CSE574 Assignment 2

PA Group 11
Deepti Bharadwaj(50363309)
Kaumudi Moholkar (50388592)
Alexander Ma (37136560)

April 11, 2021

Report 1:

1. Run the evaluation of the implemented neural network in the notebook - PA2-Part2.ipynb and report the training and test accuracy and the run time.

For the hidden value (M) as 50 and lambda as 0 the training time and accuracies for both training set and testing set are as below:
i) Training completed : 6.81seconds.
ii) Training set Accuracy: 64.22%
iii) Test set Accuracy : 61.65%

2. Compare the performance when the number of hidden layer units (M) is increased from 10 to 100, in increments of 10. Plot the training and test accuracies and training time, as a function of M. Make your observations and state the optimal value of M that you would finally choose, along with the reason.

From the trials below (Figure 1.2.1), we selected M=100 to be our optimal M. When M=100 the Test Set Accuracy is the highest out of all the trials. To get the optimal value of M, we need to consider the higher test accuracy along with good training accuracy. The difference between the training and test accuracy is also not large, so the model is potentially a good fit as compared to other values.

M	Time	Training Set Accuracy	Test Set Accuracy
10	6.36	44.18	42.77
20	6.43	54.29	52.62
30	5.73	63.48	60.94
40	6.97	66.58	63.65
50	6.81	64.22	61.65
60	7.04	67.88	64.54
70	8.31	69.42	65.72
80	9.17	67.95	65.01
90	9.33	70.1	65.98
100	9.3	71.43	66.98
max	9.33	71.43	66.98
min	5.73	44.18	42.77
average	7.545	63.953	60.986

Figure 1.2.1 - Data from M = 10 -> 100

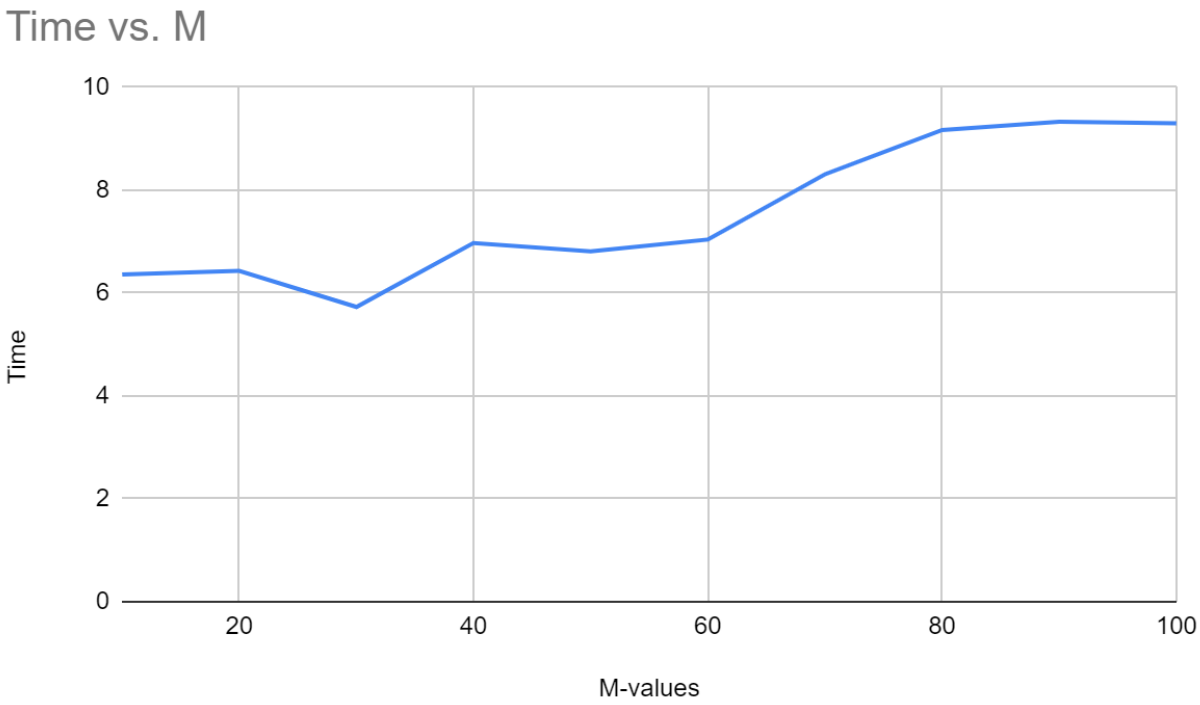


Figure 1.2.2 - Time Vs M (number of hidden layers) based of data from Figure 1.2.1

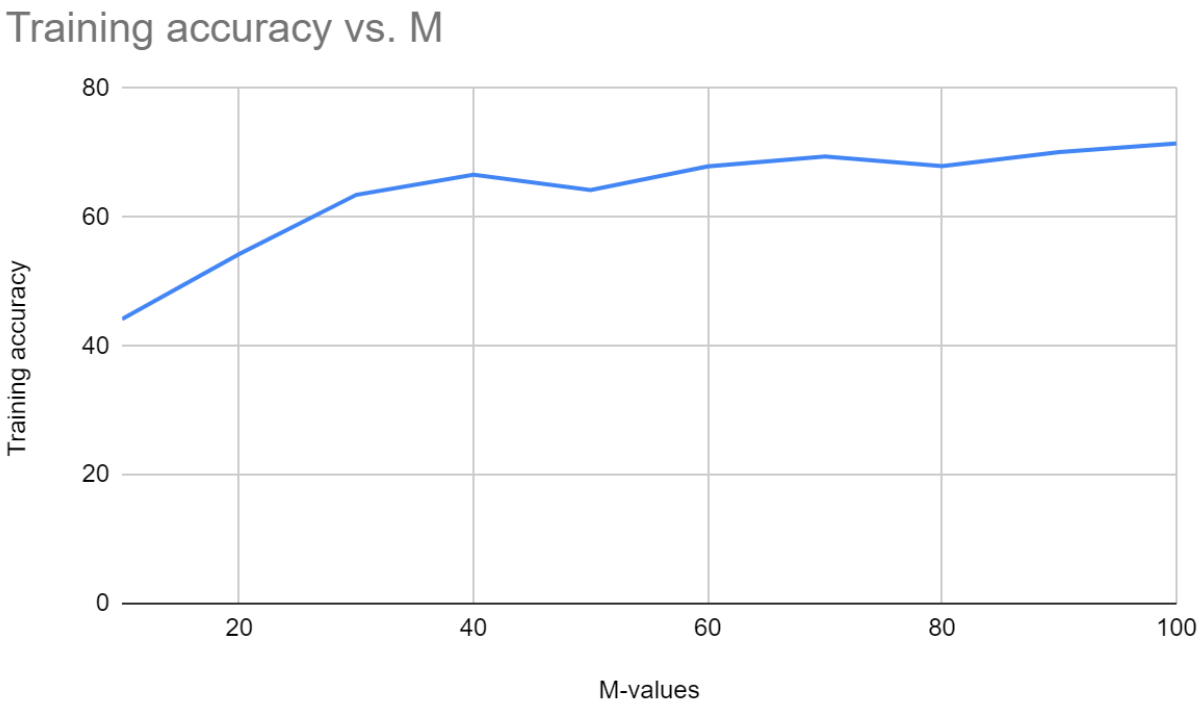


Figure 1.2.3 - Training Set Vs M (number of hidden layers) based of data from Figure 1.2.1

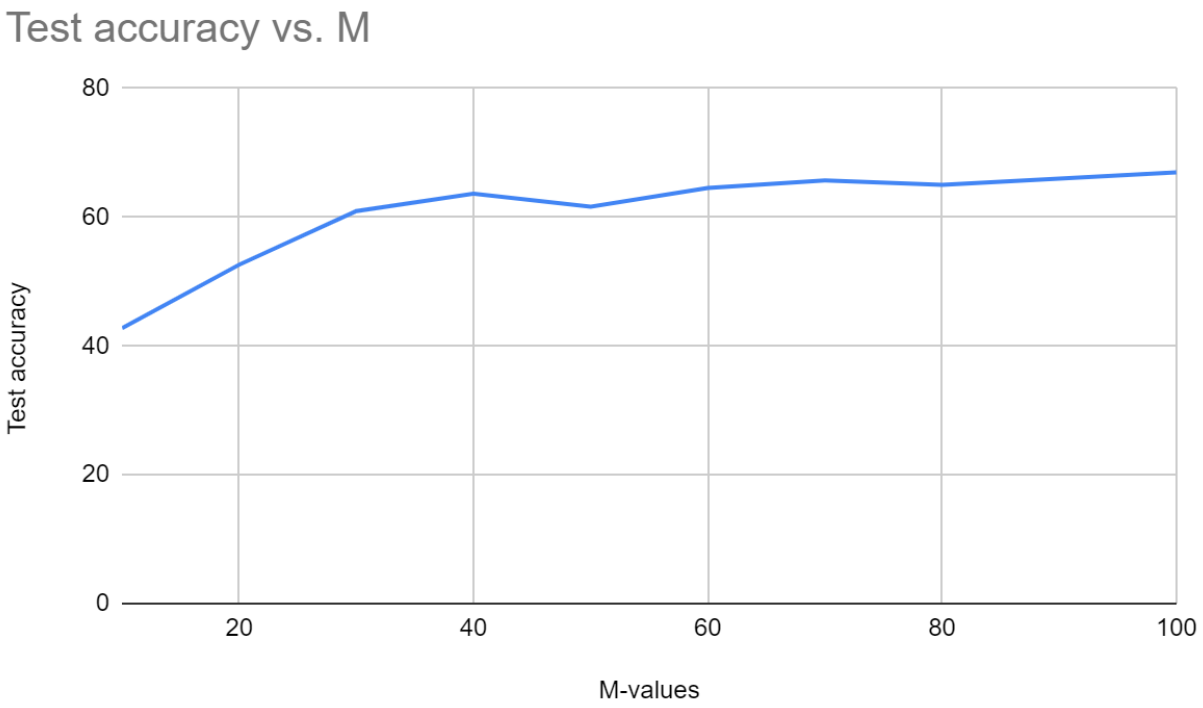


Figure 1.2.4 - Test Set Vs M (number of hidden layers) based of data from Figure 1.2.1

3. For the optimal setting of M found above, rerun your analysis by modifying λ from 0 to 20, in steps of 2. Again, plot the training and test accuracies and the training time as a function of λ and make your observations. Which value of λ is optimal and why?

After iterating λ with values from 0 to 20, we found λ = 8 to be our most optimal value. From the results below it is observed that when λ=8 ,since the test set accuracy and training set accuracy is the highest compared to the other values of λ. We stayed consistent from the optimal M by prioritizing the maximum test accuracy. A better test accuracy is a clear indication of which model has a better performance.

The graphs and the table below show the results of the model.

M	Lambda	Time	Training Set Accuracy	Test Set Accuracy

100	0	9.81	70.92	65.77
100	2	10.22	70.68	66.14
100	4	9.3	70.76	66
100	6	9.92	70.45	66.69
100	8	9.57	71.86	67.41
100	10	9.73	69.83	66.13
100	12	9.36	70.86	66.66
100	14	10.3	71.36	66.8
100	16	9.41	70.57	66.34
100	18	9.24	70.87	66.74
100	20	9.73	71.51	66.54
average		9.69	70.87909091	66.47454545
min		9.24	69.83	65.77
max		10.3	71.86	67.41

Figure 1.3.1 - Data from λ = 0 -> 20 with M=100

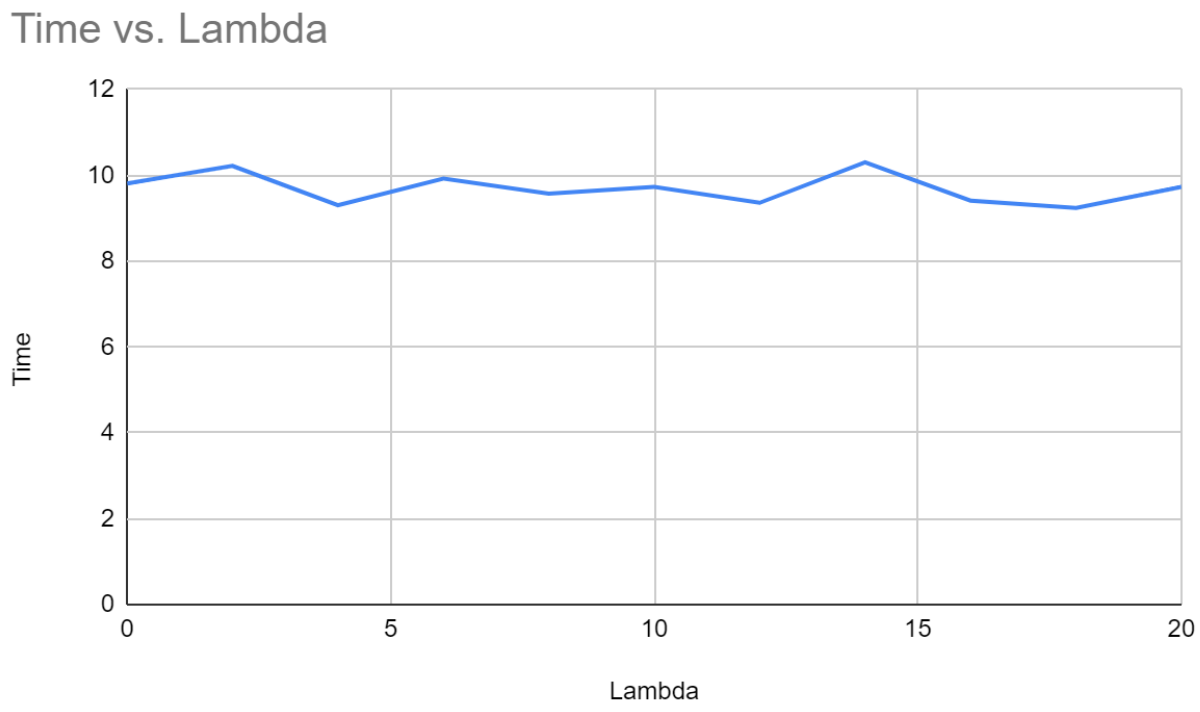


Figure 1.3.2 - Time Vs λ with M=100

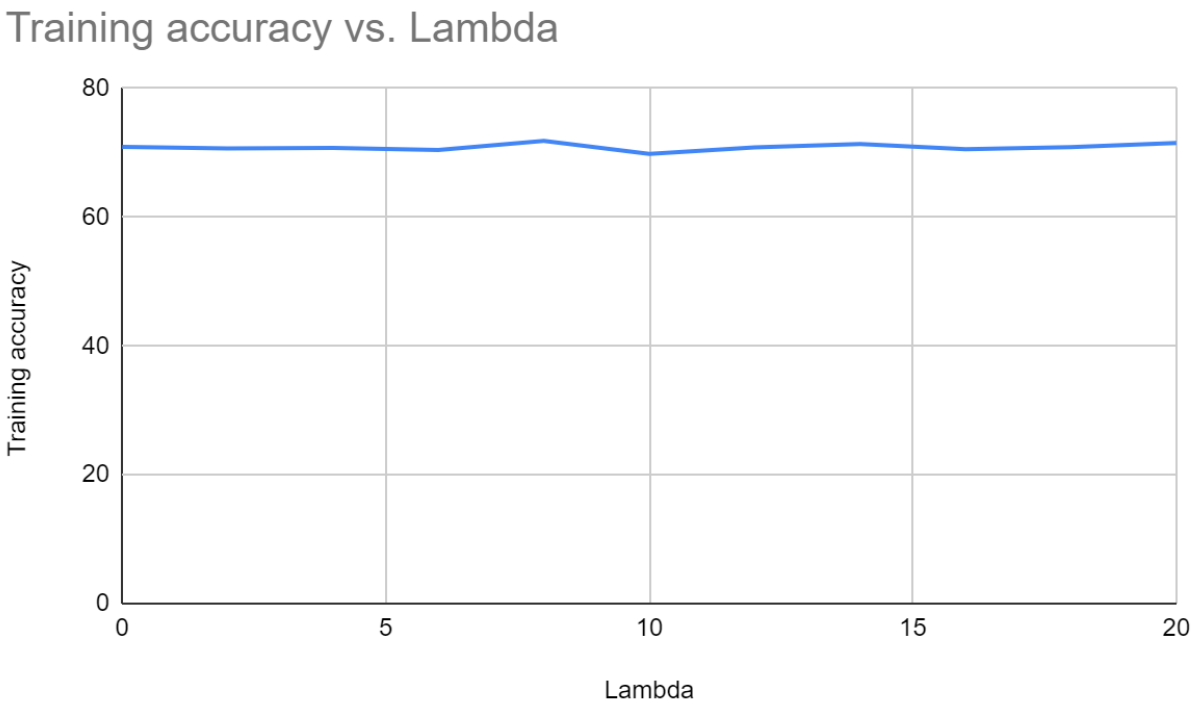


Figure 1.3.3 - Training Set Accuracy Vs λ with M=100

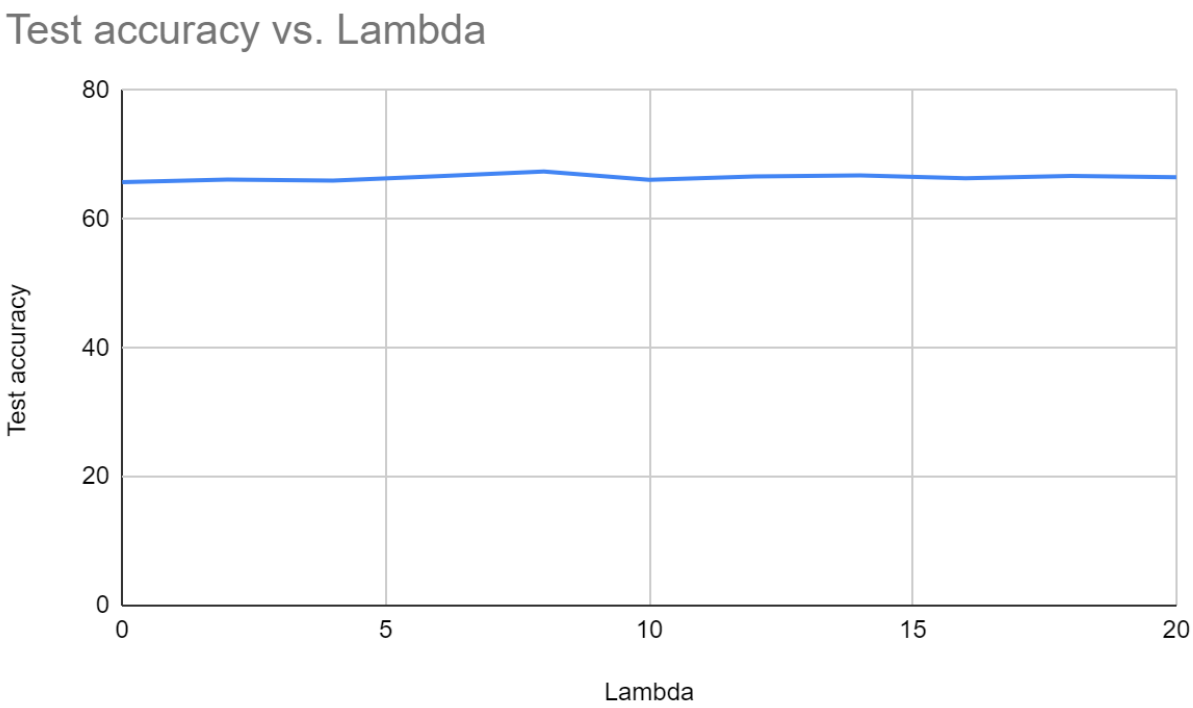
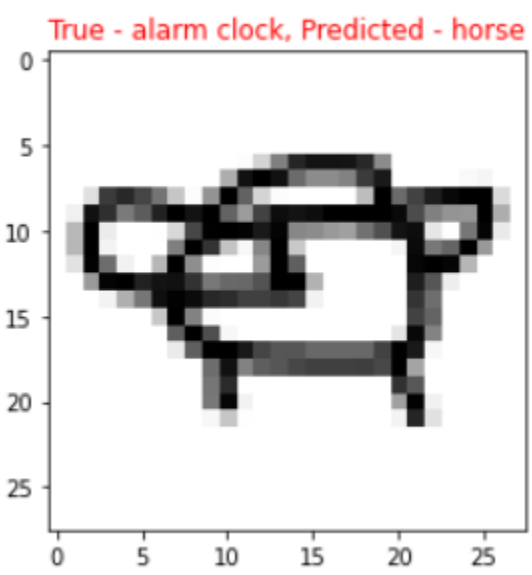
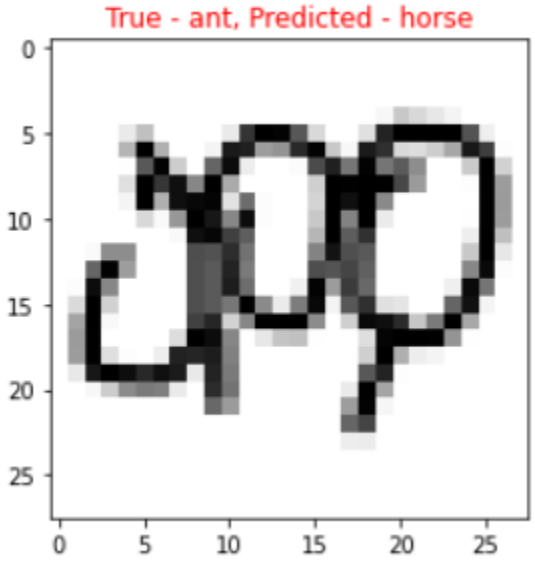


Figure 1.3.3 - Test Set Accuracy Vs λ with M=100

4. For the optimal settings for M and λ, study the performance of your model on the test data. What kind of objects does it make more mistakes on? Briefly discuss how the performance of your model can be improved further.

After running many trials using the optimal setting we found for M and λ (M = 100, λ = 8), we saw that the objects that have similar features get misclassified by the model.
From some of the misclassified images, we observed that these images had ambiguous features that even we had a hard time figuring out what the images could be. Examples below.





To overcome these misclassifications we need to reconsider how we go about training the model. Since the model doesn't understand some features properly, we can train more samples which can classify those features well. This is because complex features of this particular sample haven't been learnt by the model well.

Report 2.

1. Fixing the number of units in each hidden layer (M) to the optimal value found in Part II, run the evaluation of the implemented neural network in the notebook - PA2-Part3.ipynb for different numbers of hidden layers (L), from 1 to 5. Plot the training and test accuracies and training time, as a function of L. Make your observations and state the optimal value of L that you would finally choose, along with the reason.

Based on test accuracy and training accuracy, L=2 is the optimal value since it achieves the highest test accuracy. We selected our optimal value of L based on the highest test accuracy along with a good training accuracy A lower difference in accuracy between training set and test set would be more desirable to make sure the model is not overfitted.

M	L	Time	Training Set Accuracy	Test Set Accuracy	Difference of Accuracy (Training - Test)
100	1	78.55	76.65	73.82	2.83
100	2	78.01	79.81	76.08	3.73
100	3	88.1	79.84	76.05	3.79
100	4	97.22	78.29	74.9	3.39
100	5	100.55	78	74.81	3.19

Figure 2.1.1 - Data from L = 0 -> 20 with M=100

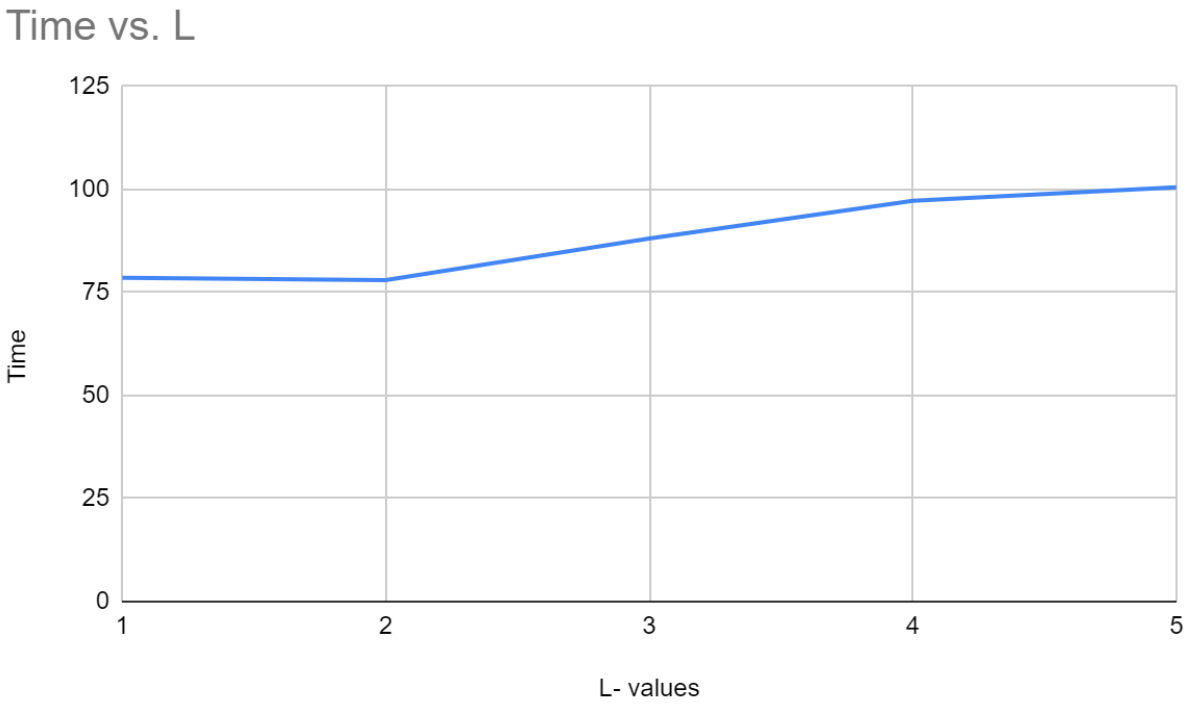


Figure 2.1.2 - Time vs L based with M=100

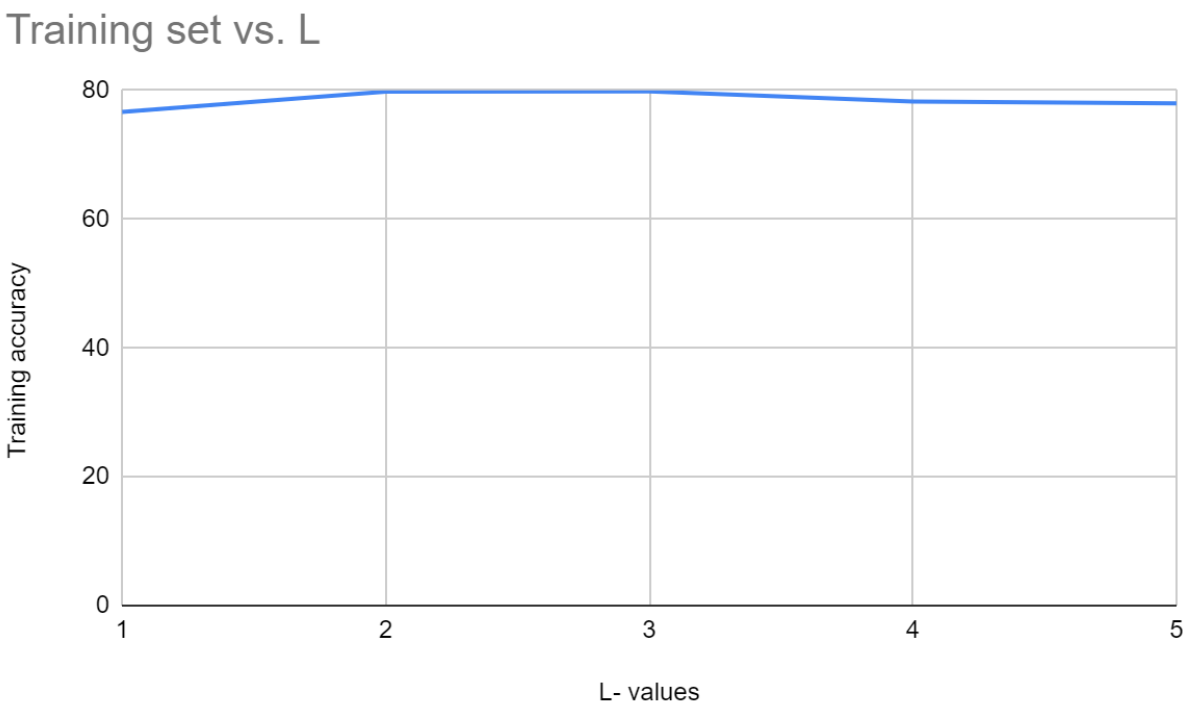


Figure 2.1.3 - Training Set Accuracy vs L with M=100

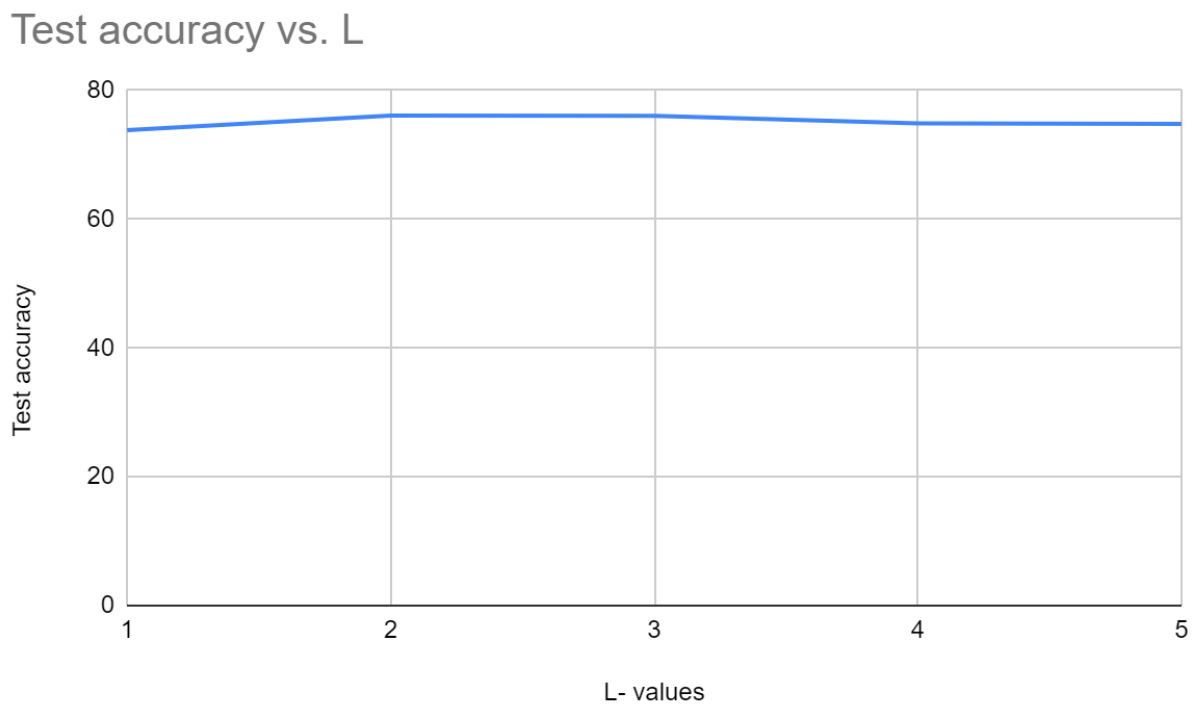


Figure 2.1.4 - Test Set Accuracy vs L with M=100

2. Using the optimal M and L from the previous part, compare the performance of the model (in terms of training and testing accuracies and the training time) for different choices of the activation function (try sigmoid, tanh, and relu). Report the best choice.

From the previous sections we selected optimal M = 100 and L = 2 to do separate trials for the different activation functions. From the data below we found that the **sigmoid** activation function is the best choice as it took less time to run and had better test accuracies when compared with the other two.

M	L	Activation	Time	Training Set Accuracy	Test Set Accuracy
100	2	sigmoid	78.91	79.81	76.07
100	2	tanh	84.01	79.82	75.59
100	2	relu	88.56	81.73	69.95

Figure 2.2.1 - Metrics from different activations