NEURAL NETWORK AND FUZZY CONTROL

EEE-1007

Predicting Diabetes with Back Propagation Neural Network using MATLAB

Team Members :-

Dhruv Gupta 20BEI0083

Aryan Shekhar 20BEI0085

Harsh Kumar 20BEI0088

Guided By :-

Dr. Himadri Lala

Related Literature -:

- Data mining is the extraction of useful information from the large volume of data .
- Data mining has been applied in various fields like medicine, marketing, banking, etc.
- In medicine, predictive data mining is used to diagnose the disease at the earlier stages itself and helps the physicians in treatment planning procedure.
- Asha Gowda Karegowda, provided the application of hybrid Genetic Algorithm and Back Propagation Network.
- They experimented for classification of PIMA dataset. They concluded that the backpropagation learns by making modifications in weight values by using gradient method starting at the output layer then, moving backward through the hidden layers of the neural network.
- So, it’s prone to lead to troubles such as local minimum problem, slow convergence pace and convergence unsteadiness in its training procedure.

- Ravi Sanakal, presented a diagnostic Fuzzy cluster FCM as well as SVM using SMO and decided which technique helps in diagnosis of Diabetes disease.

- The best result is obtained in a FCM with an accuracy of 94.3% and the positive predictive value is 88.57%. SVM has an accuracy of 59.5% which is quite low.

- These results are quite satisfactory only, due to the fact that detecting the Diabetes is a very complex problem.

- Rajesh, presented the C4.5 algorithm for classification and the success rate for classification obtained was of 91%. Future enhancement of this work includes improvisation of the C4.5 algorithms in order to improve the classification rate with greater accuracy.

- Radha, demonstrated the application of five classification techniques namely C4.5, SVM, KNN, PLR, and BLR to predict the diabetes disease in patients. They pointed out that necessary to intend an automatic classification tool.

- In this study, these five techniques were chosen based on the computing time, in which BLR has the lowest computing time with 75% accuracy and error rate of 0.27.

- The second one with more accuracy rate is SVM while comparing with other techniques. The accuracy of BLR is 75% from the results obtained. The BLR algorithm plays a vital role in data mining techniques.

- Raj Anand, presented a novel approach to Pima Indian diabetes data diagnosis using PCA (Principle Component Analysis) and HONN (Higher Order Neural Network).
- The HONN can perform diabetes classification with parsimonious representation of node architecture due to its generation of higher order terms. A lower mean square error and faster convergence is attained with PCA pre-processing.
- Veena Vijayan, suggested for the expectation of maximization of algorithms, among K Nearest Neighbour algorithm, K-means algorithm, Amalgam KNN algorithm and Adaptive Neuro Fuzzy Inference System algorithm.
- From observation EM (Expectation Maximization) possess the least classification accuracy. Amalgam KNN and ANFIS (Adaptive Neuro Fuzzy Inference system) provide the better classification accuracy results. Amalgam KNN comprises both the feature of KNN and K means. ANFIS incorporates both the features.
- Priya, proposed a method of applying neural networks for classification. The result produced by this model is higher than the other models, since it performs classification using neural networks in the rapid miner tool.

- The Objective of this project is to predict the chances of diabetes in a person, whether he/she is prone to it or not.

- Using neural network feed forward prediction model in conjunction with back propagation algorithm, and given training data set.

- We might be able to predict whether a subject is likely to have diabetes.

TIMELINE-:

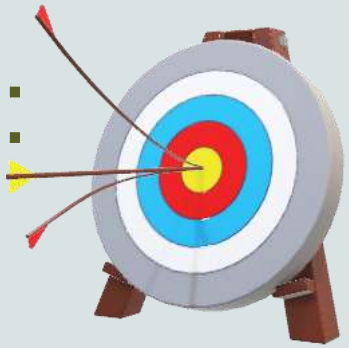
- April 2021 – Topic selection
- April 2021 – Topic approval by the faculty.

WORK STARTED.

- Learning the algorithm through various online sources.
- Learning related programming for neural networks on MATLAB.
- Designing of neural network and building it using MATLAB.
- Started training the network using sample data.

- May 2021- END of the Project

Target Deliverables -:



- This Project is to be used for diabetes diagnosis in people of all age groups.
- The target audience include: men, women and old age.
- People having Type 2 diabetes can be predicted, but it can also be used to diagnose Type 1 diabetes which is a genetic disorder
- To predict Type 2, the genetic factors can be taken into account in the learning method and hence the target audience easily covers all age groups.

SOURCE CODE-:

neural_project.m

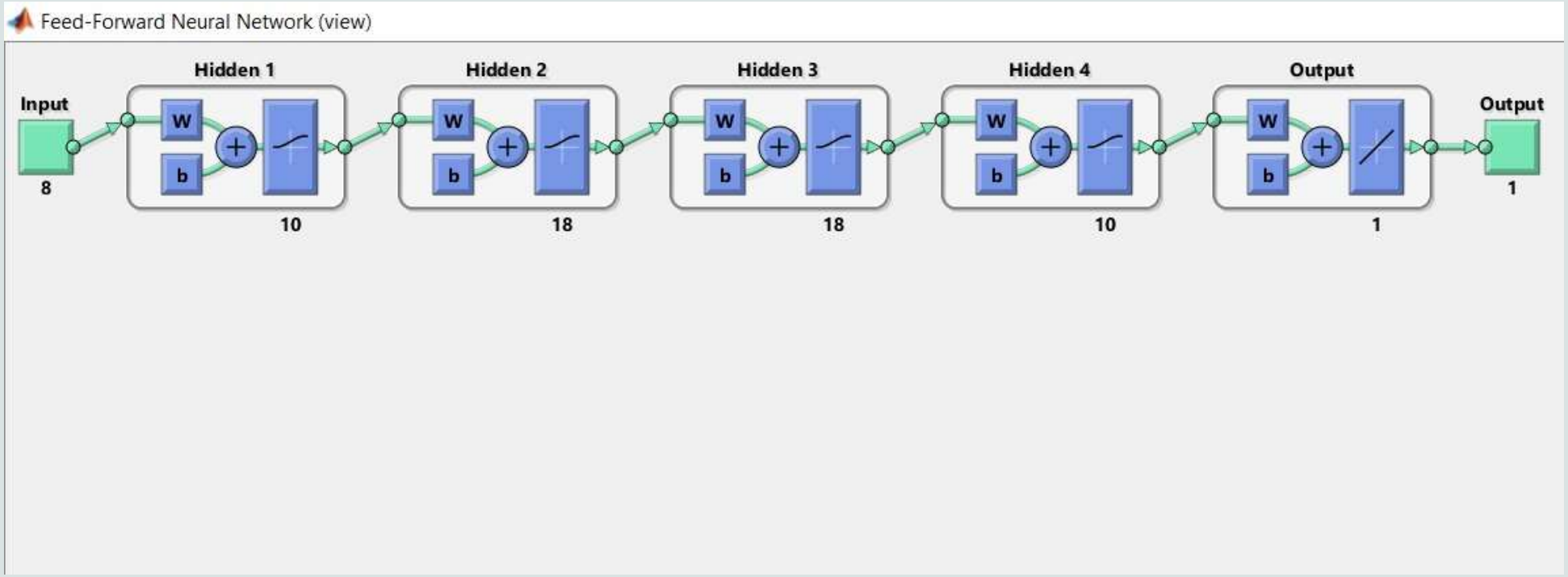
```
% This is just a test code for designing neural network with
% the help of back-propagation algorithm to predict diabetes
% in patients ..........................................
% The data collected is base on PIMA indian kaggle repository
% of machine learnig only some part of data is used here for
% testing the neural net.....................................

% Normalizing the input data in the range of [0,1].............
x_input_normalized=normalize(x_input_data,'range',[0,1]);

%Creating the network with deep learning toolbox..............
net=feedforwardnet([10,18,18,10],'traingd'); % Gradient Descent is used here

%Setting training parameters.................
net.trainparam.epochs=10000;       %No of iterations
net.trainparam.max_fail=10;         %NO of validation checks
net.trainparam.lr=0.02;            %Learning rate of network

%Setting activation function..................
for i=1:4
    net.layers{i}.transferFcn='logsig';
end
% Log-sigmoid is used as activation function for hidden layers....
net.layers{5}.transferFcn='purelin';   %At the output layer...........

%Training network.........................................
net=train(net,x_input_normalized,target_data);
y=net(x_input_normalized);       % Results from the network after training
```

neural_project.m

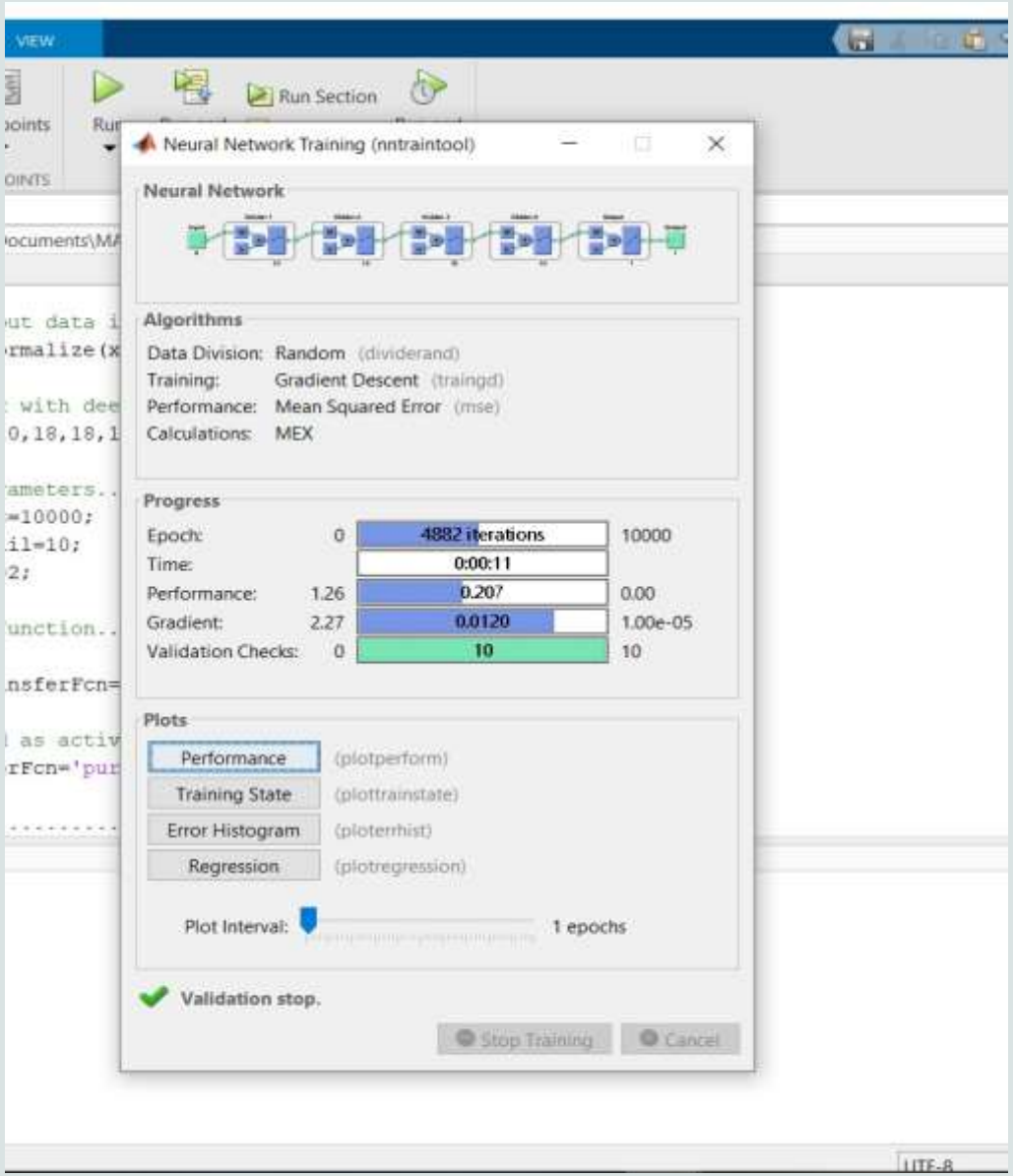
```
33      %Setting activation function.................
34 -    for i=1:3
35 -        net.layers{i}.transferFcn='tansig';
36 -    end
37      % Log-sigmoid is used as activation function for hidden layers....
38 -    net.layers{4}.transferFcn='purelin';   %At the output layer..........
39
40      %Training network.....................................
41 -    [net,tr,e]=train(net,x_input_no,target_data);
42 -    y=net(x_input_no);        % Results from the network after training
43
44      %Testing the prediction of network with seperate test data:.............
45 -    target_test=target_data(:,tr.testInd);  %Extracting test data used by
46 -    y_test=y(:,tr.testInd);                 %matlab in our model
47 -    [m,n]=size(target_test);
48 -    No_of_test_sample=n;      %Number of test samples in our model (15% of input_data)
49 -    count1=0;
50 -    count2=0;
```

```
51 -  for i=1:n
52 -      if target_test(i)==0 && y_test(i)<0.3
53 -          count1=count1+1;
54 -      end
55 -  end
56
57 -  for j=1:n
58 -      if target_test(j)==1 && y_test(j)>0.6
59 -          count2=count2+1;
60 -      end
61 -  end
62    %accuracy in terms of score.....
63 -  score=(count1+count2)*100/No_of_test_sample;
64 -  fprintf('prediction score is ....%f percent\n',score);
65
66
67
```

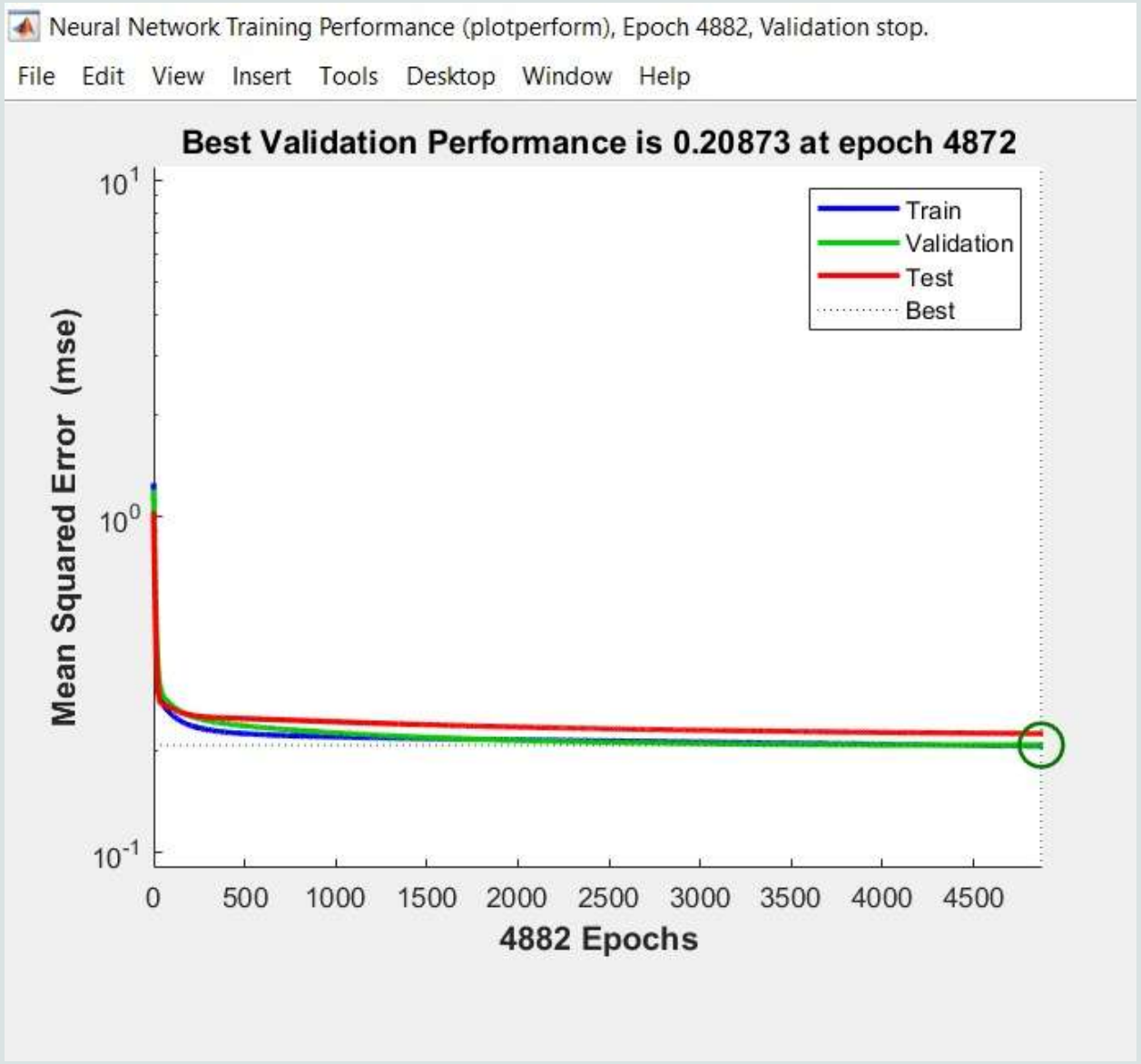
ARCHITECTURE-:

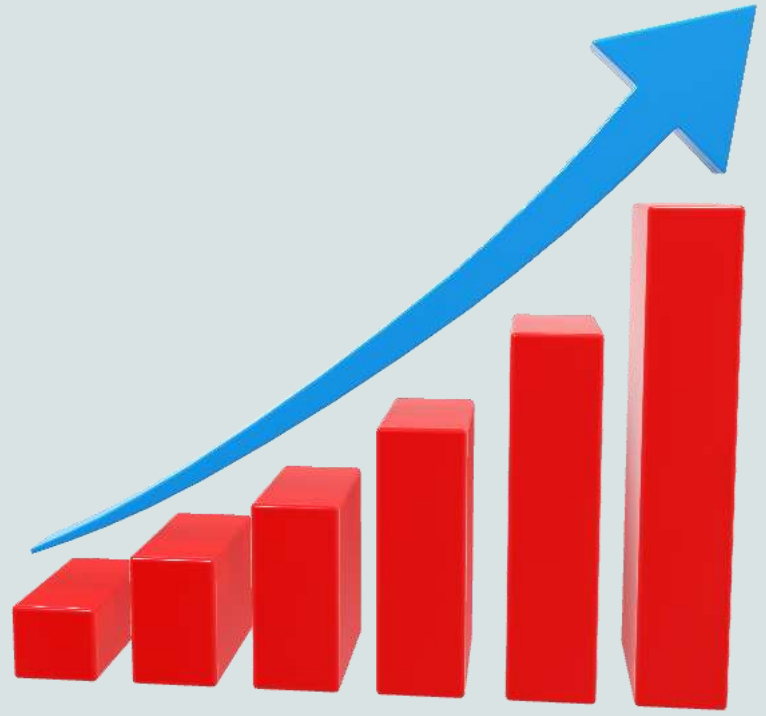


TRAINING-:

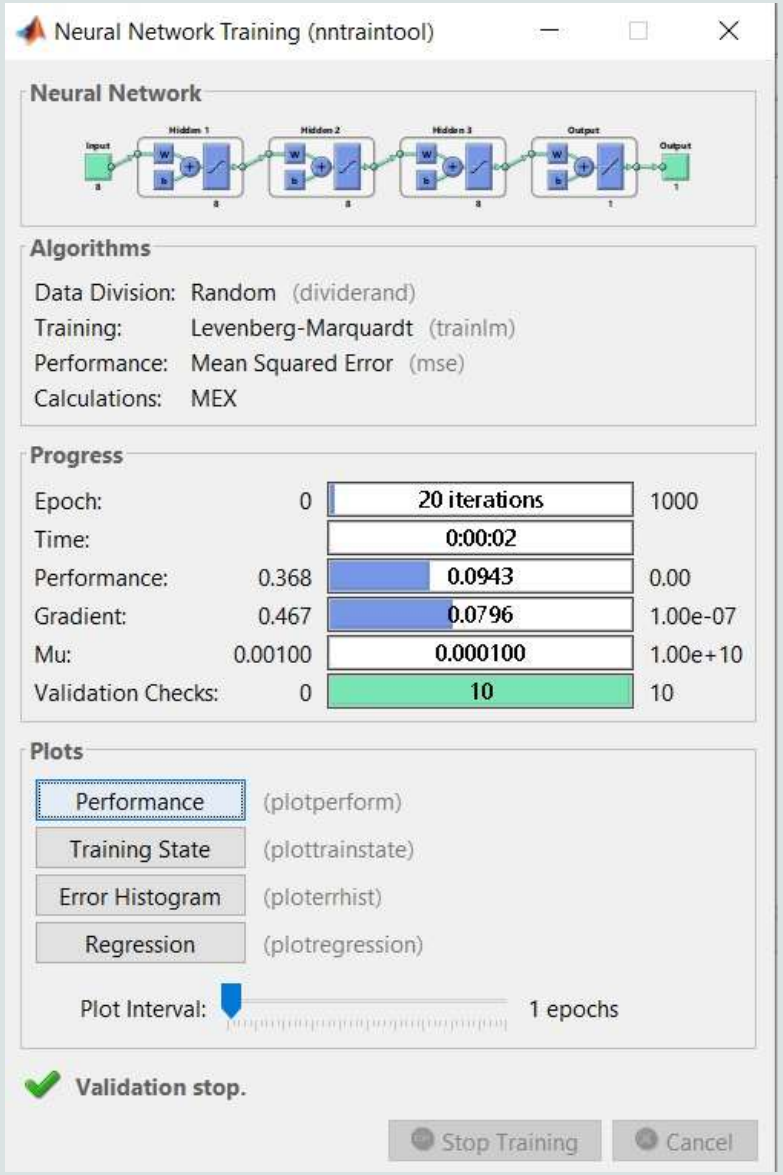

Run Section
Neural Network Training (nntraintool)
Neural Network
Algorithms
Data Division: Random (dividerand)
Training: Gradient Descent (traingd)
Performance: Mean Squared Error (mse)
Calculations: MEX
Progress
Epoch: 0 4882 iterations 10000
Time: 0:00:11
Performance: 1.26 0.207 0.00
Gradient: 2.27 0.0120 1.00e-05
Validation Checks: 0 10 10
Plots
Performance (plotperform)
Training State (plottrainstate)
Error Histogram (ploterrhist)
Regression (plotregression)
Plot Interval: 1 epochs
Validation stop.
Stop Training
Cancel

TRAINING & PERFORMANCE-:

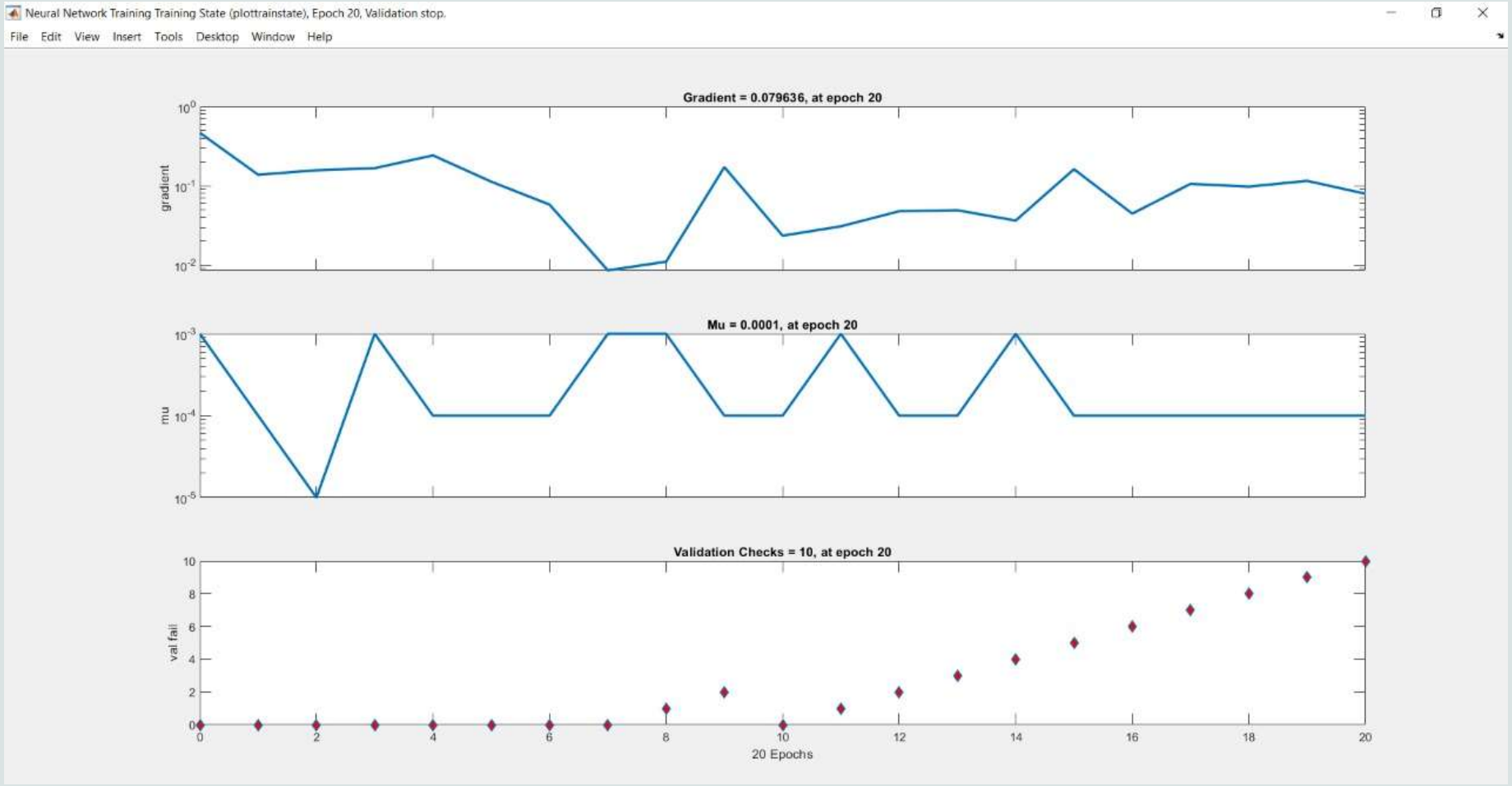




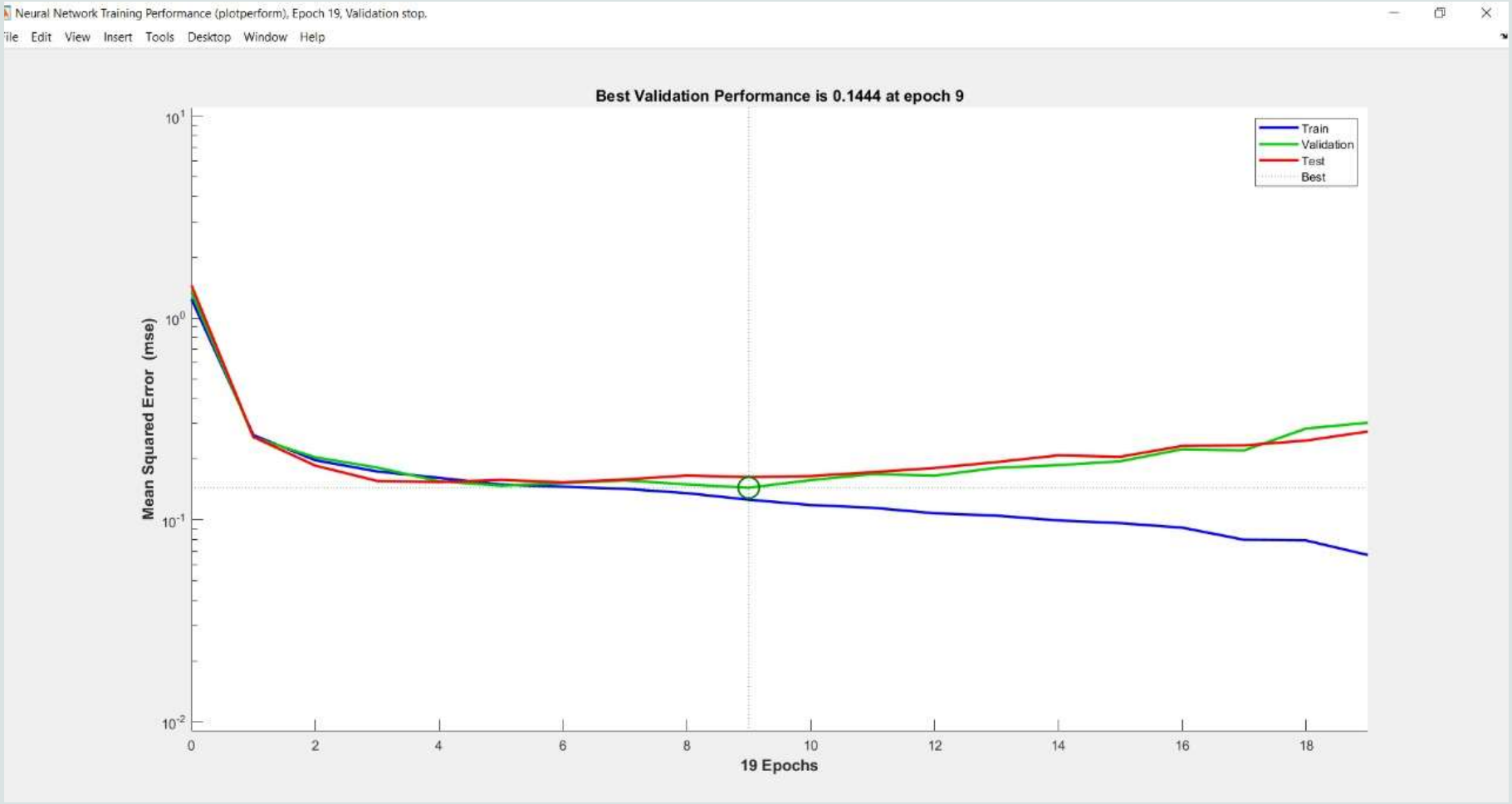
FINAL TRAINING-:



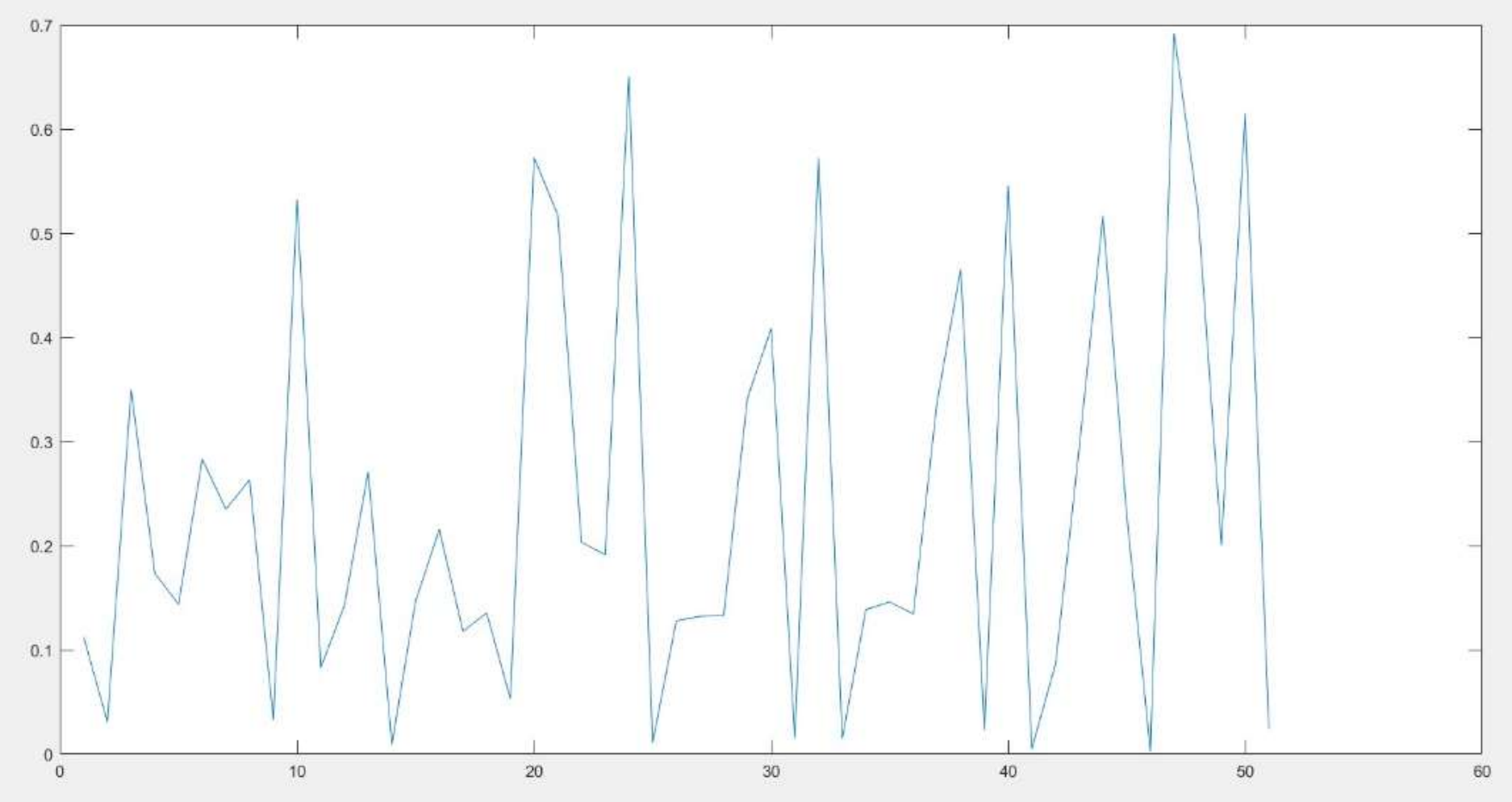
TRAINING SLATE PLOT-:



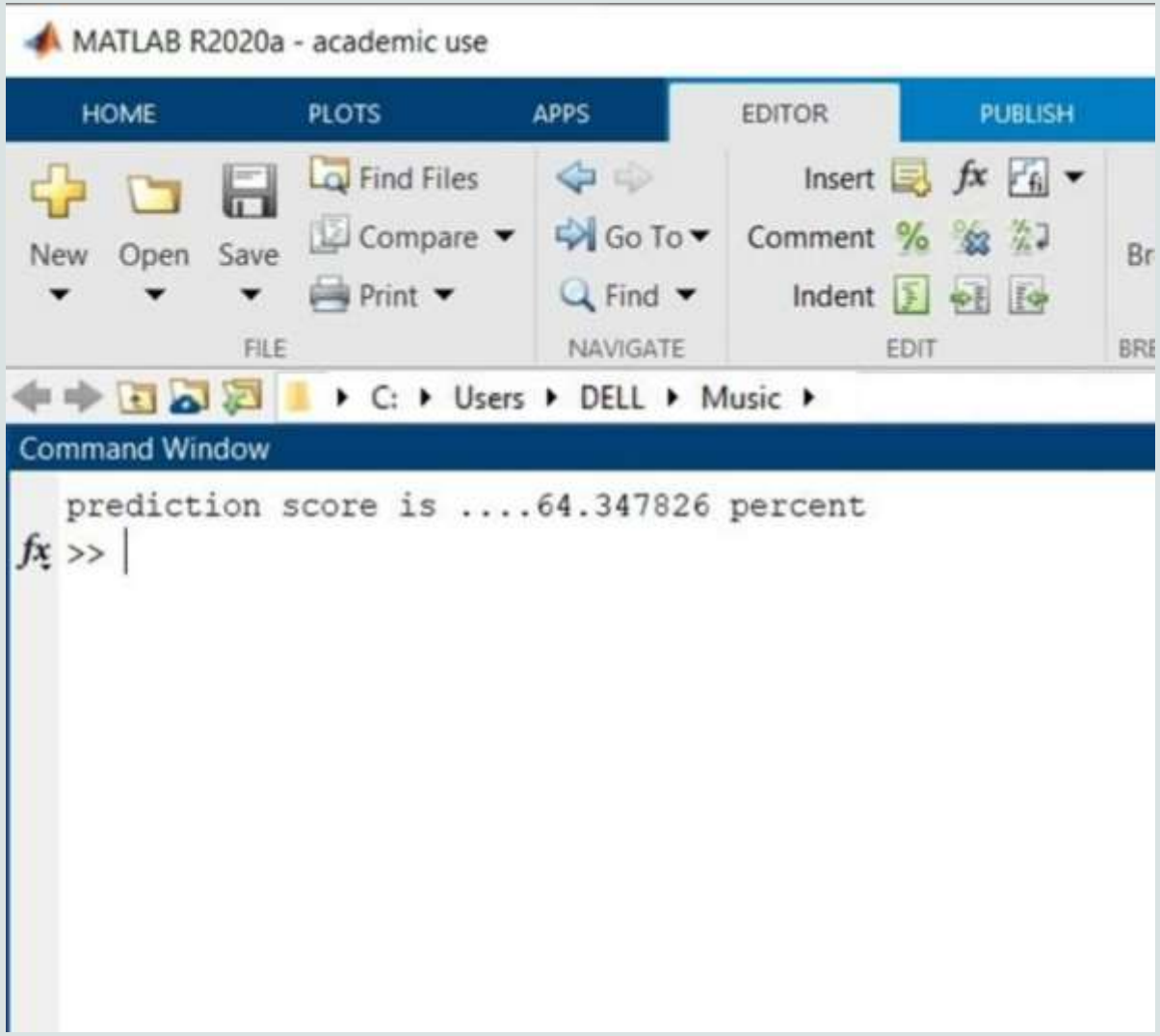
PERFORMANCE PLOT-:



PLOT FOR TARGET OUTPUT VS OUTPUT FROM NEURAL NETWORK-:



ACCURACY-:


MATLAB R2020a - academic use
HOME
PLOTS
APPS
EDITOR
PUBLISH
New
Open
Save
Find Files
Compare
Print
FILE
Go To
Find
NAVIGATE
Insert
Comment
Indent
EDIT
C: ▸ Users ▸ DELL ▸ Music ▸
Command Window
prediction score is ....64.347826 percent
>>

MSE FOR DIFFERENT SET OF DATA-:

```
      best_perf: 0.1259
     best_vperf: 0.1444
     best_tperf: 0.1631

fx >>
```

CONCLUSION POINTS-:

- MSE has been used as performance function15% data was used for validation by MATLAB
- The MSE for test data achieved is 0.1631
- Some data was separated from the main data and was used for testing.
- Error graph was plotted (Target vs Network Output) Accuracy is still a factor to be improved.

Lots of modification were done to improve the accuracy.

- Before trained training function was being used which was changed to trainlm after which the accuracy that was below 50% went up to 70%.
- Training has faster convergence rate.
- The neural network is decent enough to be used as a diabetes predictor.

REFERENCES-:

YOUTUBE:

- https://youtu.be/Ilg3gGewQ5Uhttps://youtu.be/X9ivR4y03DE
- https://youtu.be/uXt8qF2Zzfo
- https://youtu.be/Ilg3gGewQ5U
- _https://youtu.be/X9ivR4y03DE

GOOGLE:

- https://www.investopedia.com/terms/n/neuralnetwork.asp#:~:text=A%20neural%20network%20is%20a,organic%20or%20artific_ial%20in%20nature.
- https://en.wikipedia.org/wiki/Backpropagation

WIKIPEDIA
The Free Encyclopedia

Emotions are enmeshed in the neural networks of reason.

— Antonio Damasio —