Handwritten Digits Classification

Introduction:

The first part of project involves implementing a Multilayer Perceptron neural network and evaluating its performance in classifying handwritten digits. The other part of this assignment is involved in analyzing a more complicated face dataset that contains celebrity images by which the Neural network classifies whether individual is wearing glasses or not wearing glasses and compared the performance of the neural network against a deep neural network using TensorFlow library.

We experimented with the hyper-parameters to find the suitable values for the performance of the neural network. We varied the number of hidden nodes between 5,10,20,30,40 and 50. We measured the training time and the accuracy of the neural network for the given datasets and used these values to choose the optimal hyper-parameters to give best performance. We also varied the lambda value from 0 to 50 in increments of 10.

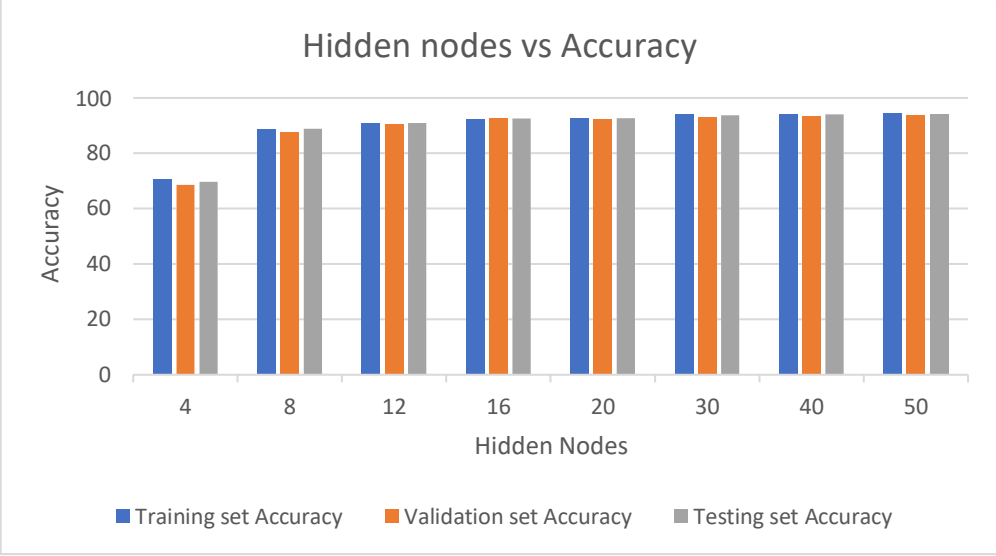
Observations and Results:

Tuning the hyper-parameters for neural network for Handwritten Digit classification (number of hidden nodes and Regularization factor (lambda)):

We experimented on the hyper parameters to determine they can influence the accuracy and training time of the neural network. In general, choosing a hyper parameter can be done on a random search basis, grid search, etc. We tried to vary the hyper parameters and number of hidden nodes to study the consistency we can get in training the model.

1. **Relation between number of hidden nodes and the performance of the neural network:**

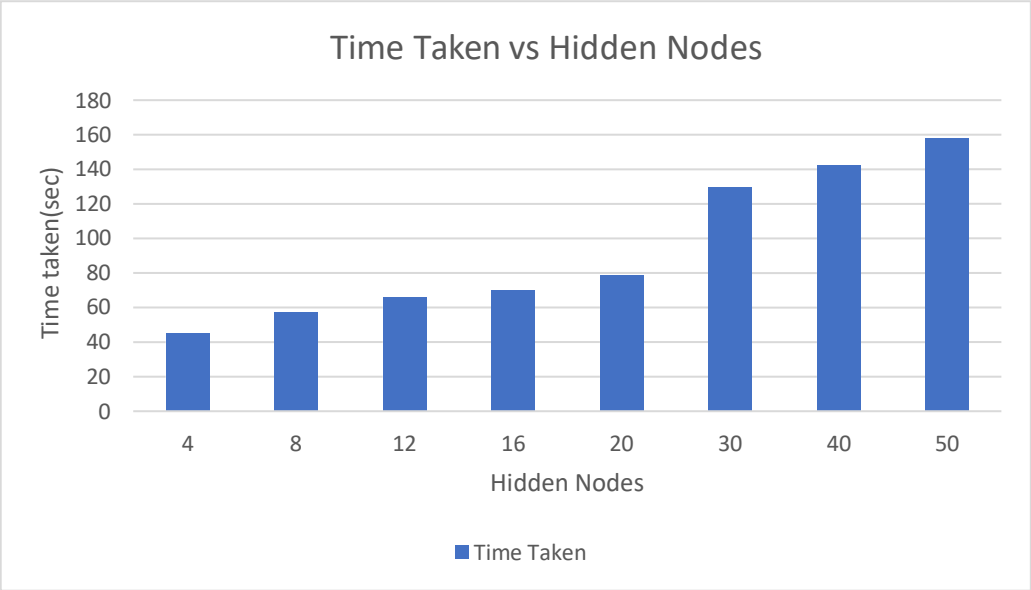
   We varied the number of hidden nodes from 4 to 20 in increments of 4 and from 20 to 50 in increments of 10.



As the number of hidden nodes increases, the accuracy has also increased. The accuracy increases because more number of hidden nodes results in higher precision. So, when the accuracy of the system is desirable, we can increase the number of hidden nodes. But there will not be any significant increase in the accuracy of the neural network after a certain value of the hidden nodes.

We started with 4 hidden nodes where we got an accuracy of 69.65% which is quite low. So, we increased the number of hidden nodes in steps of 4 until 20 and found an increase in the accuracy. The accuracy became consistent after 12 hidden nodes until 50 hidden nodes. The training dataset, test dataset and validation dataset had the best performance when the number of nodes was set to 50 with an accuracy of 94.43% after which the accuracy began to dip due to over fitting.

2. **Relationship between number of hidden nodes and time taken to train the network:**

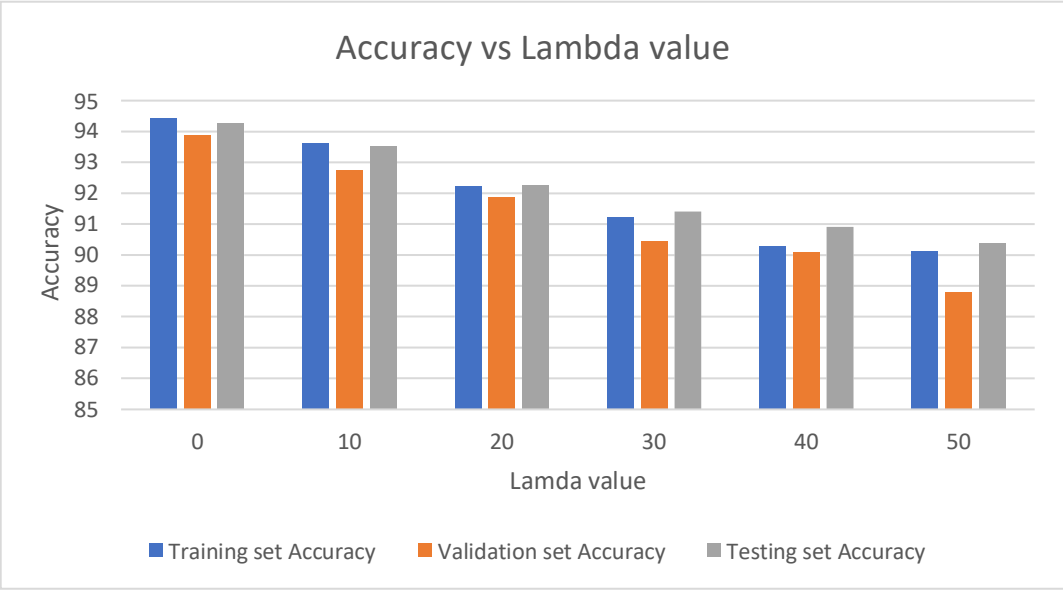


We checked how the time taken to train the network varies based upon the number of hidden nodes present in neural network. We see that, as the hidden nodes were increased the time taken to train the network also increased.

This is because when there is more number of hidden nodes, we have more weights and gradients to compute, resulting in more complexity in the computations and a longer time is needed to converge.

3. **Relationship between regularization parameter λ and accuracy:**

We observed that the results are optimal when the number of hidden nodes is set to 50. We then varied the λ values and tried to determine how it influences the accuracy of the training, validation and test data.



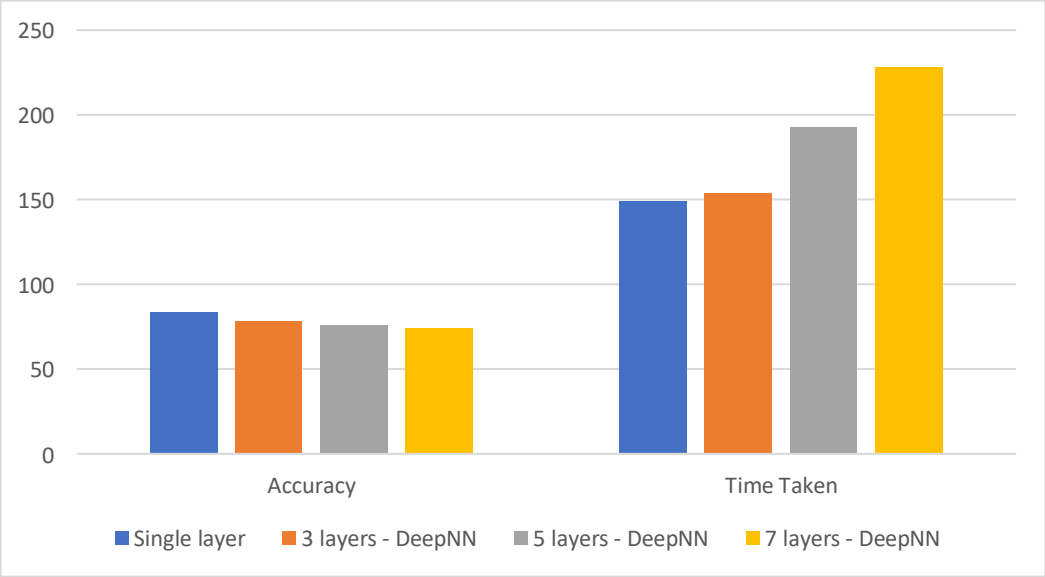
We can see that as the regularization parameter λ increases, the accuracy of the neural network model decreases. This is because regularization involves learning rate, momentum while calculating gradient descent, etc. We were able to evidently see the decrease in accuracy as the hyper parameter crosses the value of 10.

According to the experiment, we found that $\lambda < 10$ gave us consistent results on accuracy. Also, increasing the values of λ to larger values lead to reduction in the accuracy.

Comparison of neural network and deep neural network on the Celebrity data set:

We used the neural network with single hidden layer on a complex face dataset to distinguish between two classes - wearing glasses and not wearing glasses. Firstly, we ran the file FaceNN.py by implementing the sigmoid function, objective function and predict function of the previous implementation for the Celebrity data set. We used a deep neural network(DeepnnScript.py) also on the same dataset and compared the results for 1,3,5 and 7 layers using TensorFlow library.

The obtained results of accuracy and time taken for the experiments were as follows:



We notice that the accuracy decreases, and time taken to train increases as the number of layers were increased. As the model gets deeper, it becomes more difficult to obtain good generalization using a deep neural network and it becomes hard to train.

More number of hidden layers results in more number of gradients which results in the increase in computation resulting in an increase in the training time.

Conclusive results:

Test data Accuracy on Digits Classification - 94.28% ($\lambda = 0$, Hidden nodes = 50)

Test data Accuracy on Facial - Glasses recognition - 83.64%