# CSE 574: Programming Assignment 2

# Handwritten Digits Classification

**Team:**

**Abhinav Kumar: 50247660**

**Raman Keswani: 50246297**

**Explanation of how to choose the hyper-parameters for Neural Network (number of hidden units, regularization term λ).**

1. We would want our training to be faster so that we can experiment more with hyper-parameters. So, first step should be to create a light weight version of problem at hand.

2. We can speed up training by working on grayscale images and removing features which are similar for all the classes of data.

3. Another way to increase performance is to train images in batches (say on 10,000 images instead of 50,000), but at the same time we should ensure that our training model sees sufficient amount of images.
(All these above steps are just to make training faster so that we can have much more rounds of experimentation)

4. **Number of hidden nodes** depend upon the complexity of problem and size of input. Too few nodes can result in under fitting and too many nodes can result in over fitting and slower training. The basic rule is that the more the number of hidden units, the more features we can capture. But increasing beyond a certain limit will result in over-fitting.

5. A known general thumb rule is to that the number of hidden nodes should be between number of input nodes and number of output nodes. Another good practice is to start with minimum number of hidden nodes and keep increasing until you get the best results. Number of hidden nodes are usually determined empirically through cross validation.

6. We know **that regularization Parameter Lambda** is used to control over fitting. So one strategy is to start with value 0 for lambda and gradually increase it. We stop increasing it when we observe that test/validation accuracy goes down. Picking lambda value is a heuristic strategy and can be picked as one which gives us maximum accuracy for both testing and validation data.

- **Accuracy of classification method on the handwritten digits test data :** 94.95%
- **Accuracy of classification method on the CelebA data set :** 86.7146101438%

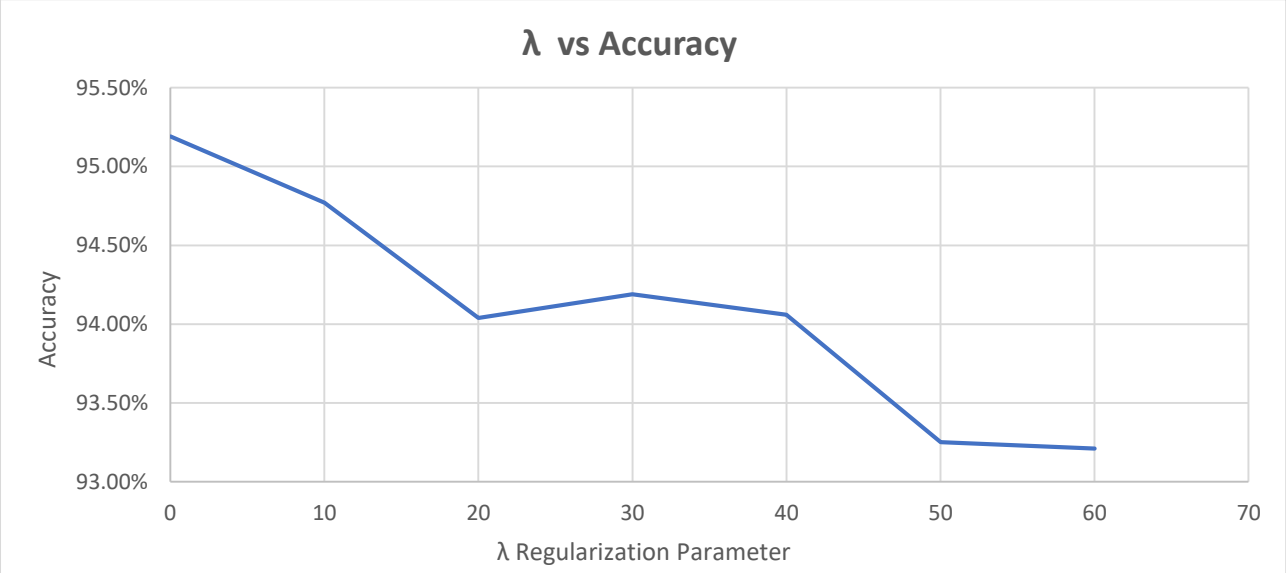
**We use regularization in Neural Network to avoid overfitting problem (more about this will be discussed in class). You are expected to change different value of λ to see its effect in prediction accuracy in validation set. Your report should include diagrams to explain the relation between λ and performance of Neural Network. Moreover, by plotting the value of λ with respect to the accuracy of Neural Network,**

**You can vary λ from 0 (no regularization) to 60 in increments of 5 or 10**

In order to assess effect of **λ.** we first optimized the model for number of hidden nodes and number of iterations/epochs. And we got best accuracy for the **hidden nodes to 100** and **number of epochs to 50**.

Now we calculate the below validation accuracy by tuning **λ from 0 to 60** for above values of hidden nodes and iterations**:**

| Lambda | validation accuracy |
|---|---|
| 0 | 95.19% |
| 10 | 94.77% |
| 20 | 94.04% |
| 30 | 94.19% |
| 40 | 94.06% |
| 50 | 93.25% |
| 60 | 93.21% |

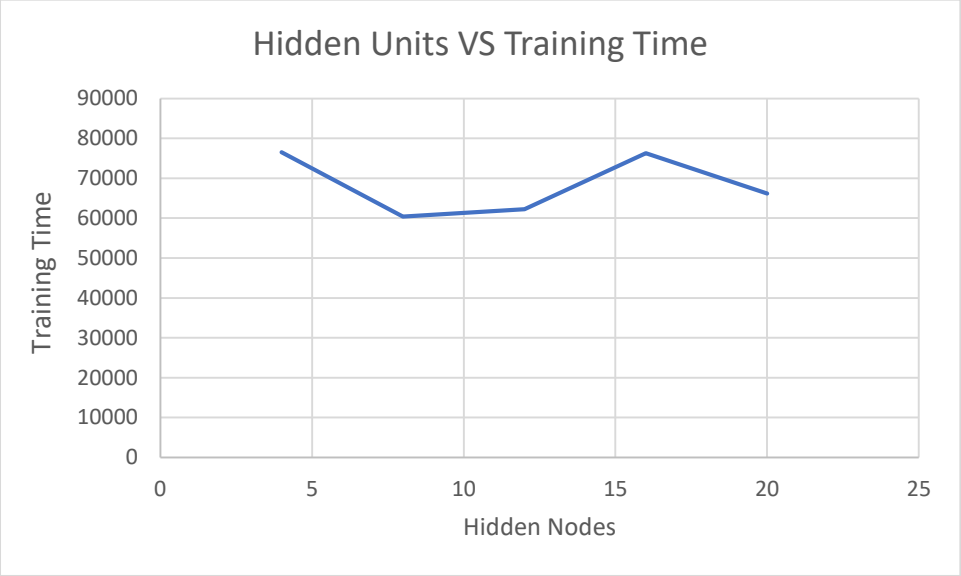



So as evident, **optimal value of λ is set at 30**, after which the accuracy dips.

**You are also expected to try different number hidden units to see its effect to the performance of Neural Network. Since training Neural Network is very slow, especially when the number hidden units in Neural Network is large. You should try with small hidden units and gradually increase the size and see how it effects the training time. Your report should include some diagrams to explain relation between number of hidden units and training time. Recommended values: 4,8,12,16,20**

We get the following results:

| Hidden Units | Training Time |
|---|---|
| 4 | 76574 |
| 8 | 60410 |
| 12 | 62244 |
| 16 | 76329 |
| 20 | 66186 |



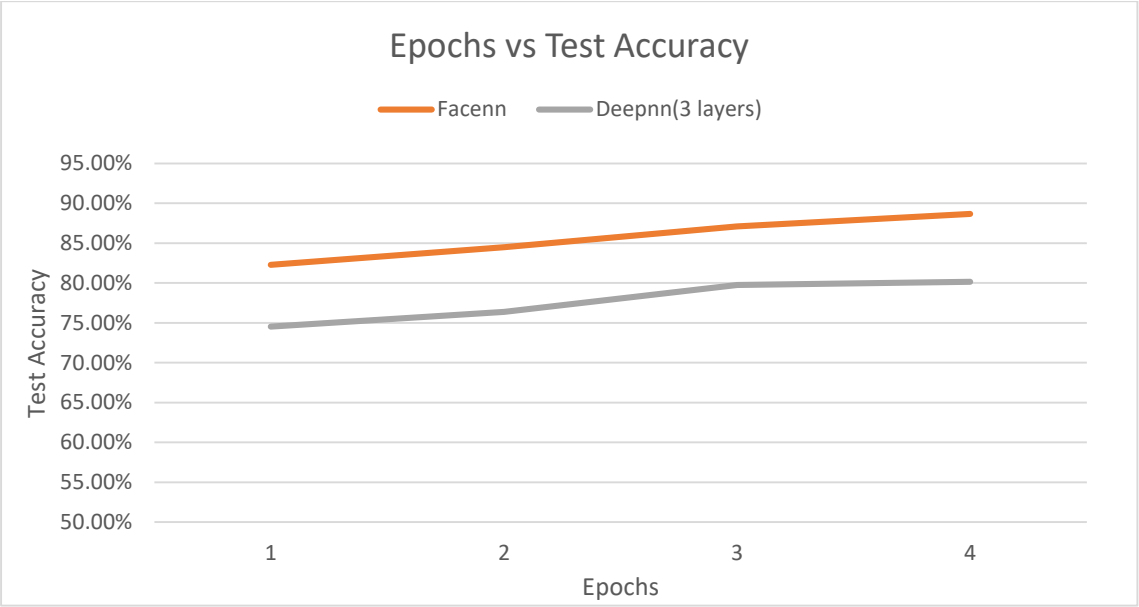

We don't see that increasing hidden nodes directly increases training time. We get lowest training time for hidden nodes = 8 and highest training time for hidden nodes = 4. This could probably because of the fact that: Small number of hidden nodes cannot capture the appropriate number of features because of the simplicity of the model and hence it takes more time to converge.

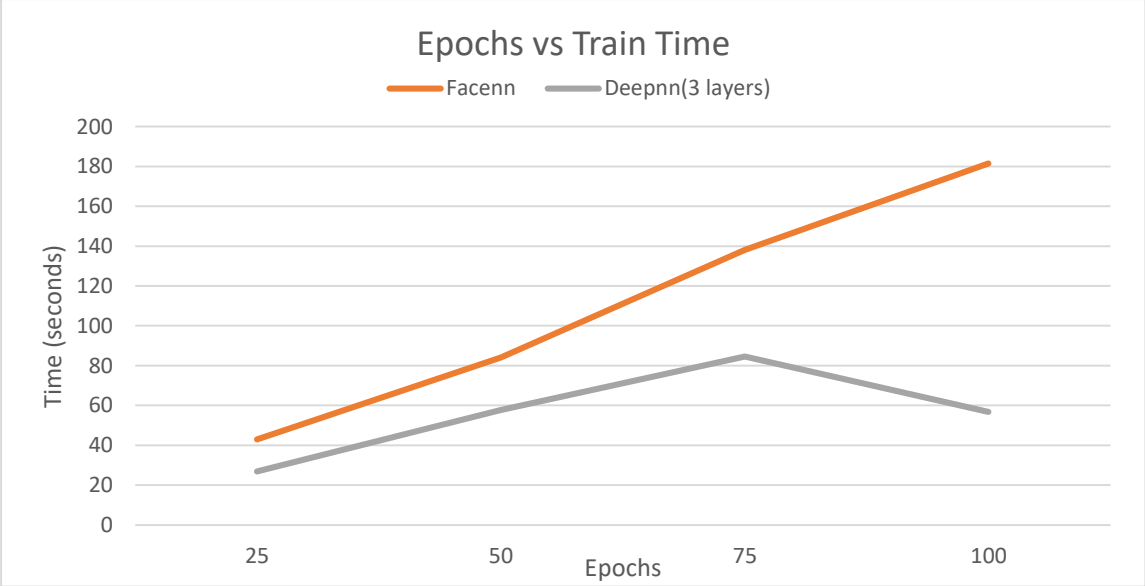
**Compare the results of deep neural network and neural network with one hidden layer on the CelebA data set**.

For conducting this experiment, we set the number of hidden nodes to be 100 for both single layer and deep neural networks. We make 3 comparisons between 1 hidden layer neural network and deep neural network.

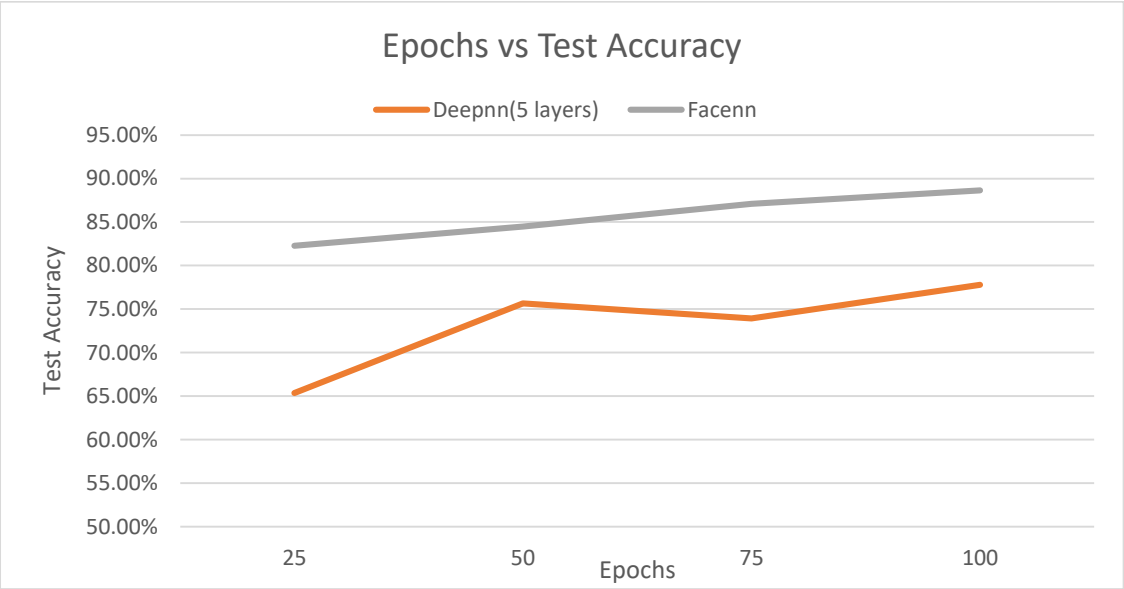
We plot accuracy of model and train time vs number of epochs for each comparison

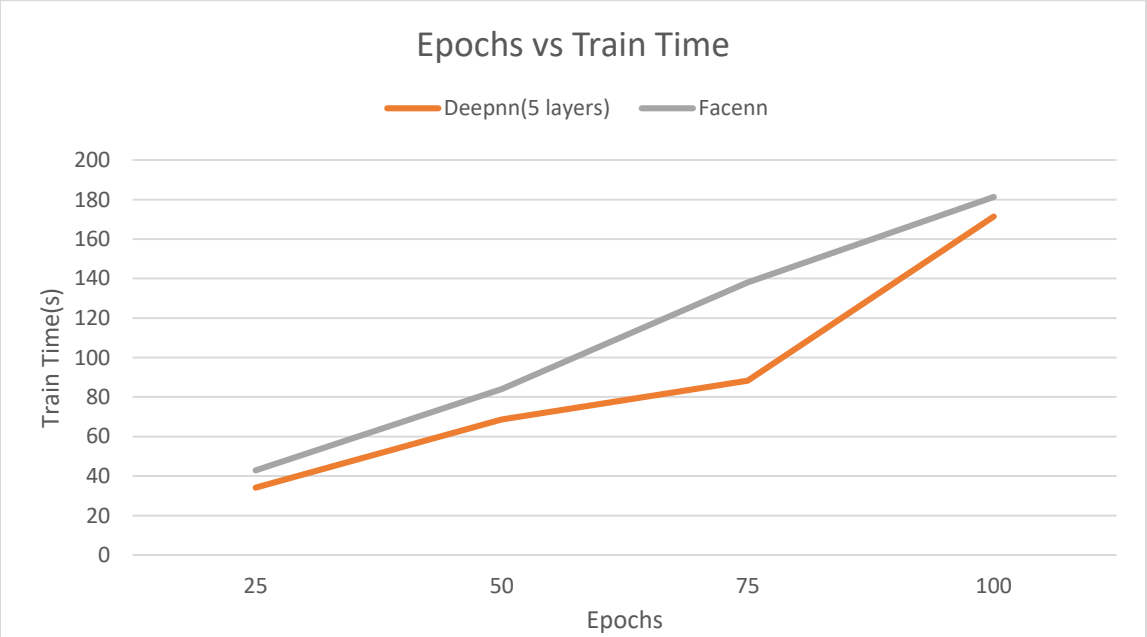
1. Single Hidden Layer Neural Network VS 3 Hidden Layer Deep Neural Network



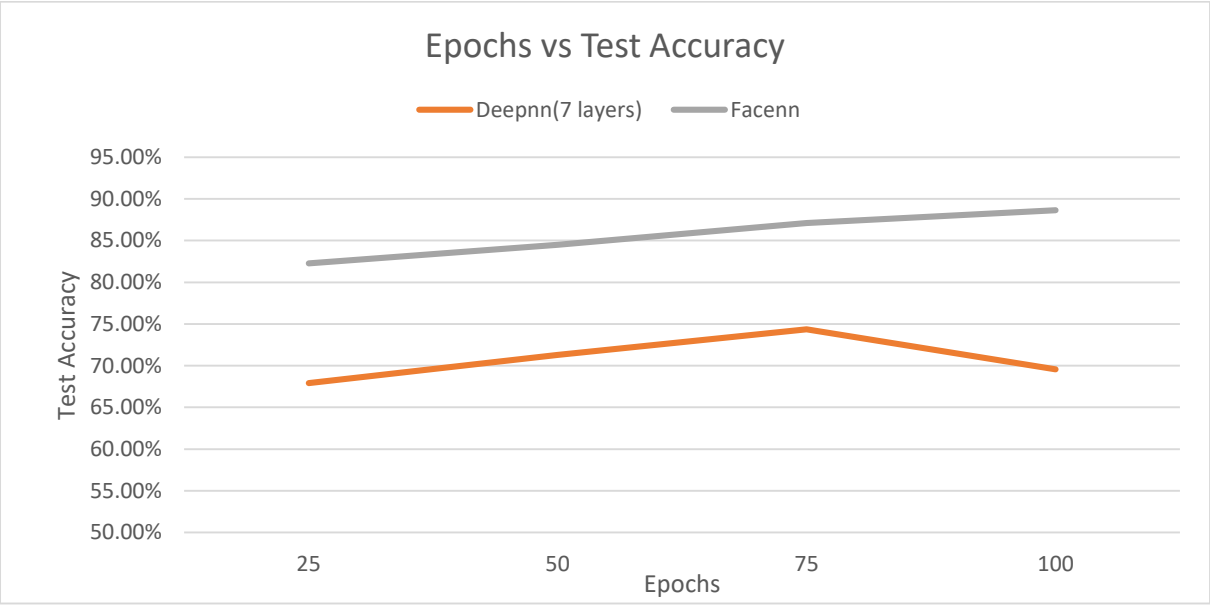



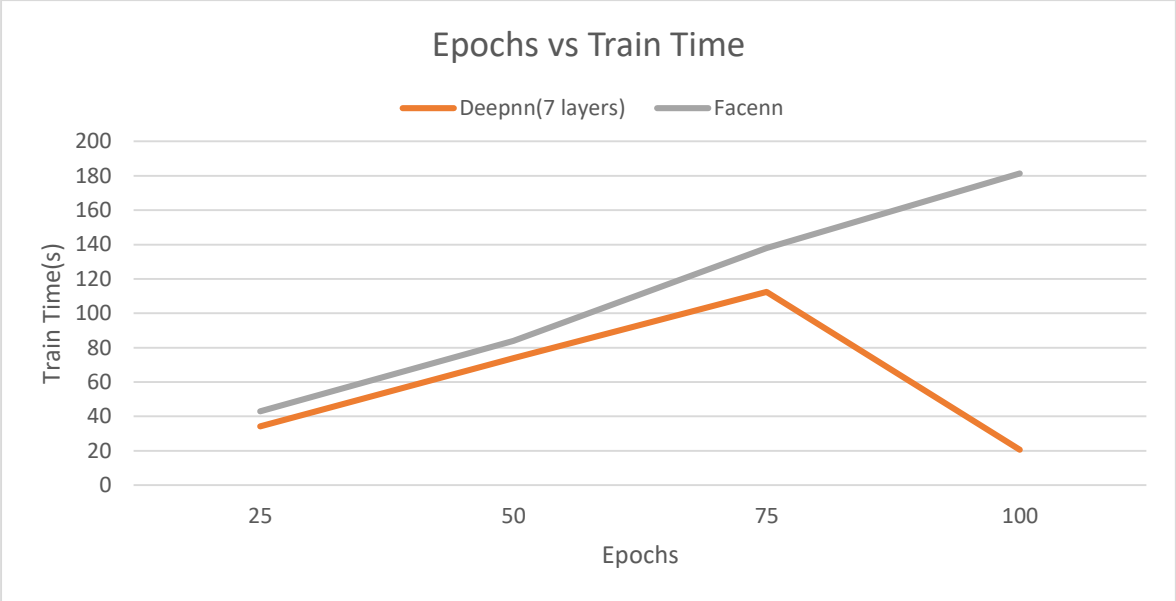
2. Single Hidden Layer Neural Network VS 5 Hidden Layer Deep Neural Network



3. Single Hidden Layer Neural Network VS 7 Hidden Layer Deep Neural Network

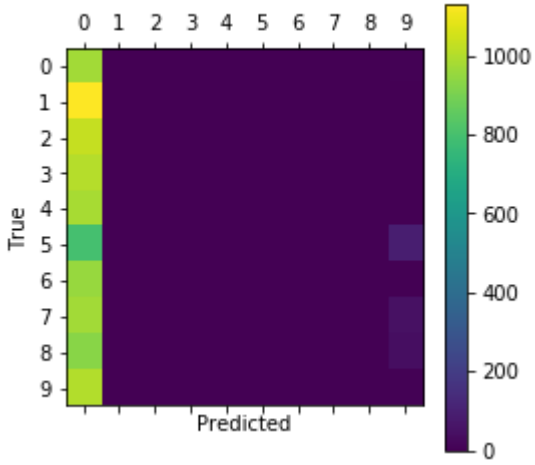


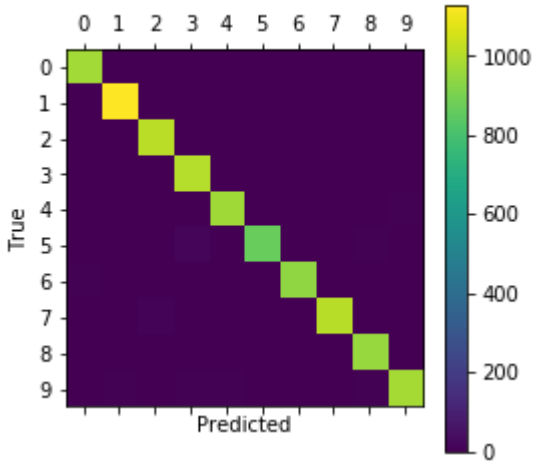
**Extra Credit:**

Results from Convolution Neural Network: (With Confusion Matrix)

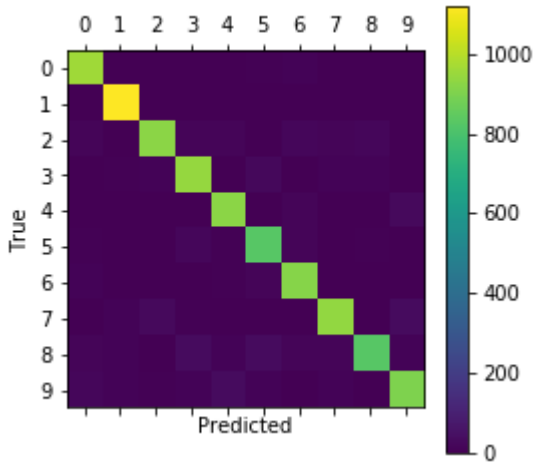
num_iterations=1
Time usage: 0:00:00
Accuracy on Test-Set: 11.4% (1138 / 10000)




num_iterations=99
Time usage: 0:00:08
Accuracy on Test-Set: 65.8% (6581 / 10000)



num_iterations=900
Time usage: 0:01:35
Accuracy on Test-Set: 93.8% (9385 / 10000)




num_iterations=9000
Time usage: 0:12:02
Accuracy on Test-Set: 98.7% (9874 / 10000)