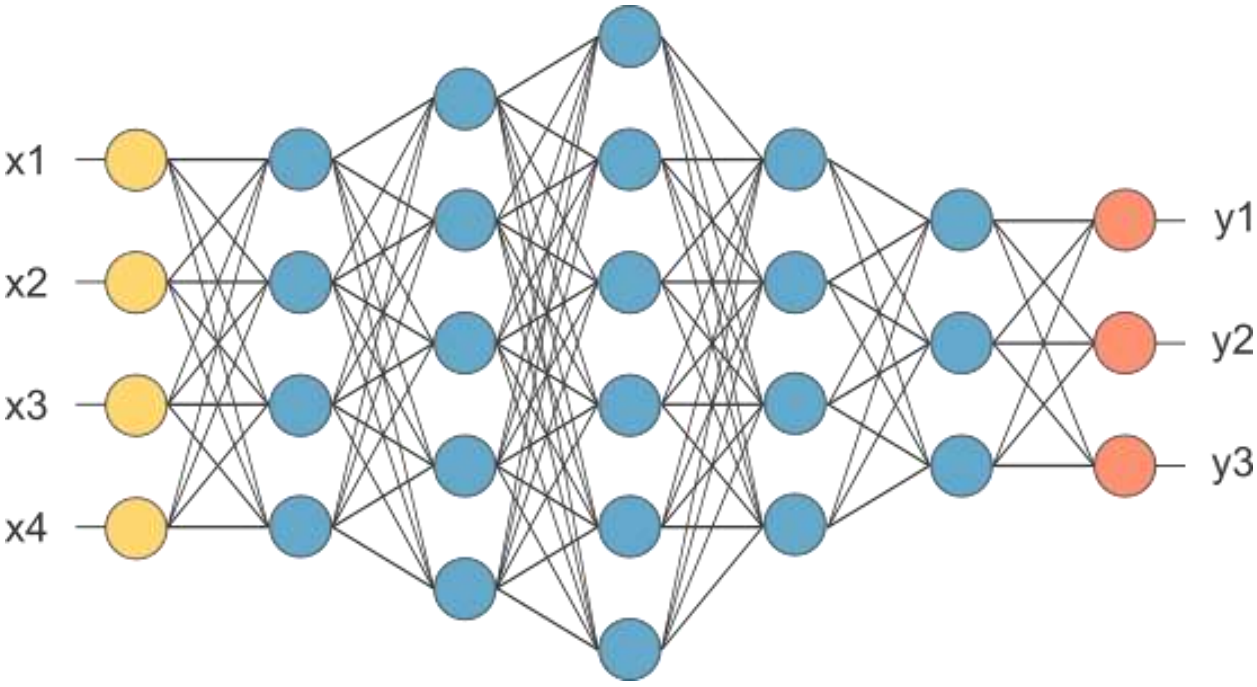
CSE574 Introduction to Machine Learning
Programming Assignment 2
Handwritten Digits Classification
Group 150

Ashutosh Ahmad Alexandar Harish Ganesan

In this assignment, we design and develop a Neural Network with one hidden layer as well as a MultiLayer Perceptron Neural Network and check the performance using handwritten digit data which we have obtained from the MNIST data set. We then use a multi-layer neural network on a more complex Celeb face dataset by making use of TensorFlow.

In this report, we will be discussing the impact of feature selection (removing unnecessary data), the selection of number of hidden nodes, optimal regularization factor (lambda), performance with multiple hidden layers, etc.



Example of a Multi-Layer Perceptron Neural Network

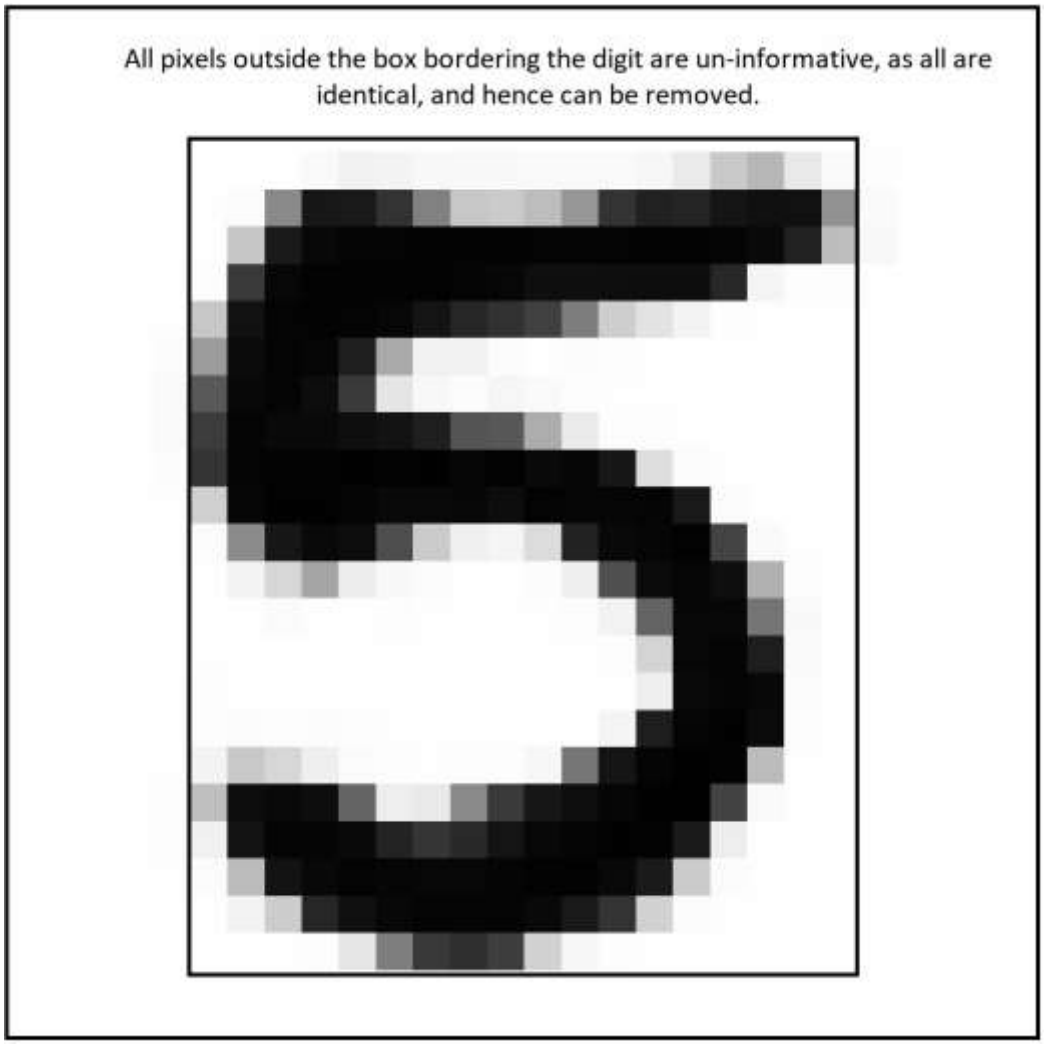
Feature Selection :

This is one of the first steps we need to perform. We need to decide what pixels contain useful information, and which pixels contain redundant information.

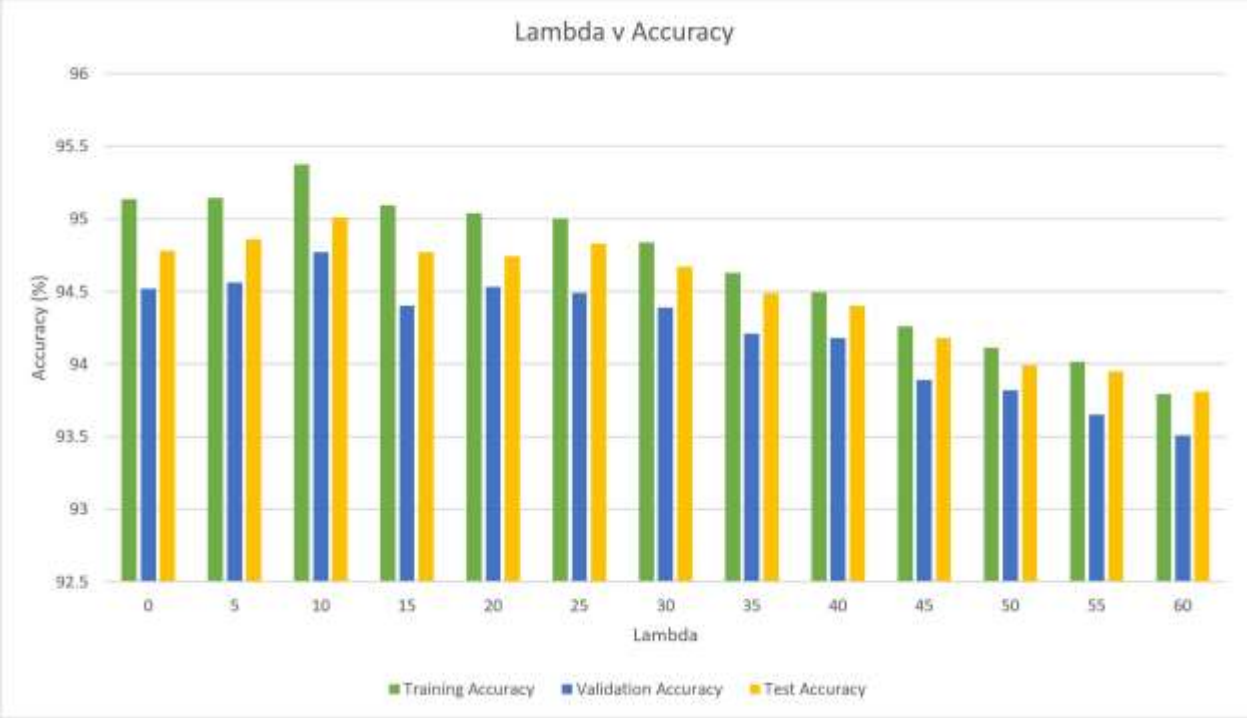
To do this, we see which features are the exact same as the background colour on the edges, which contain the redundant information. We then delete these pixels/features from the dataset.

These are the removed values from the dataset :
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 52, 53, 54, 55, 56 , 57, 59, 82, 83, 84, 85, 111, 112, 140, 141, 168, 476, 560, 644, 645, 671, 672, 673, 699, 700, 701, 727, 728, 729, 730, 754, 755, 756, 757, 758, 759, 780, 781, 782, 783]

Selected features are all indices minus the above indices, and it is provided in the params.pickle file.



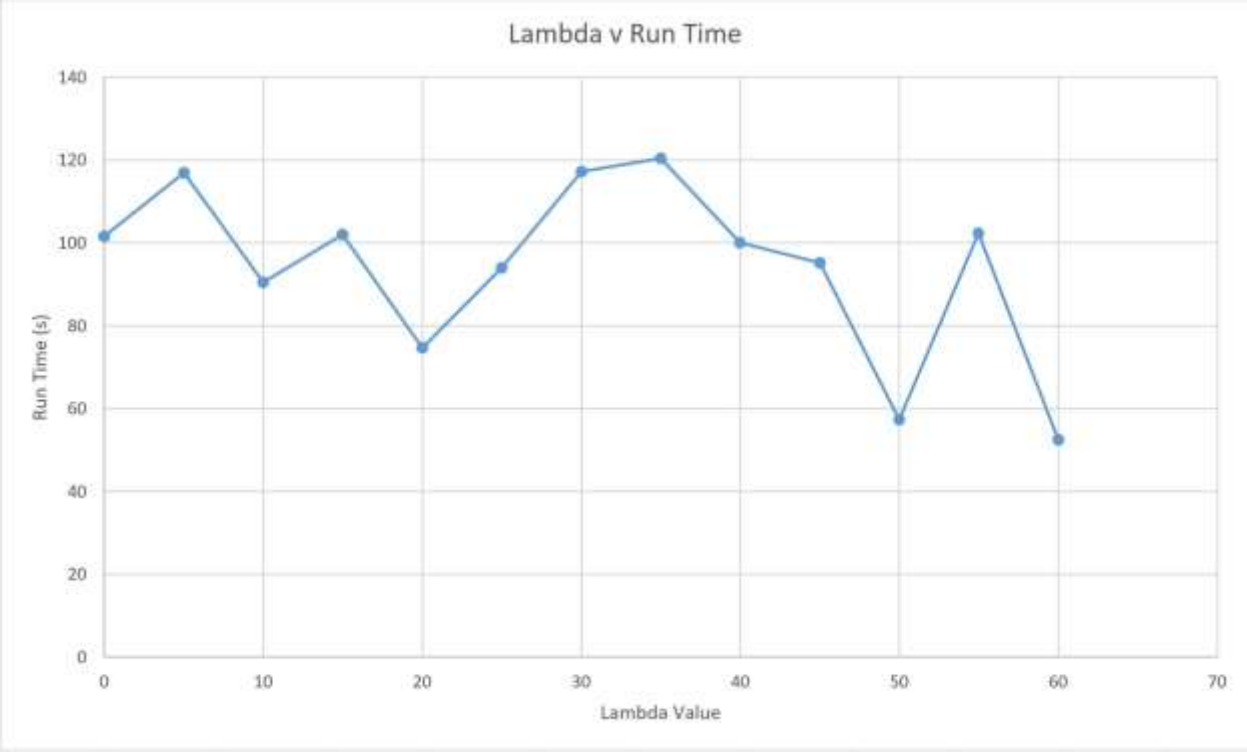
Relationship between Regularization Factor (λ) and Classification Accuracy



As we can see from the bar-chart shown above, we can clearly see that as the lambda increases from 0 to 5 to 10, the training, validation and testing accuracy also increases, and it is **maximum at lambda = 10**. From there on, as we increase the value of lambda, the accuracies keep decreasing, and this is because the values of the weights in the neural network are being changed by too large a value (for larger lambdas) and hence we get lower accuracies.

Now that we have found the optimal value for the regularization parameter, we must use this value of lambda = 10 for further experiments.

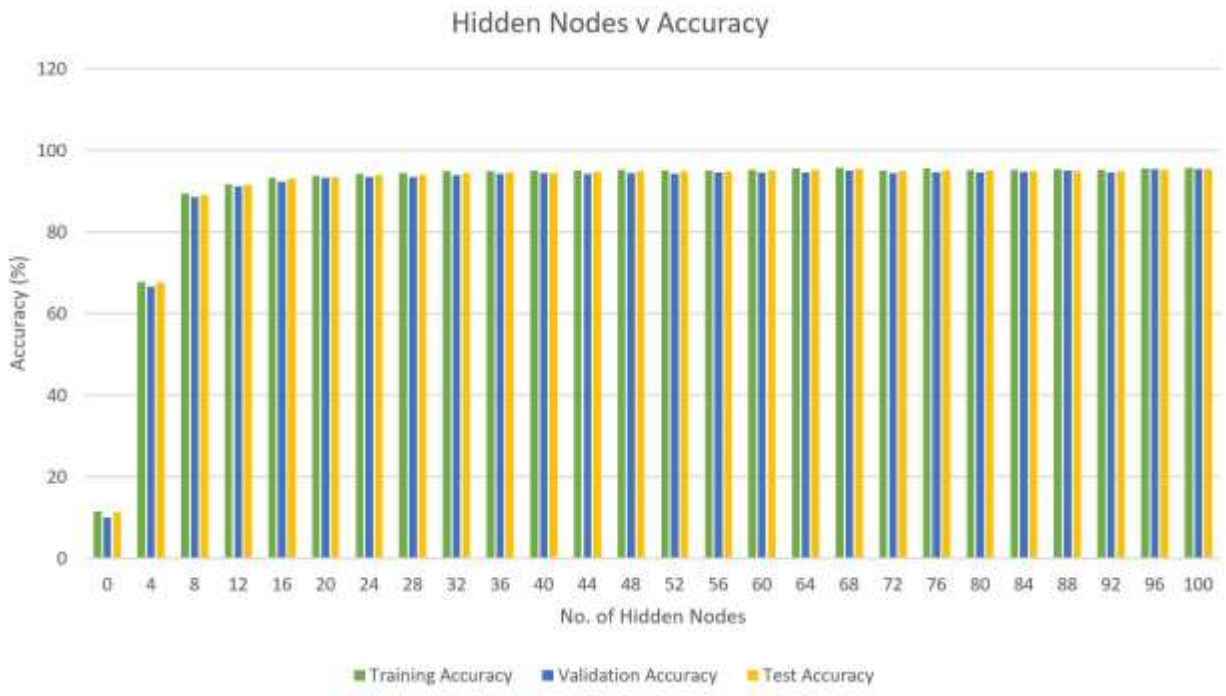
Relationship between Regularization Factor (λ) and Run-time



As shown in the graph above, the minimal times are present for larger values of lambda such as 50, 60. But if we take these higher lambda values into consideration, the accuracy drops sharply.

The fourth lowest run-time is when lambda = 10, and this is our optimal value in terms of accuracy, so we can decide furthermore that this is indeed our optimal lambda value.

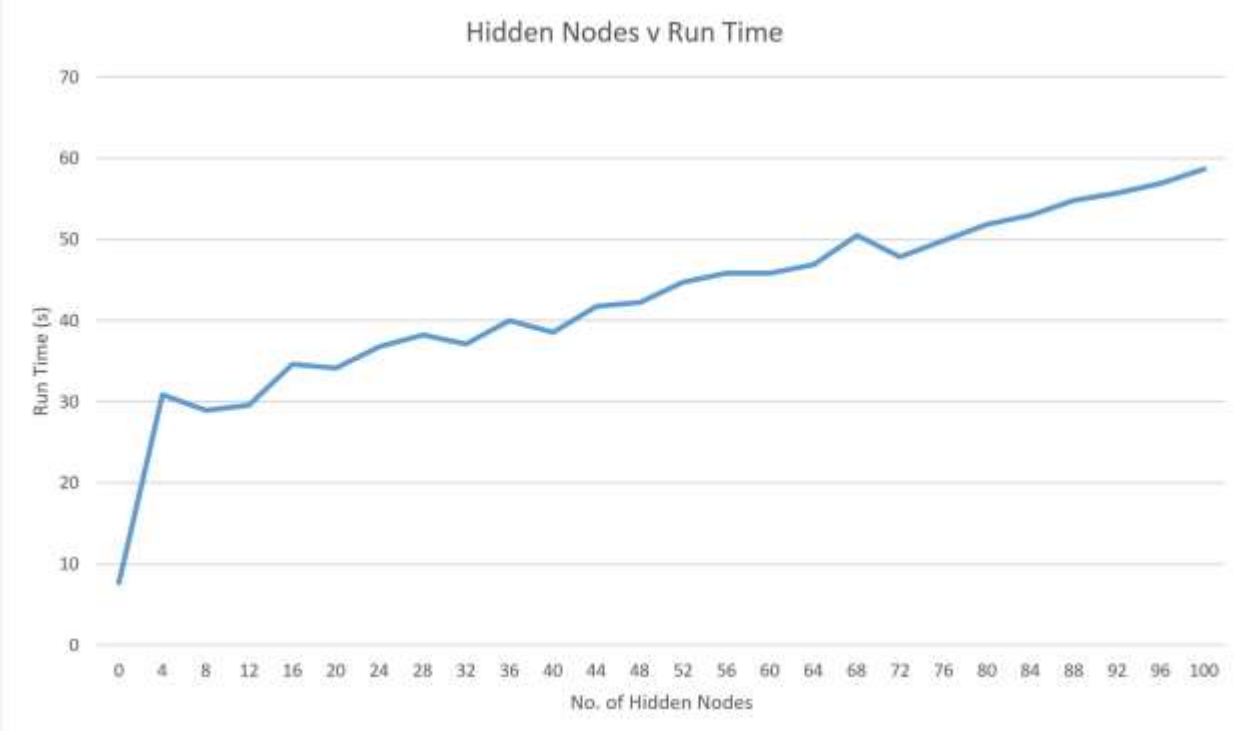
Relationship between No. of Hidden Nodes and Classification Accuracy



In this experiment, we fix the lambda value to 10 (optimal value).

As we can see from the above graph, all 3 accuracies for training, testing and validation increases as we increase the number of nodes in the hidden layer. In the beginning, the accuracies are low for smaller values such as 4 and 8, but then it increases by miniscule amounts thereon. This is because when we have more weight values, we can capture more variation in the input data. **The maximum accuracy achieved is in the case of no. of hidden nodes as 100 = 95.33% .**

Relationship between No. of Hidden Nodes and Run Time



In this experiment, we fix the lambda value to 10 (optimal value).

From the above graph, we can observe that as we increase the no. of hidden nodes, the run time also increases. This is intuitive, because with the addition of mode hidden nodes, the number of computations will increase and naturally the run time too. **The highest run-time is when no. of hidden nodes = 100.**