STAT430 – Assignment 5

Claire Grady

Student Number: 220256023

2a)

Initially, the Shapiro-Wilk test was performed on each predictor to assess whether the data was normally distributed and the results of these tests can be seen in Table 1. Given that all of the variables had p-values well below 0.05 we can infer that the data was not normally distributed. As a result of this, the min-max normalization method was used which transforms the data into a range between 0 and 1 while retaining the original distribution of the variables. Table 2 highlights the mean values of the variables prior to and after normalization. Before normalization, the mean values of the variables ranged from 1.393 up to 2007 and afterwards they ranged from 0.208 to 0.673.

Table 1: P-Values from the Shapiro-Wilk's Test

Predictor	P-Value
X	< 2.2e-16
X3	< 2.2e-16
X4	< 2.2e-16
X5	< 2.2e-16
X6	< 2.2e-16
X11	< 2.2e-16
X12	2.134e-08
X13	< 2.2e-16
X14	< 2.2e-16
X15	< 2.2e-16
Y	< 2.2e-16

Table 2: Numerical Summary of Minimum and Maximum Values for all Variables Before and After Normalization

Predictor	Training Mean Before Normalization	Test Mean Before Normalization	Training Mean After Normalization	Test Mean After Normalization
X	2007	1979.5	0.502	0.495
X3	18.11	17.95	0.208	0.222
X4	3.459	3.489	0.353	0.393
X5	354	357.1	0.390	0.406
X6	349.4	350.9	0.665	0.673
X11	3.041	2.992	0.510	0.497
X12	43.2	42.91	0.448	0.431
X13	1.393	1.391	0.248	0.247
X14	1.629	1.616	0.353	0.347
X15	9.683	9.652	0.441	0.437
Y	4.977	5.005	0.311	0.359

Artificial neural networks are versatile methods that are capable of prediction, classification and finding patterns in unsupervised learning. Initially, I chose to train nine feed-forward neural networks using different combinations of the Backpropagation, RPROP+ and RPROP- algorithms with the Sigmoid, Logisitic and Tanh activation functions and an overview of these networks can be seen in Table 3. The neural network that used the RPROP+ algorithm with the Sigmoid activation function produced one of the lowest squared sum of errors (9.714) and highest correlation coefficients (0.862) and achieved this with significantly less steps than the other neural networks that produced similar results. Therefore, this neural network was chosen and a graph depicting its topology can be seen in Figure 1. RPROP standards for resilient backpropagation and it is a variation of the backpropagation algorithm that also uses gradient descent to calculate the weights and biases. Unlike backpropagation which has a fixed learning rate, RPROP uses different learning rates for each weight and bias. The Sigmoid activation function is also known as a squashing function because it produces an output between zero and one. It takes a weighted sum of inputs and uses a non-linear function to produce the output.

The neural network displayed in Figure 1 has a network topology consisting of ten input nodes, one hidden layer and hidden node and one output node. The weights along the synapses show that the predictors that have the biggest impact on the output are X4, X5 and X6. Both X4 (-3.03) and X6 (-2.16) have negative weights which means they will reduce the output value whereas X5 (1.74) has a positive weight which will result in an increased output value. The weights for X (0.03) and X11 (0.03) are both very low, indicating that they will have significantly less impact on the output than the other predictors. The bias node in the hidden layer has a value of 3.32 whereas the bias node in the output layer has a value of 1.15.

Table 3: Overview of Nine Neural Networks

Algorithm	Learning Rate	Activation Function	SSE	Steps	Correlation Coefficient
Backpropagation	0.0001	Logistic	9.733	708609	0.863
Backpropagation	0.0001	Tanh	9.715	261479	0.863
Backpropagation	0.0001	Sigmoid	9.713	159303	0.862
RPROP+	Variable	Logistic	9.74	8735	0.863
RPROP+	Variable	Tanh	37.23	43	0.822
RPROP+	Variable	Sigmoid	9.714	15179	0.862
RPROP-	Variable	Logistic	9.739	7705	0.863
RPROP-	Variable	Tanh	11.98	992516	0.822
RPROP-	Variable	Sigmoid	9.724	8330	0.861

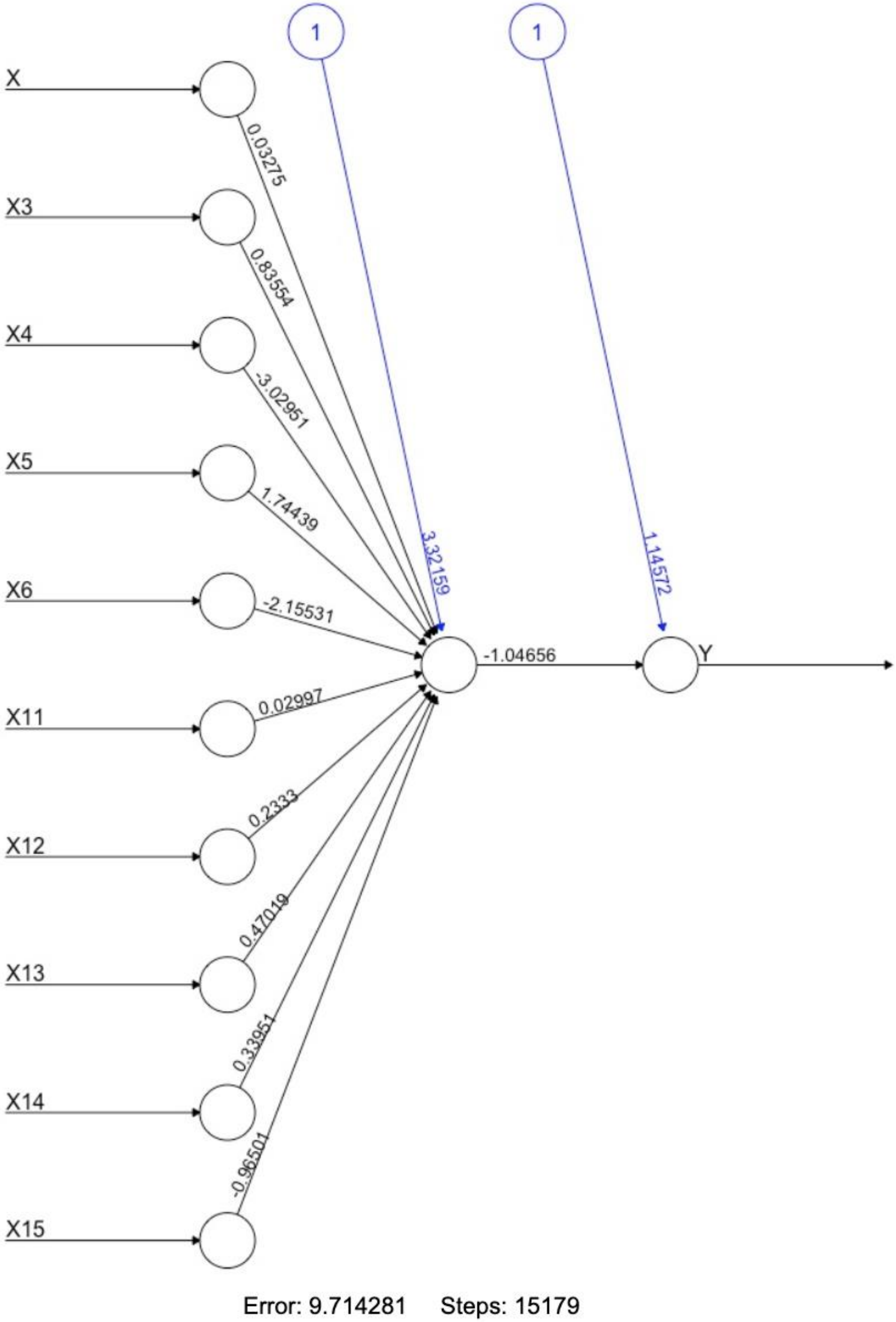

Figure 1: Trained Neural Network with Ten Input Nodes, One Hidden Layer and Hidden Node and One Output Node

2b)

Given the efficiency and relatively high accuracy of the neural network that used the RPROP+ algorithm with the Sigmoid activation function discussed earlier, the same approach was used here. Figure 2 shows the correlation coefficients of seven different neural networks, each trained with an additional hidden node. The network that produced the strongest correlation was the one with eight hidden nodes. The correlation coefficient of 0.931 indicates a strong correlation between the predicted and actual response variable (Y). Figure 3 shows the network trained with eight hidden nodes. The SSE (5.12) is significantly lower than the SSE in the network trained with only one hidden node (9.17) which indicates that with the given data set we are able to reduce the SSE by introducing eight hidden nodes as opposed to one. The number of steps required to train the network with eight hidden nodes was roughly seven and a half times as many as the network with only one hidden node which is unsurprising given how much more complex the network is. Table 4 provides an overview of all the networks trained. Apart from the networks with five and seven hidden nodes, the SSE decreased as nodes were added and the correlation coefficient increased.

Figure 4 shows the actual vs predicted response variable (Y) for the neural network trained with eight hidden nodes. It appears that the observations that contained a Y value of 5 or less were more accurately predicted than the observations with Y values greater than 5. It is possible that this is because there were a larger number of observations with Y values less than 5 than those with ones greater than 5 which can be seen in Table 5.

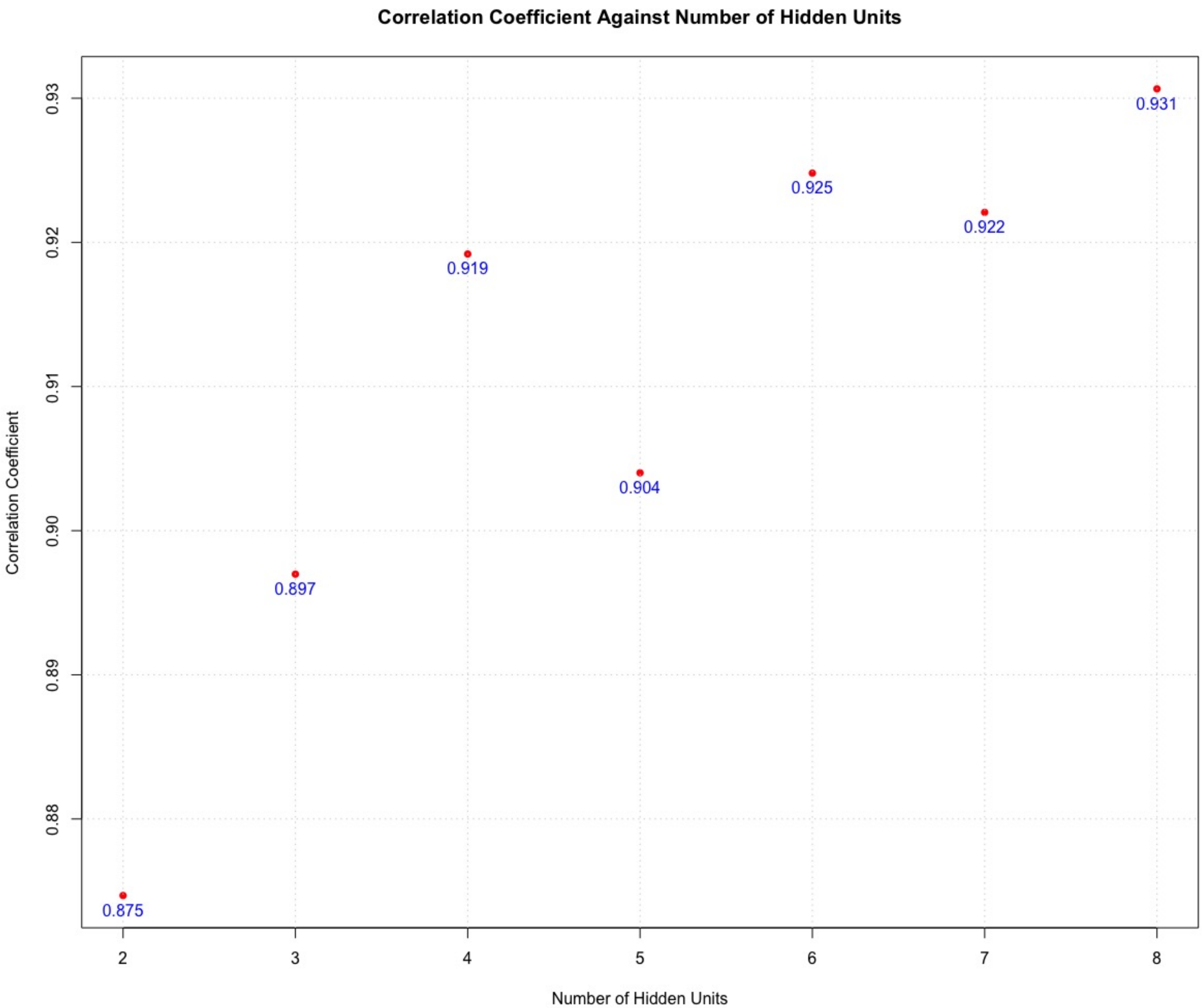

Figure 2: Plot Displaying the Correlation Coefficients Against the Number of Hidden Nodes.

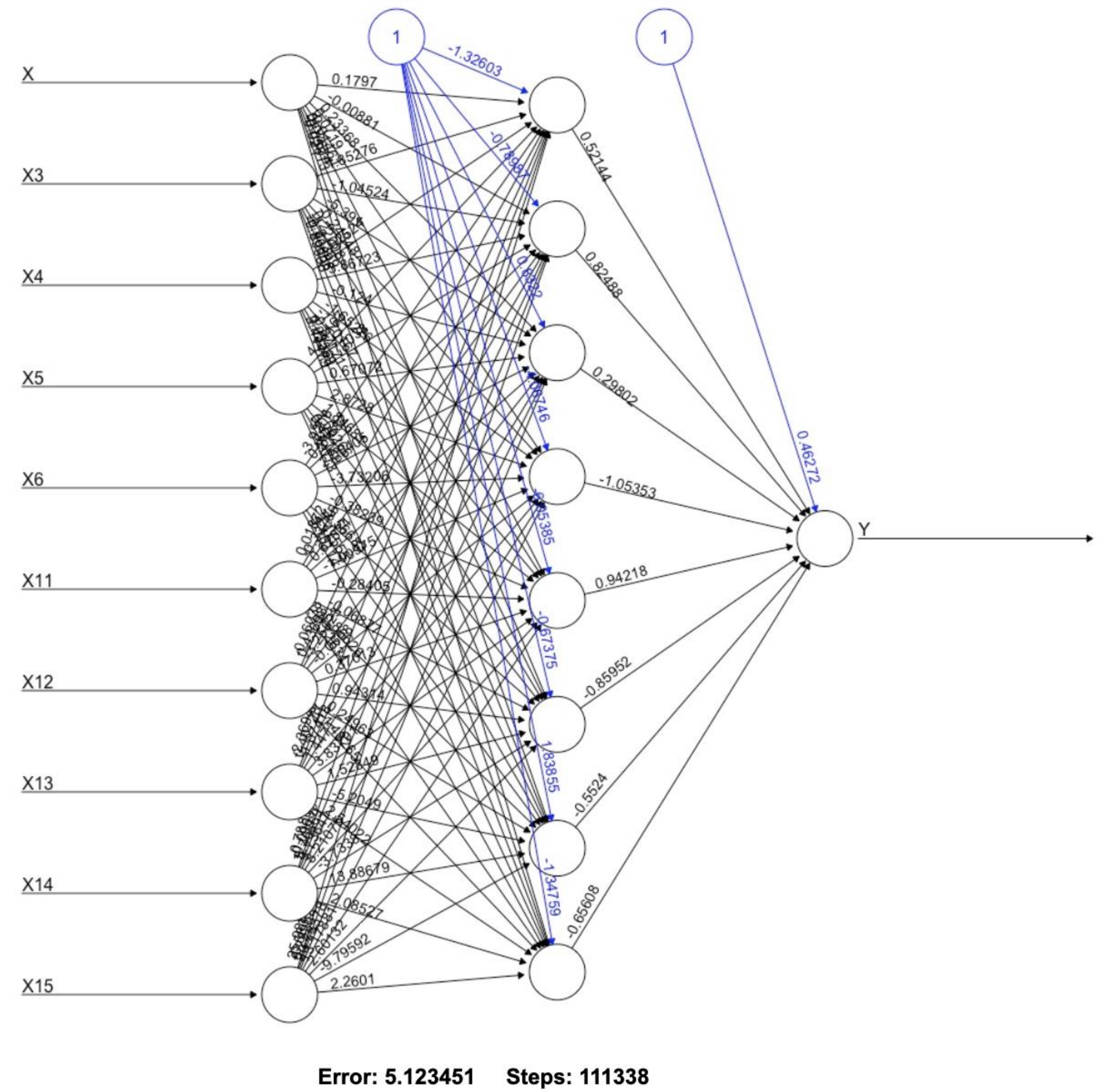

Figure 3: Neural Network Trained with RPROP+ with Sigmoid Activation Function and Eight Hidden Nodes

Table 4: Overview of Neural Networks Trained with RPROP+ and Sigmoid Activation Function

Algorithm	Activation Function	No. Of Hidden Nodes	SSE	Steps	Correlation Coefficient
RPROP+	Sigmoid	2	8.742	42506	0.875
RPROP+	Sigmoid	3	7.505	18703	0.897
RPROP+	Sigmoid	4	6.151	142594	0.919
RPROP+	Sigmoid	5	6.752	23202	0.904
RPROP+	Sigmoid	6	5.24	186833	0.925
RPROP+	Sigmoid	7	5.391	259491	0.922
RPROP+	Sigmoid	8	5.123	111338	0.931

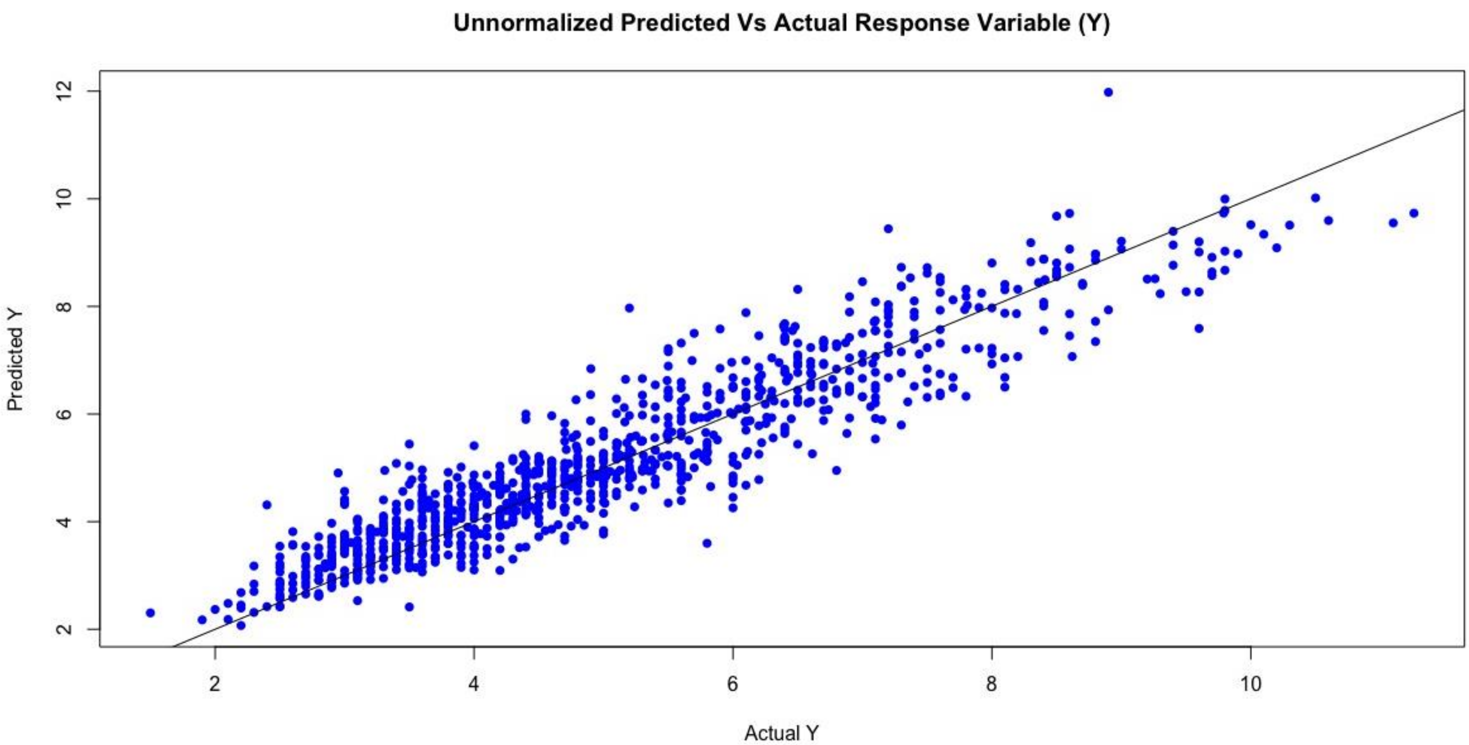

Figure 4: Actual Vs Predicted Response Variable (Y) with Unnormalized Data

Table 5: Number of Observations in the Training Data that Contained Y Values Less than and Greater than 5

Y Values Less than 5 in the Unnormalized Training Data	Y Values Greater than 5 in the Unnormalized Training Data
1708	1241

2c)

As the network that produced the highest correlation coefficient in the previous question was the one with eight hidden nodes, I opted to use the same number of hidden nodes for this question.

Figure 5 shows the correlation coefficients of four different neural networks created with varying amounts of data used for training. Interestingly, the correlation coefficients are either the same or very similar for the networks trained with 20%, 40% and 60% of the data. The correlation coefficient for the network trained on 80% of the data was the highest but was only 0.01 higher than the networks trained on 20% and 40%. In addition, the MSE which can be seen in Table 6 is roughly the same for all four networks. This suggests that there was not a significant difference in the network's ability to predict the response variable (Y) when the percentage of data used for training was increased.

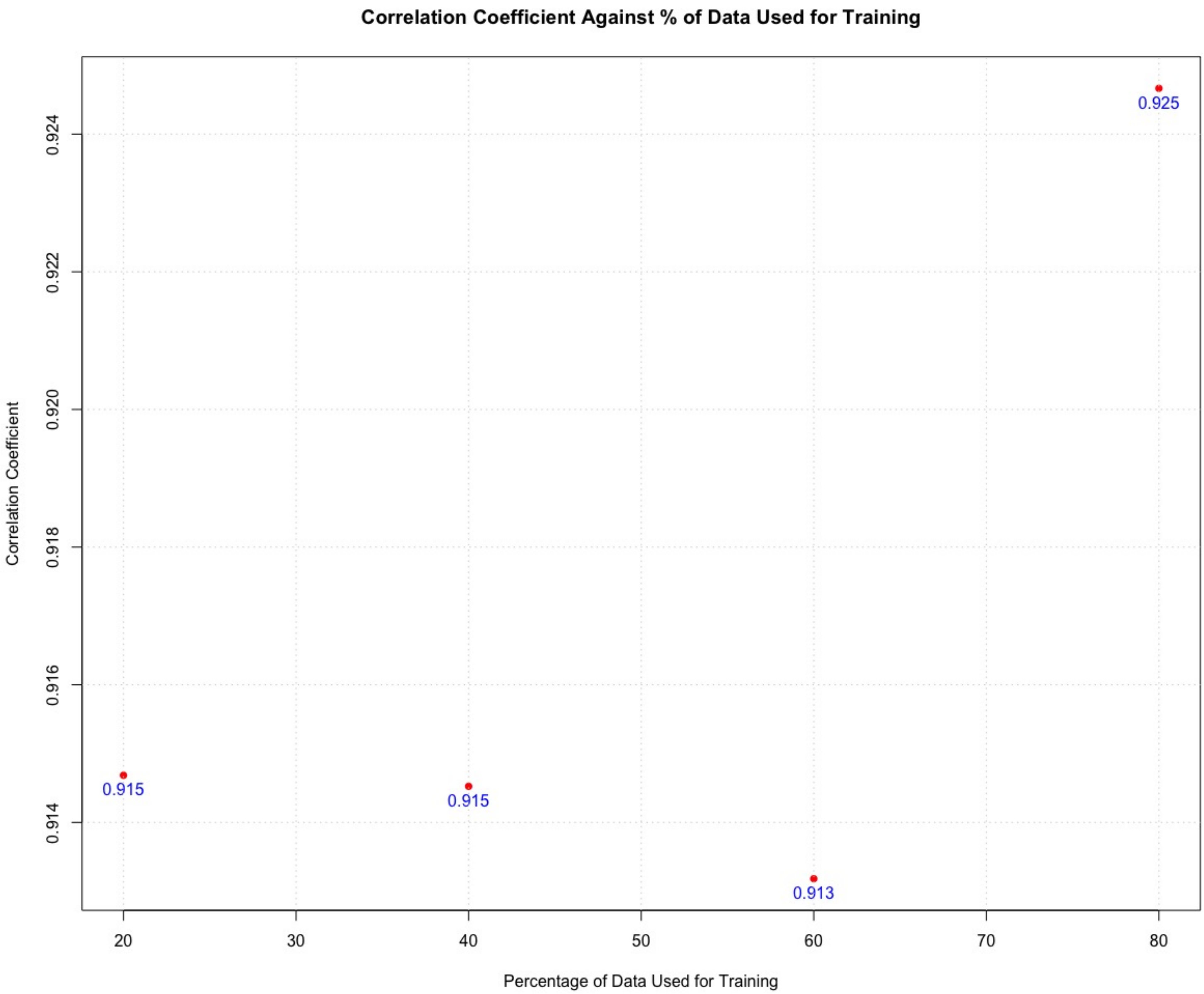

Figure 5: Correlation Coefficients of Neural Networks Produced with Varying Amounts of Data Used for Training

Table 6: Overview of Neural Networks Trained with Varying Amounts of Training Data

Algorithm	Activation Function	% of Data Used for Training	MSE	Steps	Correlation Coefficient
RPROP+	Sigmoid	20	0.0016	14695	0.915
RPROP+	Sigmoid	40	0.0015	198235	0.915
RPROP+	Sigmoid	60	0.0017	103073	0.913
RPROP+	Sigmoid	80	0.0017	208231	0.925