

PA 1- Handwritten Digits Classification Report

CSE 574: Introduction to Machine Learning

Team 7:

Firnaz Lutzian Adiansyah - firmazlu@buffalo.edu

Sumedha Salil Nashte - sumedhas@buffalo.edu

Anushree Naveen Gupta - agupta38@buffalo.edu

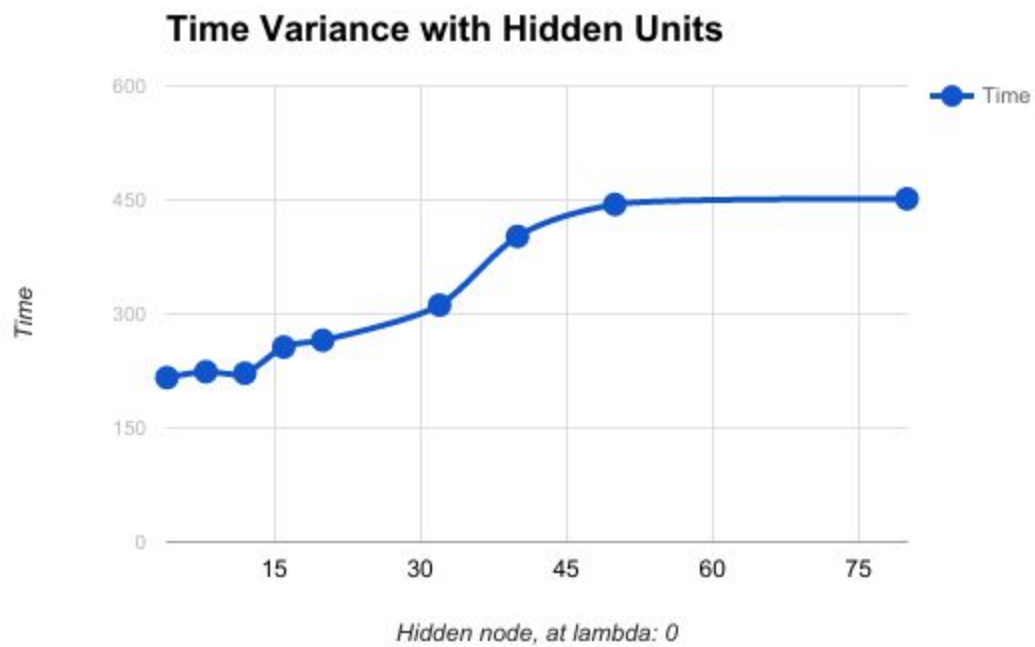
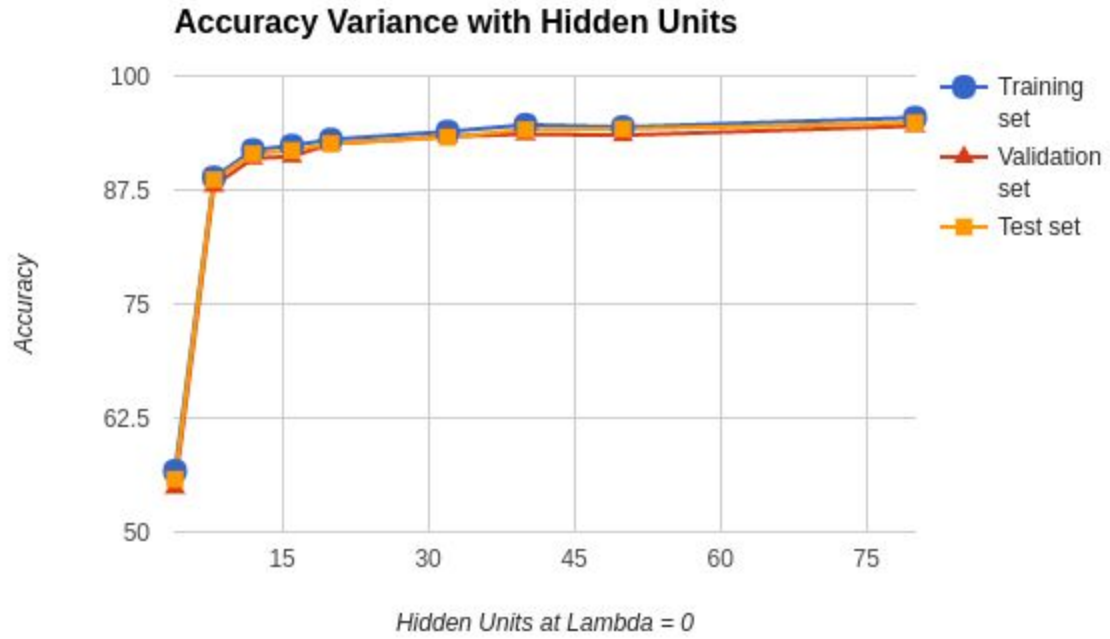
UNIVERSITY AT BUFFALO

PART 1: Explanation of how to choose the hyper-parameters for Neural Network using nnScript.py

Experiment 1: Finding the optimal number of Hidden Nodes by keeping Lambda constant

- Lambdaval: 0
- Hidden Units: [4, 8, 12, 16, 20, 32, 40, 50, 80]

Hidden Units	Lambda	Training Set	Validation Set	Test Set	Time (secs)
4	0	56.688	54.85	55.78	216.280596
8	0	88.874	87.95	88.63	224.0690348
12	0	91.874	90.96	91.46	222.0112495
16	0	92.356	91.17	91.82	256.4870369
20	0	93.036	92.55	92.57	265.3556936
32	0	93.9	93.37	93.27	311.6394339
40	0	94.646	93.61	94.11	402.0598094
50	0	94.398	93.5	94.2	444.1534941
80	0	95.428	94.51	94.86	451.5657842



Inference:

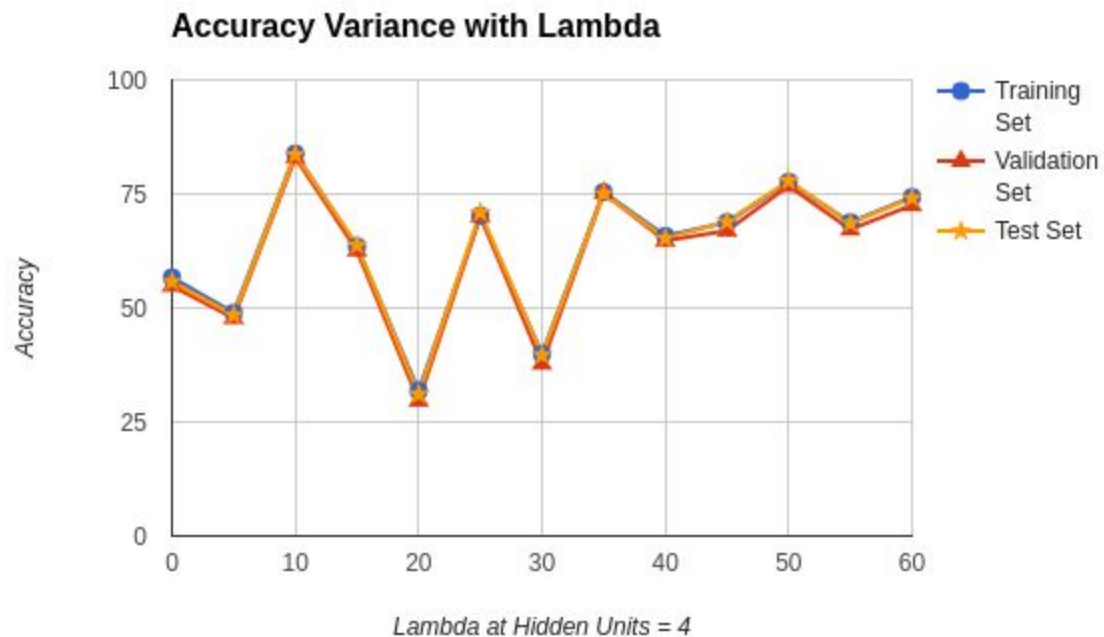
- Increasing number of hidden nodes, significantly increases the accuracy of the system.
- As a drawback, time taken for computation also increases significantly with the number of hidden nodes

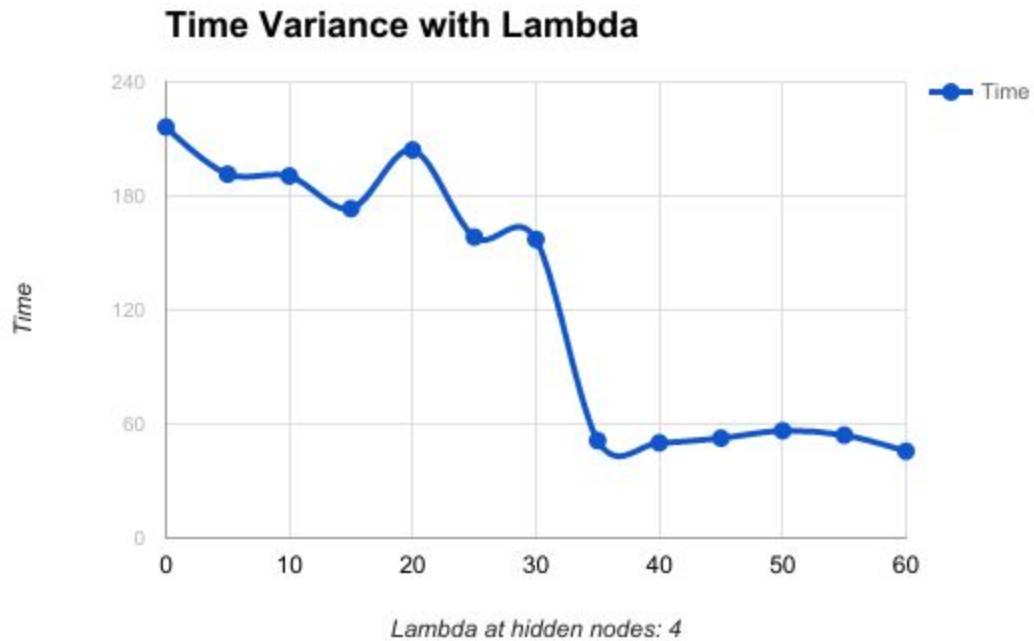
Thus if accuracy is more important than the time taken for computations we should choose higher number of hidden nodes.

Experiment 2: Finding the optimal Lambda by keeping Hidden Units constant

- Lambaval: [0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60]
- Hidden Units: 4

Hidden Units	Lambda	Training Set	Validation Set	Test Set	Time (secs)
4	0	56.688	54.85	55.78	216.280596
4	5	48.926	47.78	48.44	191.4631741
4	10	83.846	83.02	83.63	190.5121696
4	15	63.454	62.62	63.78	173.362231
4	20	31.858	29.67	31.08	204.1811707
4	25	70.246	70.16	70.93	158.3313904
4	30	39.938	37.78	39.67	157.1357303
4	35	75.386	75.19	75.1	51.26734829
4	40	65.796	64.75	65.31	50.07892132
4	45	68.768	66.95	68.92	52.52586603
4	50	77.628	76.76	77.89	56.42496443
4	55	68.776	67.22	68.59	54.08878446
4	60	74.324	72.53	74.12	45.6342442





Inference:

- In above experiment we also compared the accuracies for different hidden units by keeping different values of λ constants. We observed that by keeping higher values of λ as constant the accuracies for the entire dataset went down.
- Higher accuracies are achieved for the lower values of λ due to under-fitting. This happens as higher regularization parameter gives more importance to the weights at the expense of the error function while the opposite follows for lower values of regularization parameter.
- With increasing regularization parameter the time taken to learn the system decreases.

Conclusion:

Considering all the above observation we have chosen the number of Hidden Units and Lambda value considering a trade-off between the accuracy as well time.

Hyper-Parameters	Default Values	Optimal Values
Regularization Parameter	0	20
Hidden Units	50	80
Accuracy	Training Set - 94.40% Validation Set - 93.50% Test Set - 94.20%	Training Set - 95.258% Validation Set - 94.94% Test Set - 94.89%
Time	444.153 seconds	547.629 seconds

Note:

We have implemented all the functions as per the Programming Assignment 1 pdf and run all scripts on Metallica Server, which has resulted in long computation times.

PART 2: Compare the results of deep neural network and neural network with one hidden layer on the CelebA data set

facennScript.py:

We implemented all the three functions from nnScript.py to facennScript.py and ran the program for CelebA data set. Below are the observations for default values of lambda and hidden nodes as well as optimal values derived for nnScript.py.

Hyper-Parameters	Default Values	Optimal Values
Regularization Parameter	10	20
Hidden Units	256	80
Accuracy	Training Set - 84.20% Validation Set - 83.00% Test Set - 84.41%	Training Set -85.36 % Validation Set - 84.13 % Test Set - 85.16 %
Time	238.56 seconds	293.687 seconds

deepnnScript.py:

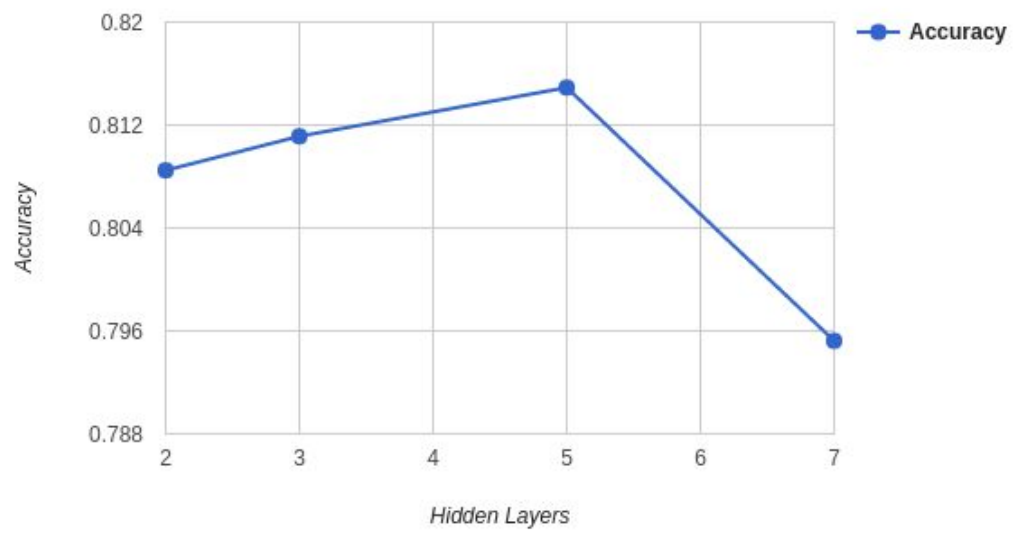
We varied the number of hidden layers in the deepnnScript.py and noted the observations as below.

Layers	Accuracy	Time
2 (by default)	0.808478	74.56 seconds
3	0.811128	90.73 seconds
5	0.814913	126.98 seconds
7	0.795231	267.78 seconds

Inference:

On comparing the above results we can say that neural network with one hidden layer and our optimized values of hyper-parameters perform with better accuracy i.e., 85.16% on the Celeb dataset as compared to deep neural network which gives a maximum accuracy of 81.49% with 5 hidden layers.

Accuracy Variance with Hidden Layers



Time Variance with Hidden Layers

