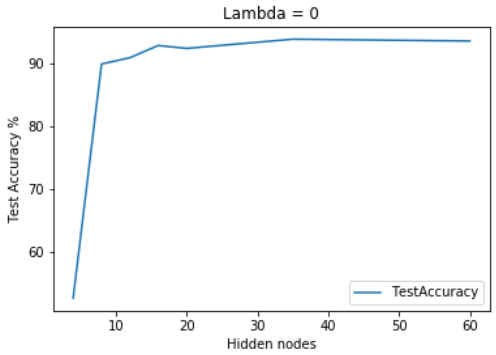
Assignment 2 report

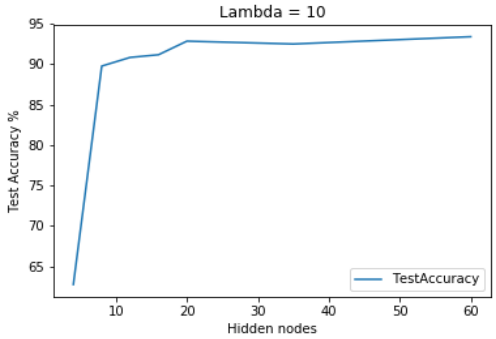
CSE 574: Introduction to Machine Learning
Members: Vineel Kurma, Preeyank Pable, Eswar Chand Jonnakuti

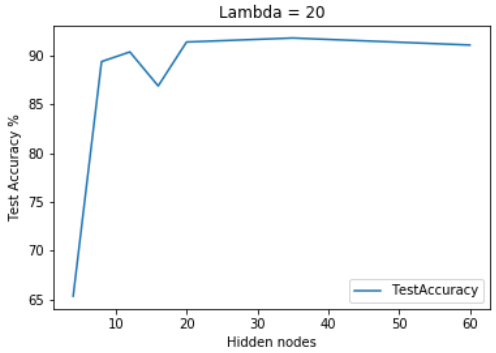
Neural network (1 Hidden layer) on MNIST Dataset:

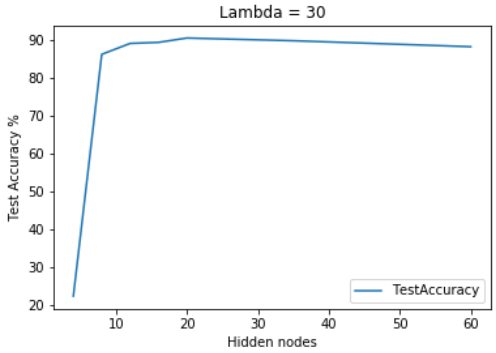
Feature Selection: All the cells in the MNIST dataset with values less than 10 have been removed to improve the processing speed and remove unwanted data. Final no. of features left after pre-processing are 647, which are stored in the pickle file. So the total unwanted noisy features amount to = 784-647 = 137.

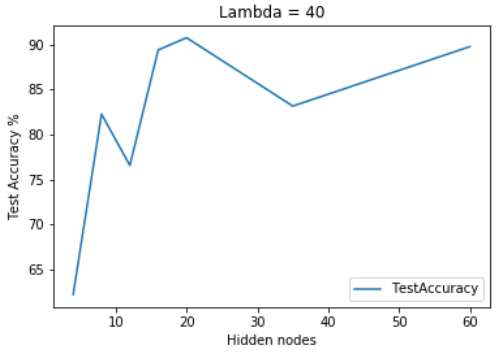
To obtain the optimal hyper-parameters for neural network with the highest accuracy, the regularization parameter (λ) and no. of hidden nodes for the hidden layer need to be iterated till we obtain the values we desire. The iterations were run for the following combinations:
λ : 0, 10, 20, 30, 40, 50, 60
No. of Hidden Nodes: 4, 8, 12, 16, 20, 35, 60

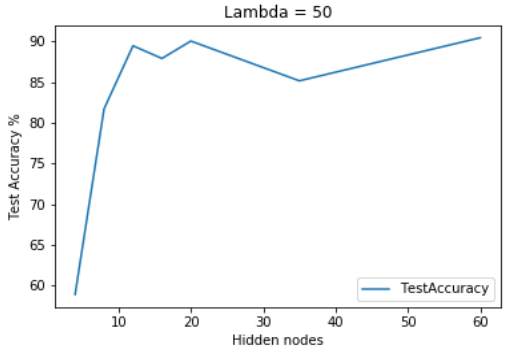


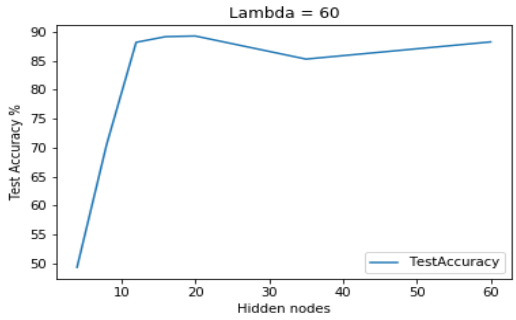




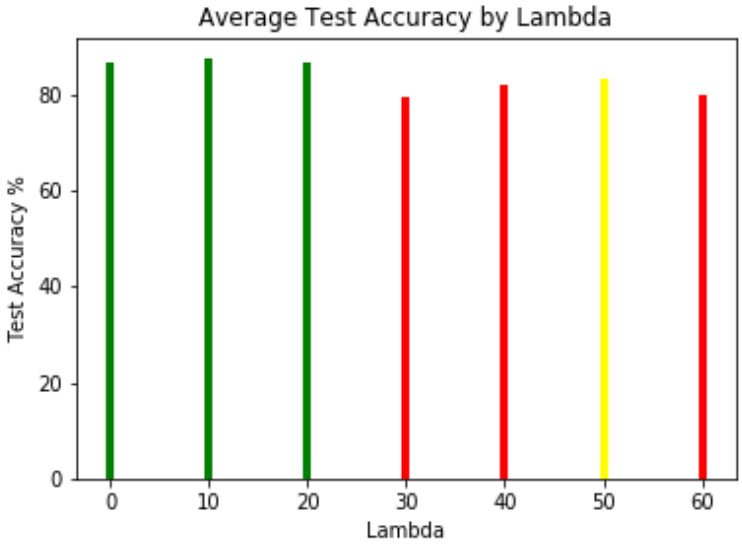


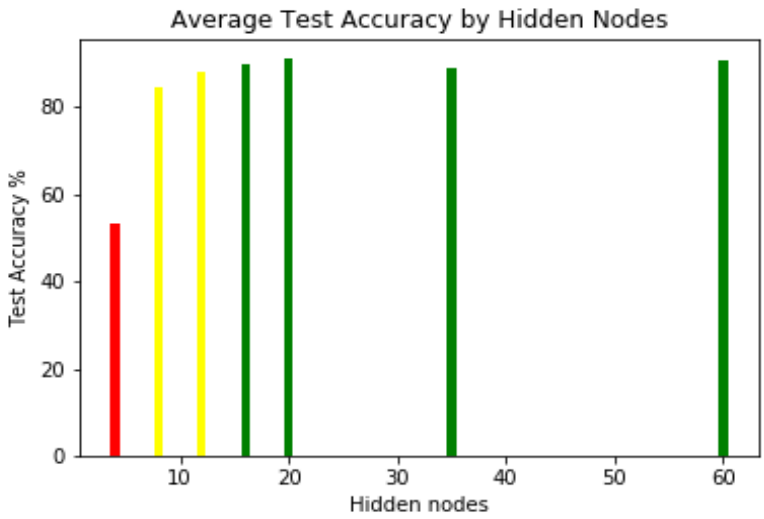






We can see from the above trends that the highest number of Hidden Nodes (60) didn't give the highest accuracy most of the times. **Now to choose the parameters, we need to look how they performed overall in all the cases instead of just considering the best-case scenario. i.e. we shouldn't just blindly take the parameters that give out the best accuracy.**





The above two plots show how the Average Test accuracy is varying among different *Lambdas* and *No. of Hidden Nodes*.

Hypothesis: We can see that the **regularization parameters with lower values** and **Hidden Nodes with higher values** tend to have higher average test accuracies. The case with the highest accuracy of 93.20% (Hidden Layer nodes = 60 & λ = 10) is falling in above hypothesis region.
Hence, we can conclude with confidence that **Hidden Layer nodes = 60 & λ = 10** would be the best set of hyper-parameters in our case.

Neural network (1-Hidden layer) on CelebA Dataset:
Classified the images from CelebA dataset using the previous algorithm to run the 1 hidden layer neural network. The No. of Hidden nodes was set to be 256. The test set accuracy came to about:
84.56% for lambda = 10 with a training time of 5.45 minutes
85.43% for lambda = 0 with a training time of 4.92 minutes.

Deep Neural Network on CelebA Dataset:
Used tensor flow for running the deep neural network on the CelebA dataset. As suggested, ran the iterations for No. of Hidden layers : 2, 3, 5 & 7 for different combination of hidden layer nodes. The following accuracies were obtained. We can see that the highest accuracy belongs to **Hidden layer=5.**
Parameters used: Learning rate = .001, Training Epochs = 100, Batch Size = 100

Deep NN				
Hidden Layers	Hidden Layer Nodes	Test Accuracies	Time in Seconds	Time in Minutes
2	256, 256	83.72%	550	9.17
3	256, 256, 200	79.45%	620	10.33
5	256, 256, 200, 150, 100	84.67%	681	11.36
7	256, 256, 200, 150, 100, 80, 70	79.14%	709	11.81

Comparison between 1-Hidden layer NN vs. Deep NN on CelebA dataset: We can clearly see from the above results that **1-Hidden layer NN (Accuracy: 85.43% ; Training time: 4.92 minutes)** is performing better than **Deep NN (Accuracy: 84.67% ; Training time: 11.36 minutes)** in terms of both accuracy and training time. This might be occurring due to overfitting in Deep NN.

CNN on MNIST Dataset:
Convolutional Neural Network was run on MNIST dataset.
Accuracy = 98.8%
Total time = 19 minutes
We can see that CNN performed way better than 1-Hidden layer NN even though CNN took a lot of time to process (15 minutes more than 1-Hidden layer NN).