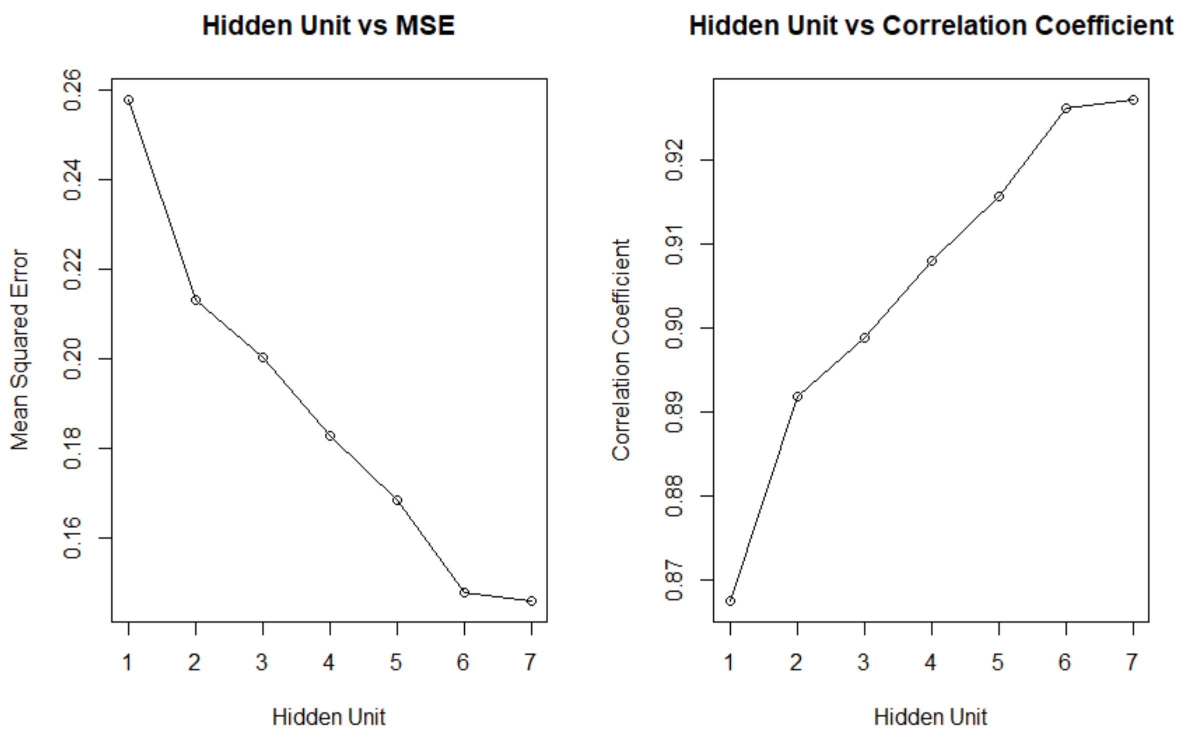


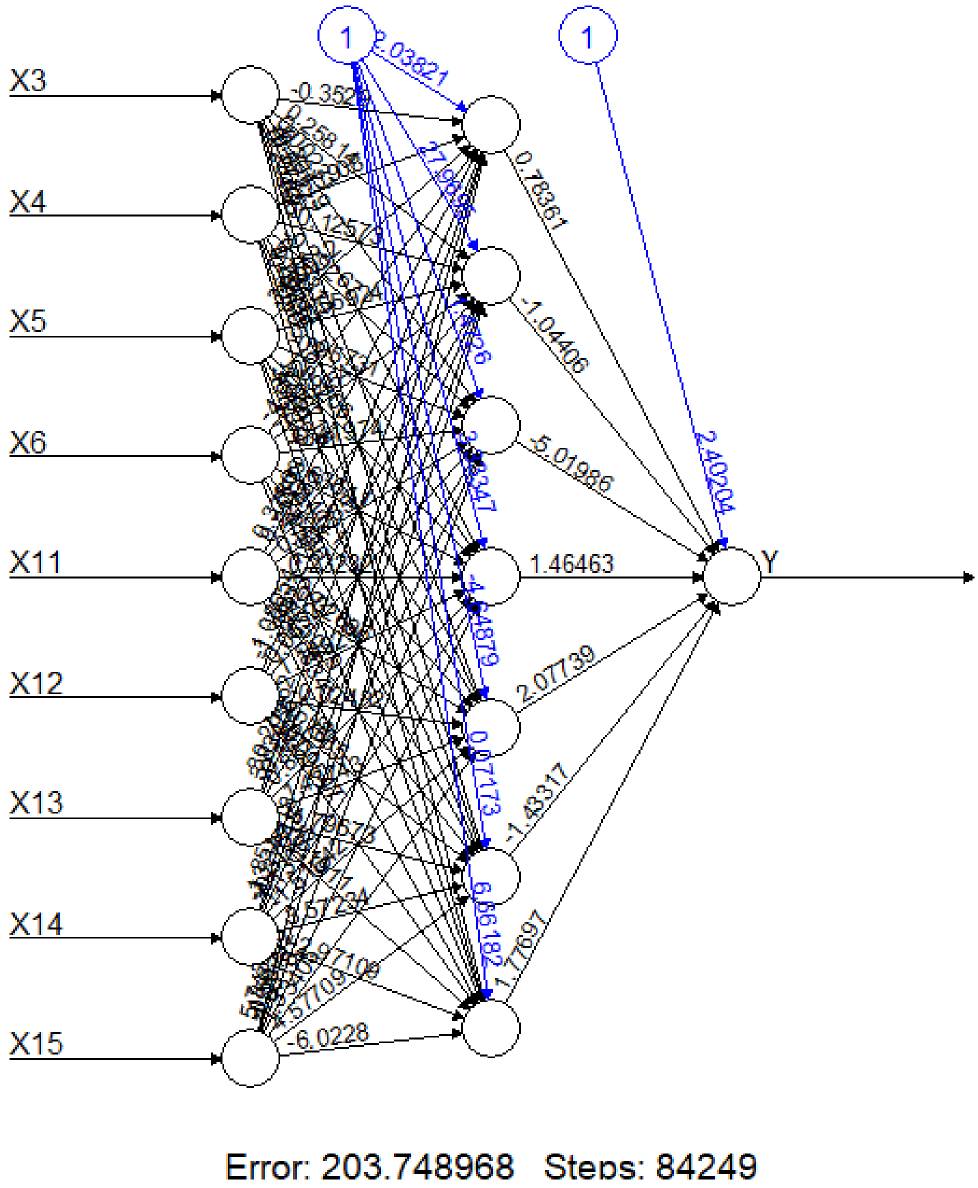
Figure 2: MSE and Correlation Coefficient

Exploring the model with 7 Hidden Units with the highest accuracy

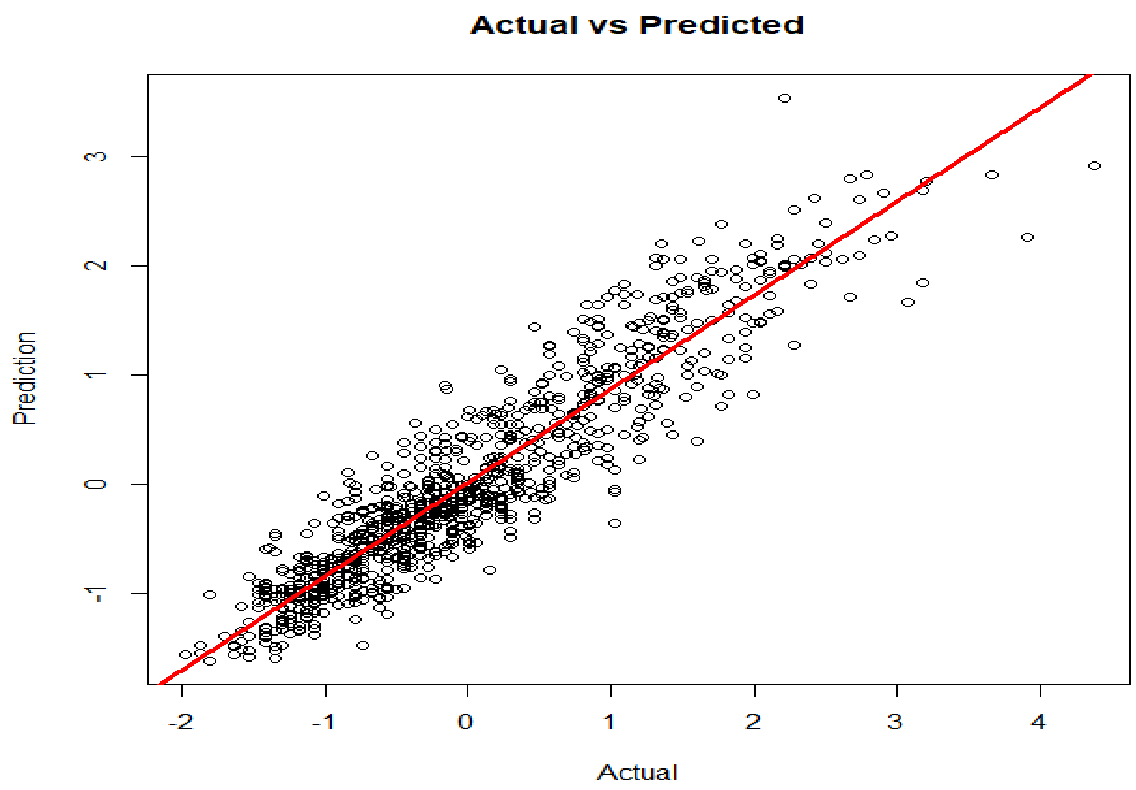
From the neural diagram below, assuming the hidden layer nodes are represented by **X1** to **X7,** Y will be given as:

Y = 2.4+0.78(**X1**) – 1.04(**X2**) **–** 5.02(**X3**) + 1.46(**X4**) + 2.08(**X5**) – 1.43(**X6**) + 1.78(**X7**)

And the correlation coefficient is 0.9285043, implying a 92.9% correlation between its predicted values and the actual data values. A lower mean square error of 0.143478 shows the risk function of the model. It will be concluded that the model with a higher number of hidden units like **(7)** will perform better than a model with lower hidden unit.



The below scatter plot gives a better visual of the actual value against the predicted values of a model with 7 hidden units, and it would be seen that the plots take a linear pattern and fall around the line (red).



Question 3

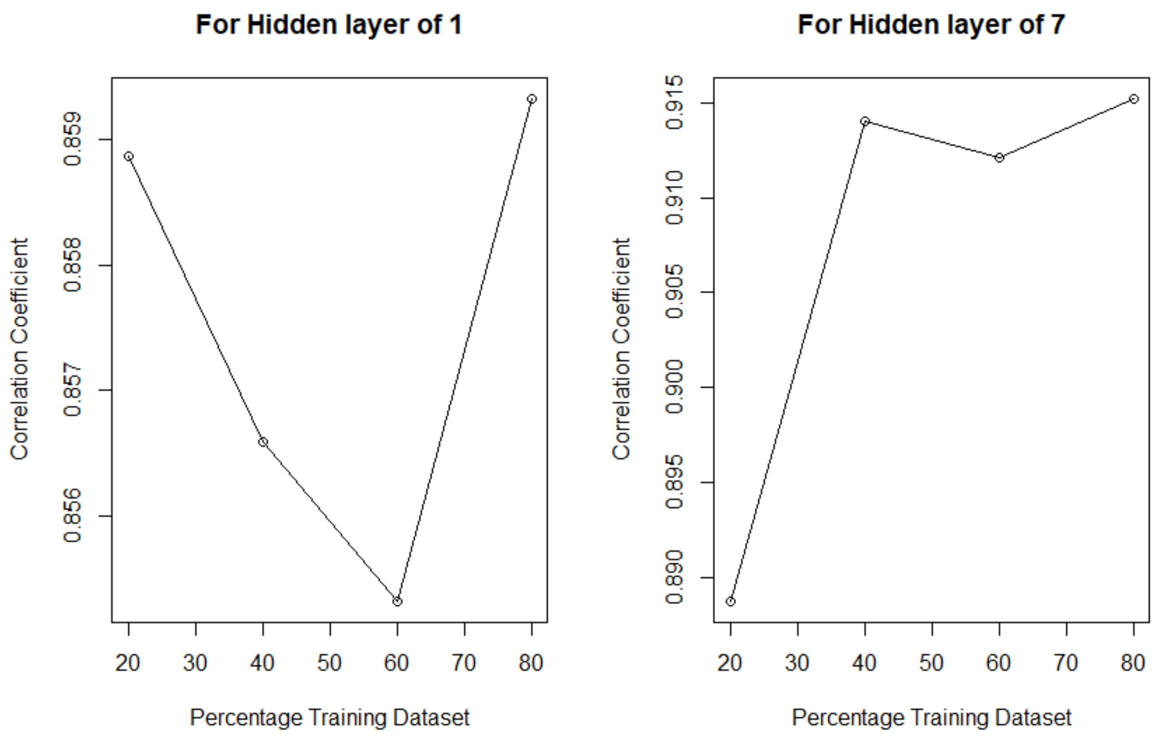


Figure 3: % Training vs Cor.Coef

The above plots show how the model reacts to a change in test/training data, the models used are the neural network model with 1 and 7 hidden units. From the above figures, it can be seen that the model with one hidden unit has a drop in correlation from 20% of the training data to 60% and keeps increasing from then. But the maximum correlation coefficient is around 86%. While for the model with 7 hidden units shows the correlation, coefficient increases drastically from 20 to 40% and a slow decrease at 60% of training data; then keeps increasing and peak at 80% of training data; but overall, it can be concluded that a training data of 60% gives a lower correlation coefficient in the 2 cases and a training dataset of 80% gives a higher correlation accuracy as there is sufficient data available to help the model handle high variance in the data and finally, the size of data used to train the model is an important factor to consider when building the model, provided the model will be used for significant predictions.