CSE 574 - Introduction to Machine Learning
PA - 1 Report

Handwritten Digits Classification Using Neural Networks

Group - 37
Arvind Srinivass Ramanathan (arvindsr)
Barath Eswer Nagasubramaniayan (barathes)
Senthil Kumar Laguduva Yadindra Kumar (laguduva)

INTRODUCTION

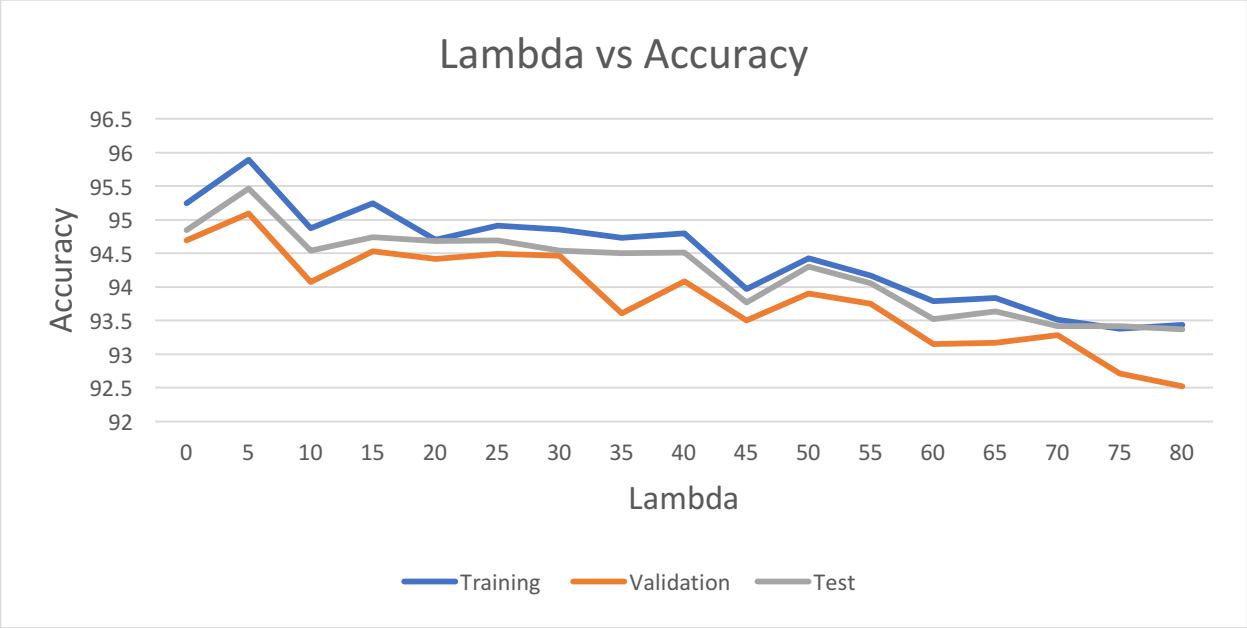
This assignment involves the implementation of a Multilayer Perceptron Neural Network and evaluating its performance in the classification of Handwritten Digits. Primarily, Feedforward and Backpropagation techniques were used in this implementation. The dataset given to us was preprocessed, and features that had matching values were removed from the training data set.

Choosing hyper-parameters for our neural network

In order to conclude on a favorable hyper-parameter value λ (regularization parameter) we applied different values for the number of hidden units and λ. The number of hidden units and λ were initially set to 50 and 0 respectively. We worked in increments of 4 between 4 and 20 and increments of 10 between 20 and 50 for the hidden units. For λ we used increments of 5 in the range of 0 to 80. The percentage accuracy of classification was measured on the training, validation as well as test data.

We analyzed the performance on the same and observed that the best performance is obtained for 50 hidden units with λ as 5.0. This value of λ performs well as it allows for the contribution of regularization to that of the loss and we set λ = 5.0 to avoid the underfitting problem where the contribution of regularization is lesser, thus the error of the test data is not accommodated. However, we also note that with increasing value of λ our performance is affected negatively thus we settle for a favorable λ value which is not too high this avoiding over fitting problem. Our observations were plotted as a graph depicting the relation between Lambda vs Accuracy for 50 Hidden Units (*Graph 1 Lambda vs Accuracy*) and Hidden Units vs Accuracy (*Graph 3 Hidden Units vs Accuracy*).

λ vs. Neural Network Accuracy %

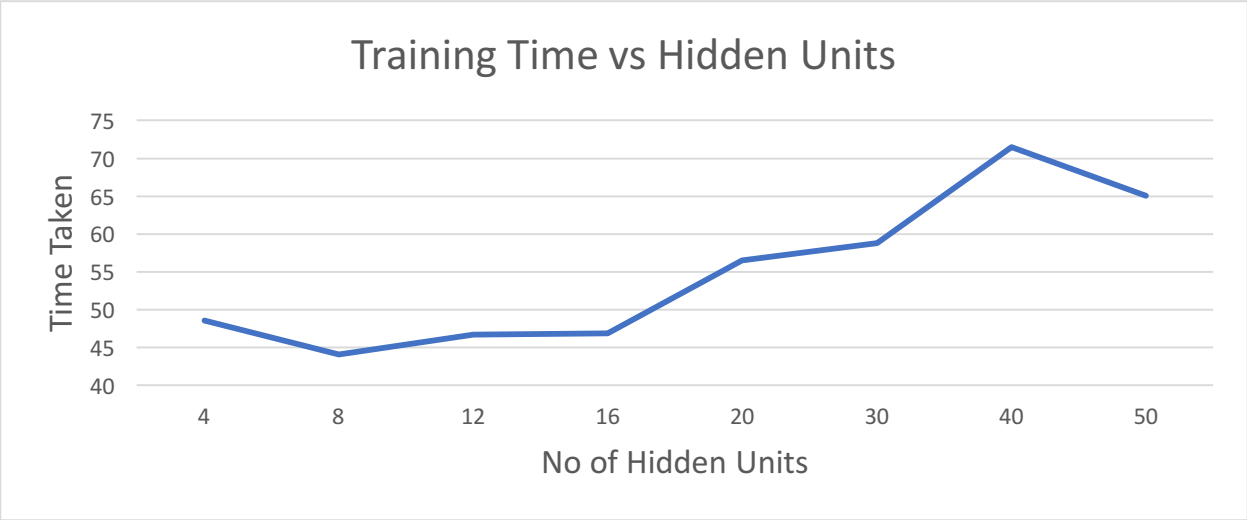


Graph 1 Lambda vs Accuracy

The Number of hidden units are : 50.

The above graph shows the accuracy of the neural network having 50 hidden units with λ in the range 0 to 80, in increments of 5. There is an increase in the accuracy at different points in the graph between 0 – 80 with an overall decrease across the range. With an increase in the λ value, we notice that the accuracy decreases gradually over the range and not steadily. We note here that with increasing λ we see a small improvement in the performance of the neural network and this increase helps us avoid the underfitting problem which occurs when the contribution of the regularization is lesser resulting in more errors. The overall decrease in each of the three values of accuracy for training, validation and test data indicates that the regularization value is not favorable for higher values of λ as the contribution of the regularization also increases thus causing overfitting. To avoid that we use a smaller value of λ. We justifiably select the λ as 5.0 to address overfitting and underfitting problem.

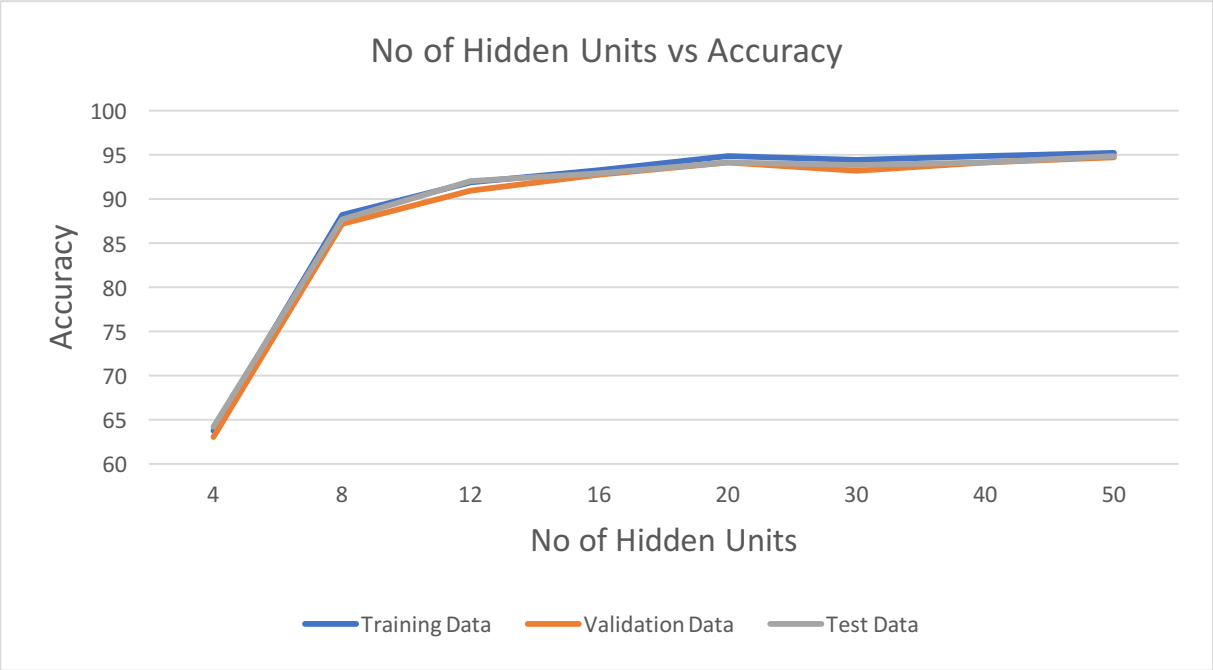
Relation between Training Time and Number of Hidden Units



Graph 2 Training Time vs No of Hidden Units

Based on the graph we observe that the training time increases with number of hidden units. The reason to this is quite intuitive as we need to process the same set of input data across more hidden units i.e., more features. This increase in the number of features cascades into computations on weight and subsequently the gradient values. This is clearly seen in how the training is slower for higher values of hidden units. We notice that the training time despite being on the higher side for 50 hidden units is favorable because of the accuracy it gives in later observations.

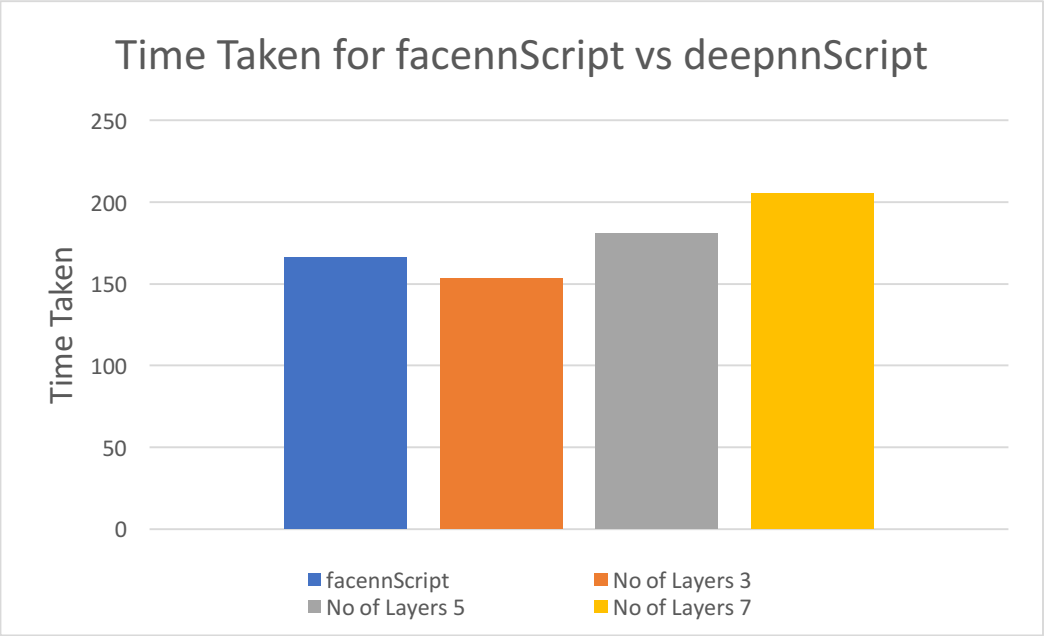
Relation between Number of Hidden Units and Accuracy



Graph 3 Hidden Units vs Accuracy

The graph above describes the accuracy of neural network for various values of hidden units over a constant λ. It is observed that the accuracy increases slowly after an initial rapid increase in the number of hidden units and is the highest for 50 hidden units. The immediate increase in accuracy between 4 and 8 hidden units is because it includes within the neural network a set of another four features or hidden units thus achieving a significant improvement in the performance of the system. However, this increase becomes relatively slower after 8 hidden units because the later features or hidden units do not contribute equally towards the accuracy of the prediction of the neural network. The highest value for performance of the test data was seen at 94.85.

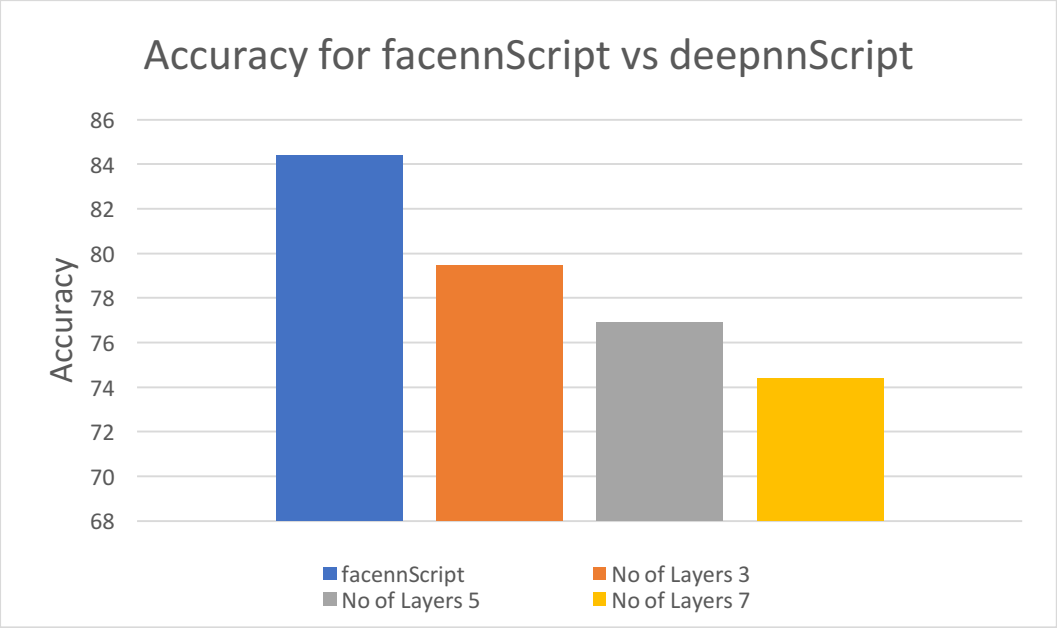
Relation between Time Taken for deepnnScript and facennScript:



Graph 4 Time Taken for Layers

The above column graph shows the time taken to train the neural network when the number of layers of the network was increased. The facennScript runs a neural network with only one layer whereas the deepnnScript runs the neural network for multiple layers. The observation from this graph is that there is an overall increase in the time taken to train the system across more layers. This observation can be reasoned as the system experiences an increase in the computations performed when the number of layers is increased.

Relation between Accuracy for deepnnScript and facennScript:



Graph 5 Accuracy for Layers

The above graph shows the variation in the accuracy for different layers of the neural network. Here the number of layers in the network is increased and we see that the accuracy decreasing steadily. Thus we see that the most accurate predictions come from the lesser number of layers.

Conclusion

We understand from the graphs that with increasing number of hidden units there is also an increase in the accuracy of the system. However, with increasing hidden units the computation also increases. This increase in the computation will lead to longer training times for the neural network. However, a tradeoff between the accuracy and the training time can be made and we choose **50 hidden units** as it gives us the most accurate value. Another trend observed is the relation between λ and the accuracy. We choose an appropriate value of λ after handling overfitting and underfitting. We find this value to be **5.0**.