SER 594: Assignment 1 - Fundamentals of Neural Network

Technical Report

Neural network is one of the latest technologies in the field of computer science and Artificial Intelligence. Neural network is built using neurons, which in turn is a fundamental part of the human brain. The neural network has many commercial applications like pattern recognition, exploratory data analysis and so on.

The neural network is composed of many neurons as mentioned earlier. The neurons are arranged in layers. The first input layer receives the input and the successive layers receives the processed input. This is done in a systematic way as our human brain does. The network is made to learn from the input and give the prediction of the output.

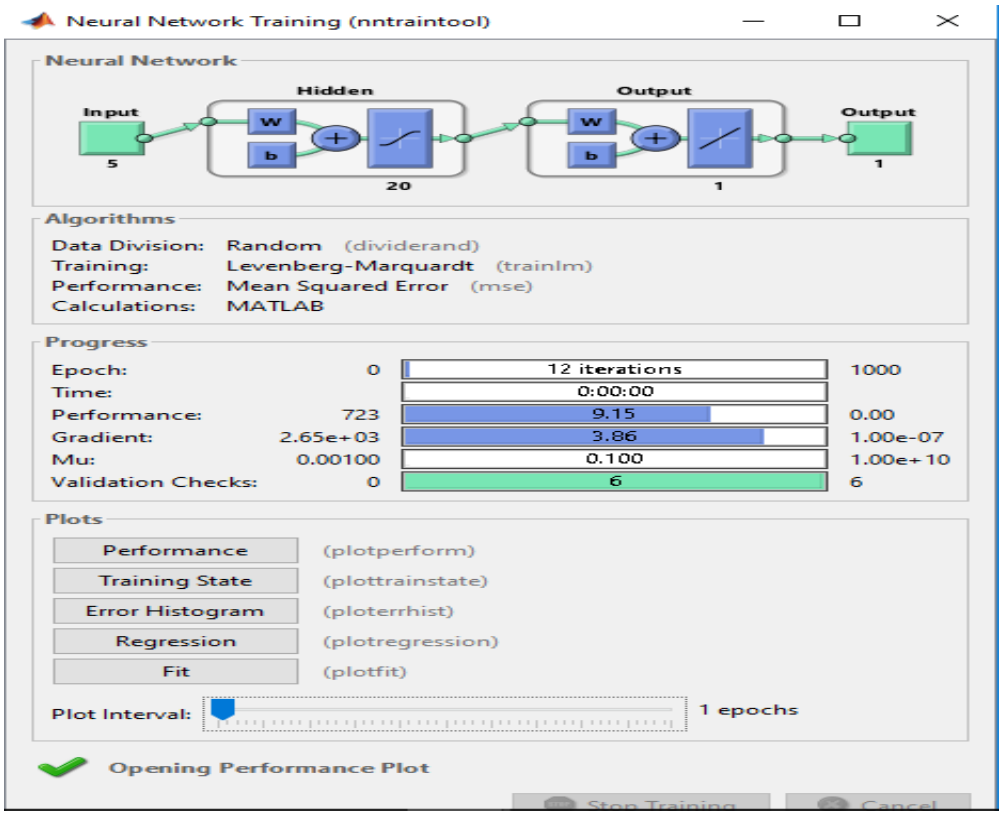
In the given problem, we need to estimate or predict the MPG (Miles per gallon) based on the provided information about cars. Given below is a screenshot of the provided information of cars.

mpg	cylinders	displacement	horsepower	weight	orign	car name
18.0	8	307	130	3504	1	chevrolet chevelle malibu
15.0	8	350	165	3693	1	buick skylark 320
18.0	8	318	150	3436	1	plymouth satellite
16.0	8	304	150	3433	1	amc rebel sst
17.0	8	302	140	3449	1	ford torino
15.0	8	429	198	4341	1	ford galaxie 500
14.0	8	454	220	4354	1	chevrolet impala
14.0	8	440	215	4312	1	plymouth fury iii
14.0	8	455	225	4425	1	pontiac catalina
15.0	8	390	190	3850	1	amc ambassador dpl
15.0	8	383	170	3563	1	dodge challenger se
14.0	8	340	160	3609	1	plymouth 'cuda 340
15.0	8	400	150	3761	1	chevrolet monte carlo
14.0	8	455	225	3086	1	buick estate wagon (sw)

The first column gives the information on the MPG, in the second column, number of cylinders for each car is given, the third fourth and the fifth column gives the information about the displacement, horsepower and the weight of the car respectively. We know that the predicted MPG should be the output of the neural network that we build. Hence, the input data will be all other columns except the MPG column.

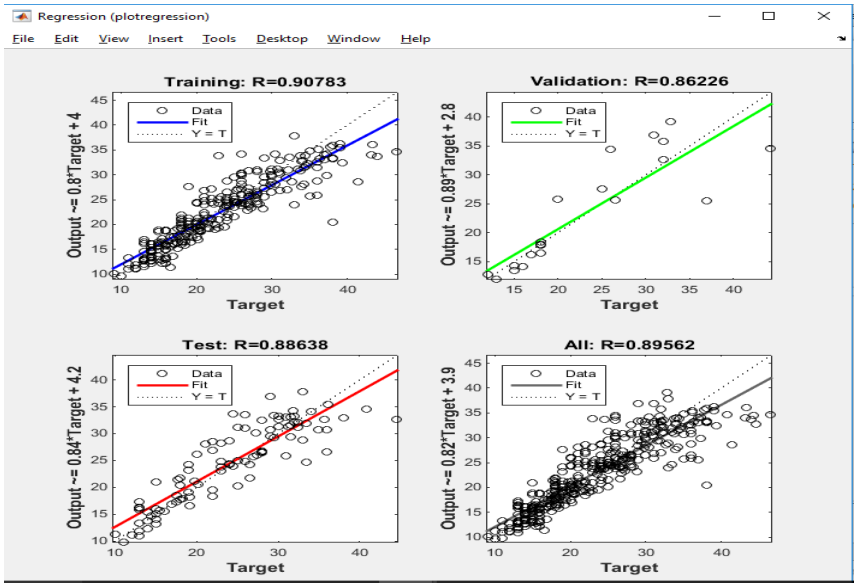
Now, let us divide the given data into two parts. Training data (70%) and Test data (30%). There are 398 rows in the given data set. The first 279 rows are used as training data and the remaining 119 rows are used as the test data. The MPG of the first 279 rows is used as the 'Training output' in MATLAB and the first 297 rows for the columns, cylinders, displacement, horsepower and weight are used as 'Training Input'. Likewise, the rows for the columns, cylinders, displacement, horsepower and weight of the remaining 119 rows are used as 'Test Input' and the MPG column of the remaining 119 rows are used as 'Test Output'. The next step is to transpose all these different sets of data in order for the neural network to accept them. It is as similar as transposing a matrix.

We now start building our neural network in MATLAB to predict the MPG of the cars. To build the neural network we need to train the network using the training data that we have. Therefore, we give the 'Training Input' as the input data and 'Training output' as the Target data for the neural network. As mentioned previously, there are different layers of neurons present in the neural network. For the given data we take 20 neurons in the hidden layer of the neural network. We use Mean Squared error as the performance function. Our neural network also uses 'Levenberg-Marquardt optimization' as the training function and 'gradient descent with momentum weight and bias' as the learning function to train the neural network. Now, with the 'Training input' as the Input and 'Training output' as the Output, we train the neural network.



The above figure shows the neurons, the hidden layer, input and output layers and the performance of the neural network.

The output of the neural network is the measure of how much the neural network has learnt from the given input data. We measure the output using the linear regression line. The R value in the linear regression line tells us if the neural network has efficiently predicted or not. The R value is always between 0 and 1, higher the R value, better the model. The below figure shows the R value of the Test data and the train data.



We see that the R value of the training data is 0.90783. That means that the neural network that we built has learnt 90.783% of the given data and only 9.127% of the entire data set could not be deduced by the network.

We also calculate the error percentage using the test data. Test data as I mentioned above is the untrained data. The error is calculated using the formula:

$$Error = \frac{\Sigma(actual\ test\ ouput) - \Sigma(neural\ network\ test\ output)}{\Sigma(actual\ test\ output)}$$

When the above value is calculated the average error value is 0.1159. That means that 11.59% of the data is erroneous and 88.41% of data is accurate.

To conclude, the neural network that we build has predicted the MPG and the predicted value is 90.783% accurate. This tells us that the neural network is highly efficient in predicting the MPG value of the cars. In this manner, neural network can be built for many other types of data sets of various fields and predict the values.