# CSE574 Introduction to Machine Learning

# Programming Assignment 1

# Handwritten Digits Classification

Group 60

Himal Dwarakanath (himaldwa)

Manish Kasireddy (manishka)

MuthuPalaniappan Karuppayya (muthupal)

## Introduction

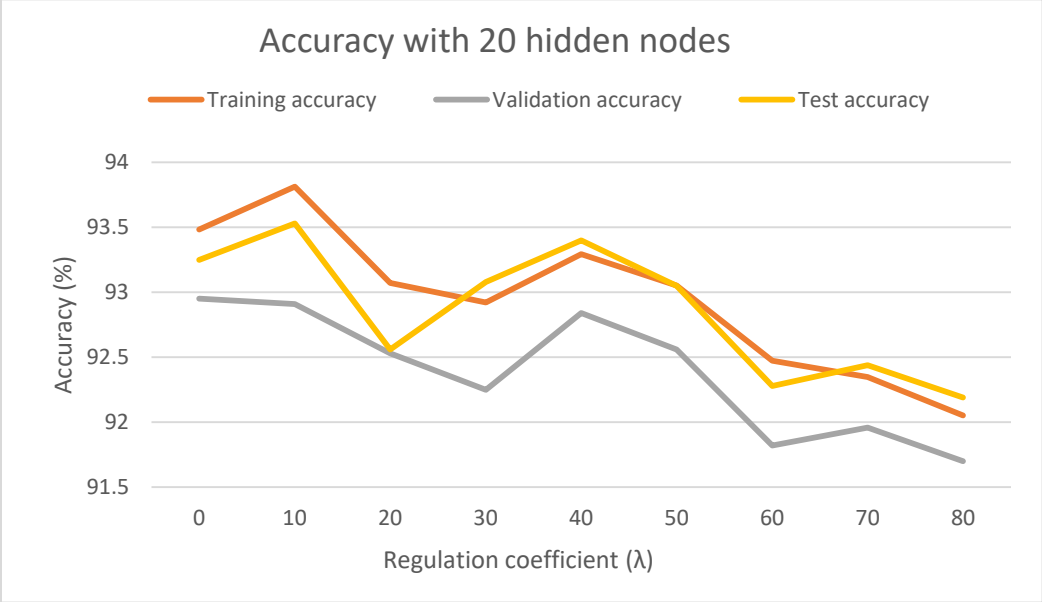
We have implemented a Multilayer Perceptron Neural Network and evaluated its performance in classifying handwritten digits. We used the same network to analyze a more challenging face dataset and compared the performance of the neural network against a deep neural network using the TensorFlow library.

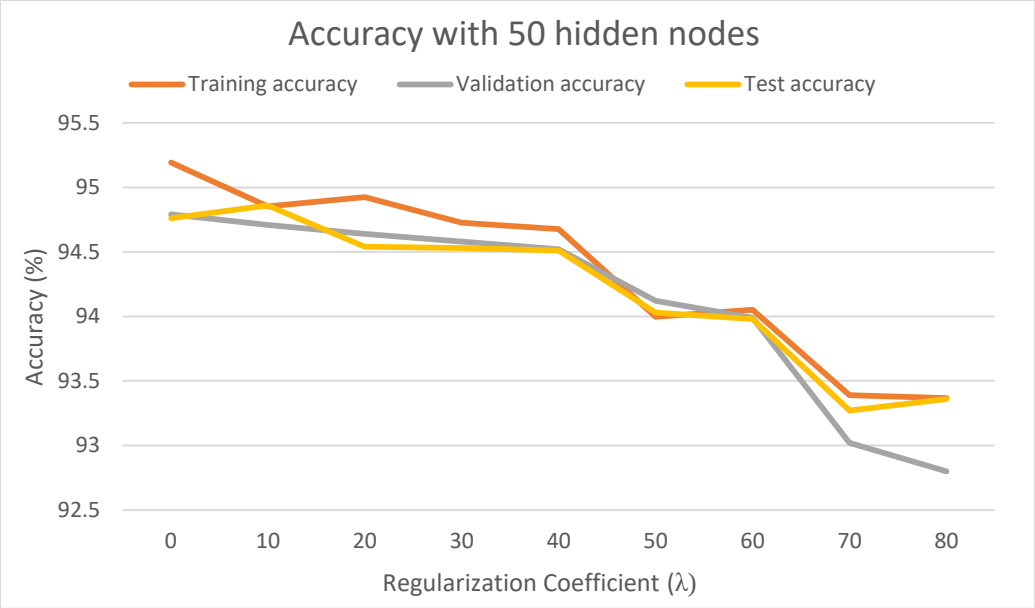
We varied the value of the hyper-parameters to find optimal values for the performance of the neural network. We varied the λ value from 0 to 80 in increments of 10. We also varied the number of hidden units between 4, 8, 12, 16, 20, 30, 40 and 50. We measured the accuracy and training time of the neural network on the given input datasets and used these values to choose the hyper-parameters to give the best performance.

## Tuning the hyper-parameters for the Neural Network for Handwritten Digits Classification

### 1. Relation between hyper-parameter λ and the performance of the neural network

We plotted the accuracy of the neural network against different values of λ and varying the number of hidden nodes.



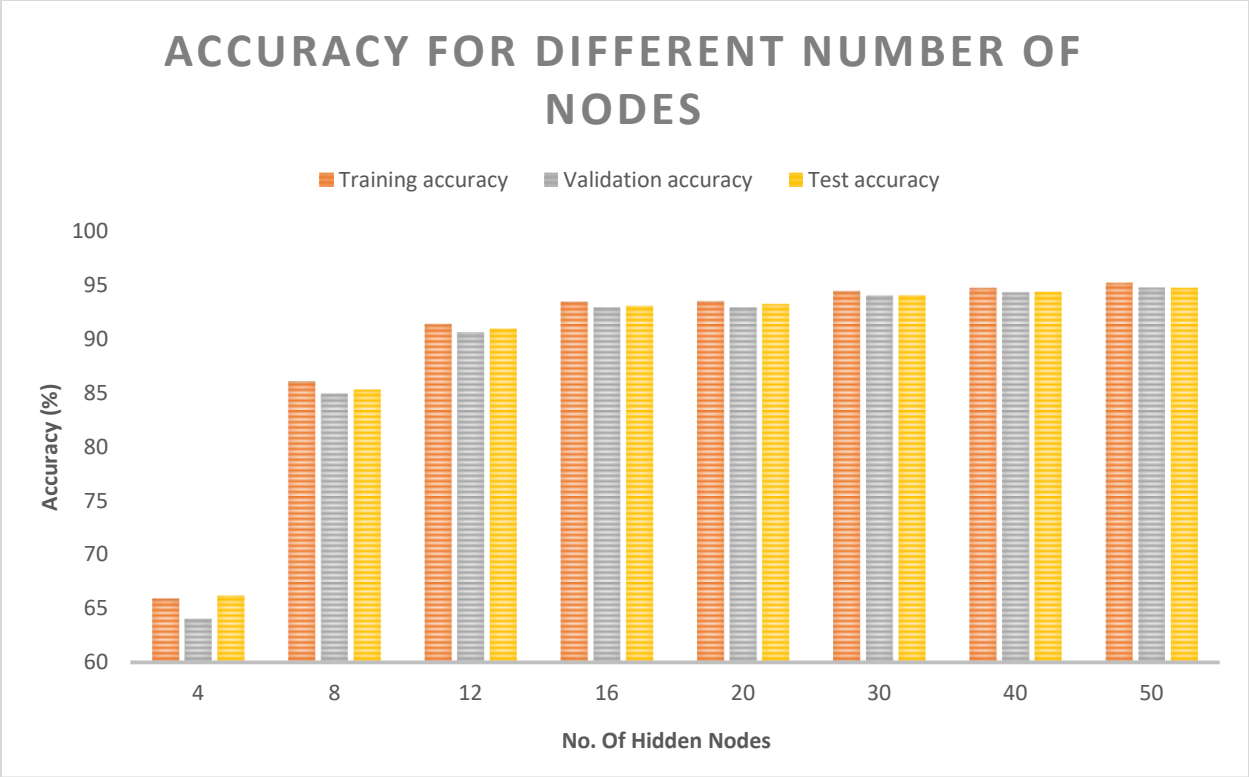



Generally, the accuracy on the training data decreases when the value of λ is increased. But as seen from the graph, the accuracy doesn't decrease continuously even when λ is increased because we chose the initial weights randomly. Overall, there is not much change in the accuracy when λ is varied. We got the highest accuracy when λ = 10 i:e 94.86%.

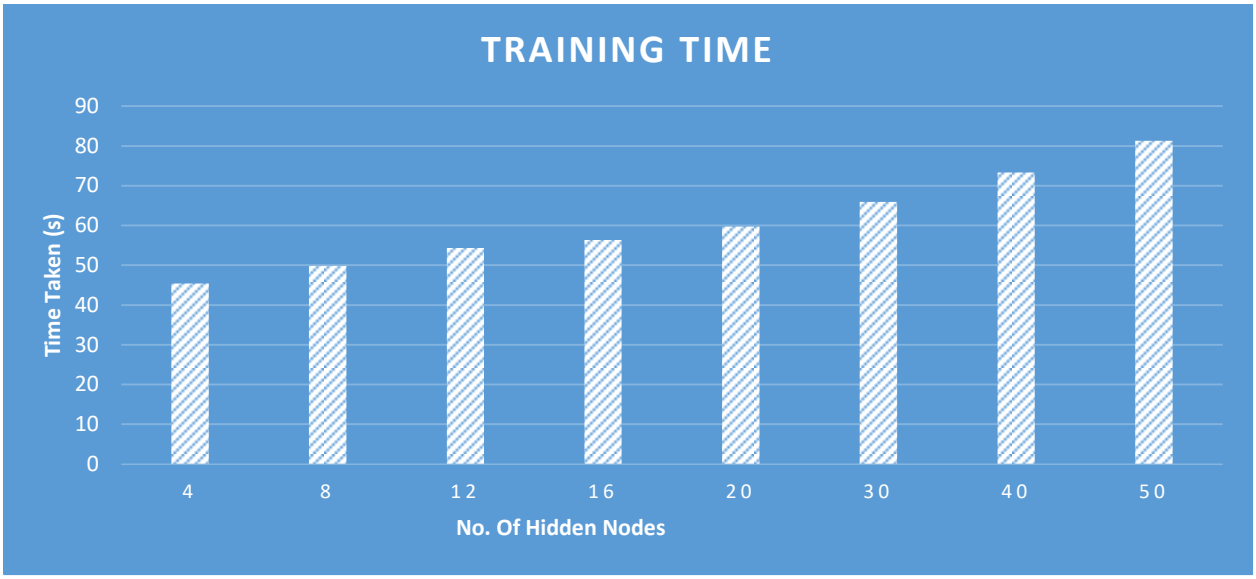
We use regularization in neural network to avoid underfitting and overfitting problem. The similarity between training and test data may cause the performance of the test set not to be affected by overfitting on the training set. Increasing the value too much will result in underfitting, so to avoid that we choose the value so that too much regularization isn't done (which results in underfitting problem.

## 2. Effect of number of hidden nodes on the performance of the neural network and training time

As the number of hidden nodes increases, the accuracy is expected to increase but the training time of the neural network also increases. 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 we keep increasing the number of hidden nodes over a certain value.




We started with 4 hidden nodes where we got an accuracy of 65% which is quite low. So we increased the number of hidden nodes to 8 and found an increase in the accuracy. From 12 nodes, we got a steady accuracy of over 90%. Finally, we chose the number of hidden nodes to be 50 to get the maximum accuracy.



As seen from the graph above, more number of hidden nodes results in higher training time. 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.

**Choosing the optimal hyper-parameter values:**

As shown in above graphs, we have varied the hyper-parameter values and we have arrived at optimal performance for the below values.

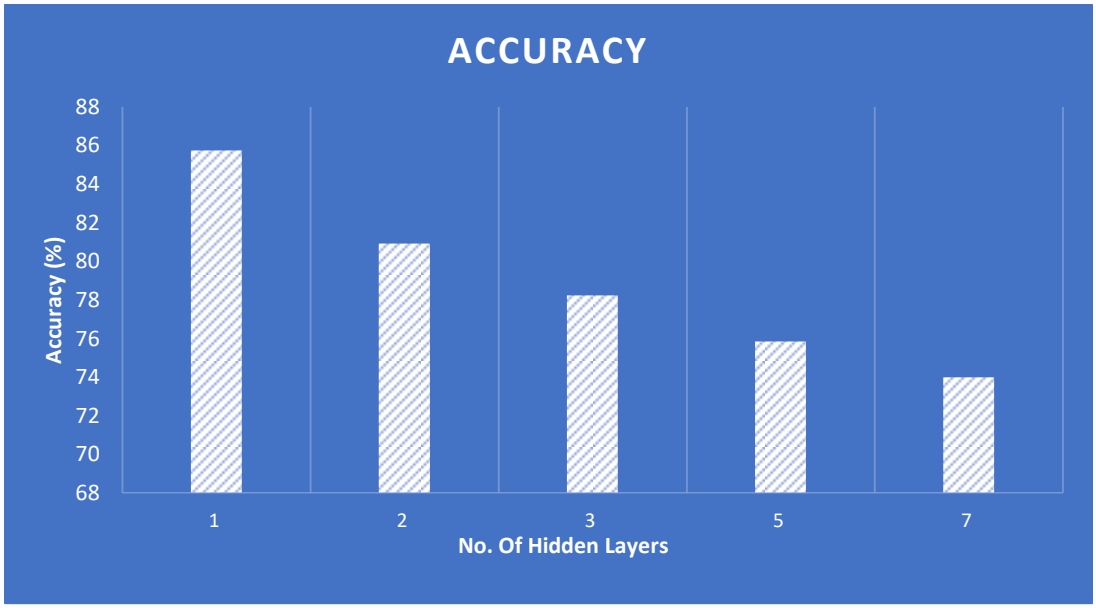
$\lambda = 10$

Hidden nodes = 50

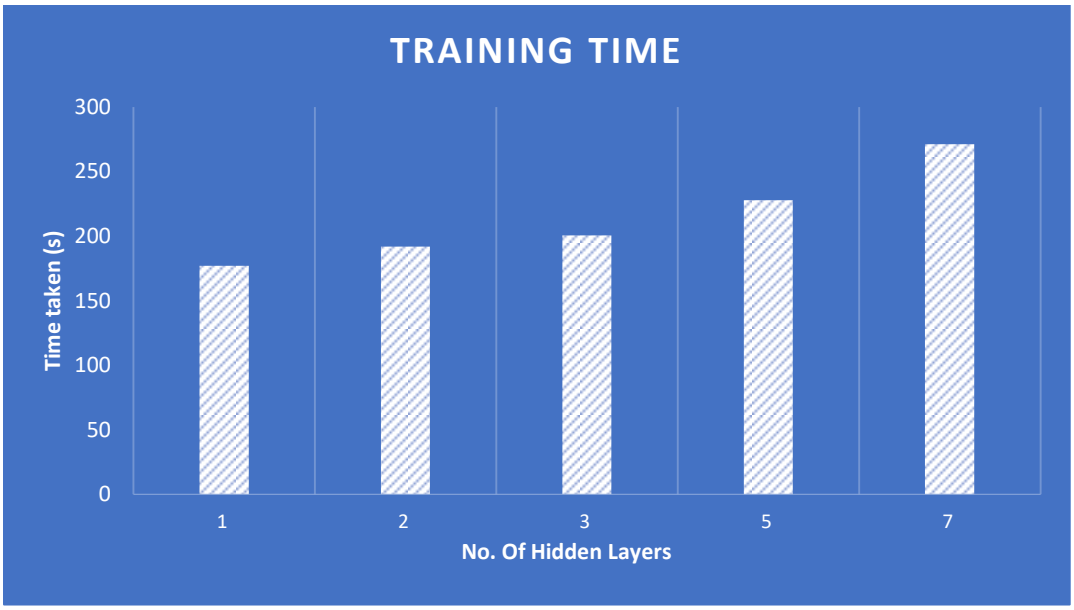
Accuracy for test data = 94.86%

## Comparison of the neural network with a deep neural network (using TensorFlow):

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. We used a deep neural network also on the same dataset and compared the results.




As seen from the graph, accuracy is more when neural network with a single hidden layer is used as compared to the deep neural network. The accuracy using the deep neural network decreases with increase in the number of hidden layers. As the architecture gets deeper, it becomes more difficult to obtain good generalization using a deep neural network.




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.

| Number of hidden layers | Accuracy (%) |
|---|---|
| 1 | 85.74 |
| 2 | 80.92 |
| 3 | 78.23 |
| 5 | 75.85 |
| 7 | 73.99 |

**Conclusion:**

For neural network on handwritten dataset, we have achieved an accuracy of 94.86% with the hyper-parameters set to λ = 10 and hidden nodes = 50. We have also compared single and multi-layered neural network on face dataset.