Group 35
Jae Hyeok Kwak
Sachin Tiwary
Manya Singh
CSE-574

PROBLEM 2: USING THE NEURAL NETWORK FOR IMAGE CLASSIFICATION

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.

```
Training completed in 16.45 seconds.
Training set Accuracy: 67.79%
Test set Accuracy    : 64.52%
```

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.

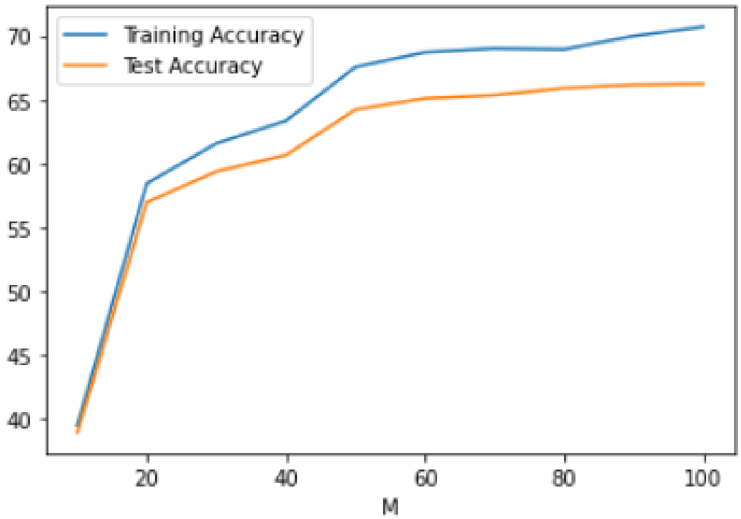


Figure 1 Training and Test Accuracies vs Hidden layer units (M)

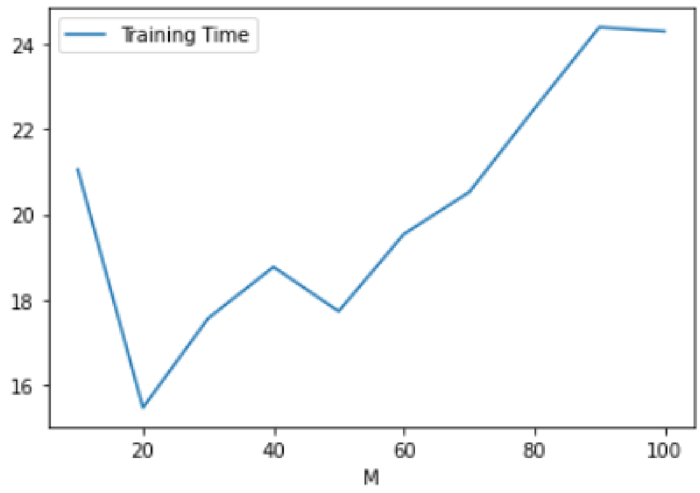


Figure 2 Training Time (in seconds) vs Hidden layer units (M)

Observations: Both training and test accuracies increase with increase in the number of hidden layer units (M). There is a sharp increase in both the accuracies initially (from M=10 to M=20). This increase in accuracy keeps getting lesser and lesser (especially for test data) as M approaches 100. The training time also increases with the increase in M.

Optimal Value: M=100

Reason: Accuracy for test data as well as for training data is highest at M=100. This is because the complexity of the model increases with increase in the number of hidden layer units which makes the model learn better for the given training set. Furthermore, the higher test accuracy at M=100 suggests that the complexity is not so much that it starts overfitting.

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?

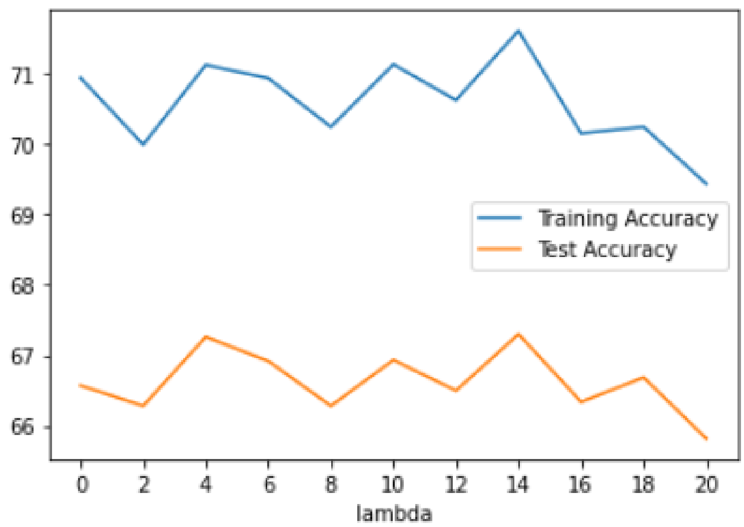


Figure 3 Training and Test accuracies vs Lambda

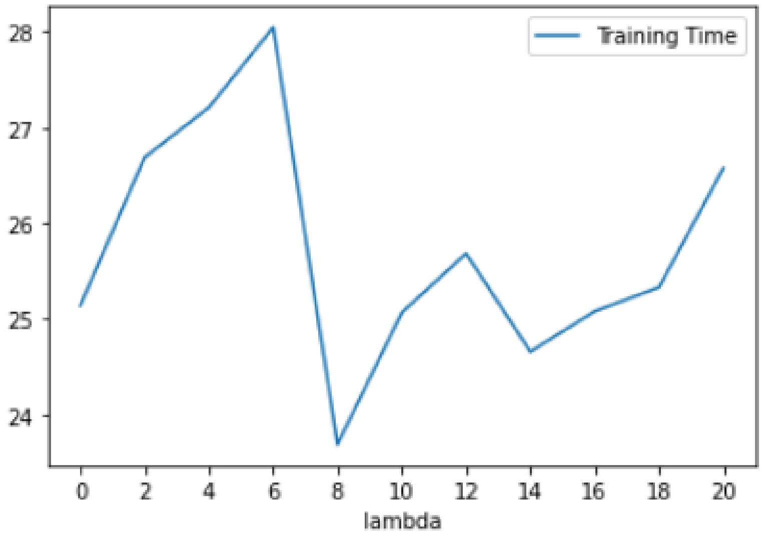


Figure 4 Training time (in sec) vs Lambda

Observations: The graphs were plotted by taking the optimal value of M (100) and varying λ between 0 and 20. The accuracy vs λ graph varies a lot with each trial (loading different training sets). The training accuracy takes values between 70 and 73 while test accuracy takes values between 66 and 68. The test accuracy is generally lower when λ is at its extremities and tend to take higher values somewhere around the middle. The training time vs λ graph is quite random and shows no particular trend.

Optimal Value: λ =14

Reason: We got the highest test accuracy (67.30%) at λ=14. This suggests that the model produces the best results with the test dataset when the regularisation constant is around this value and at this point the bias vs variance trade-off is optimal.

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.

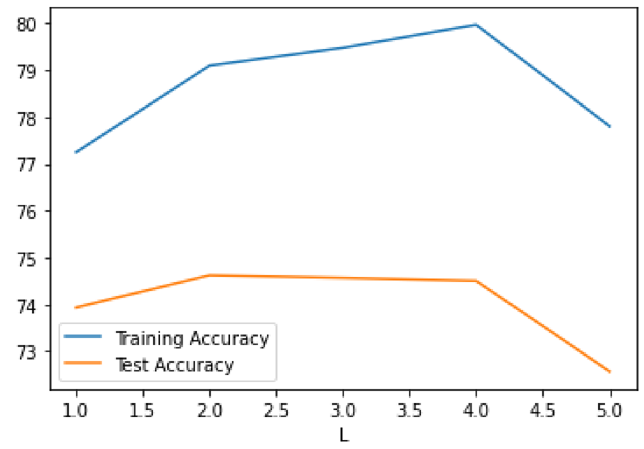
Taking M=100 and λ=14, the model was run 100 times out of which it made false predictions 29 times (71% accuracy). Most of the mistakes were made on objects having a linear shape like arm, banana, airplane and ant, while it mostly predicted correctly the objects having rounded shapes such as apple and basketball or objects having a peculiar feature like bed and axe.

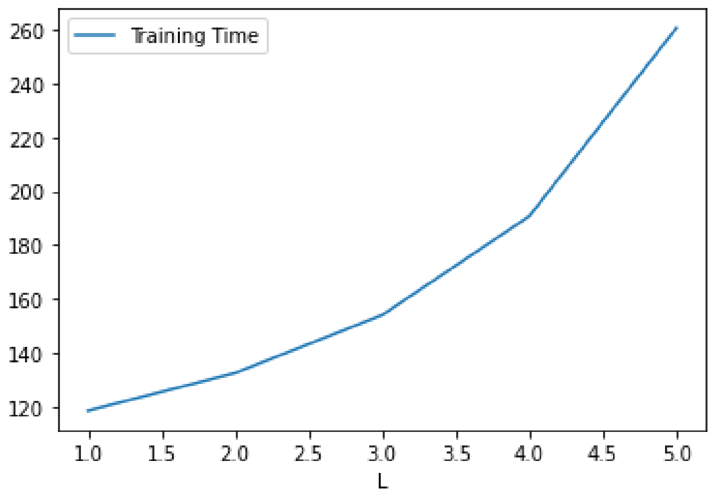
The performance can be further increased by increasing the number of hidden layers (L) in the model. We can get an optimal value for L just like M and λ to maximise the test accuracy. We can also try with different activation functions like tanh and check which one works the best for the given data. Also, since we use gradient descent method, trying with a different initialisation of weights can help converging to a different minima. Lastly, training with a larger data set can yield a better learned model and can increase the test accuracy.

PROBLEM 3: USING DEEP[ER] NEURAL NETWORKS FOR IMAGE CLASSIFICATION

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 number 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.





Observations: The graphs were plotted by taking the optimal value of M (100) and varying L between 1 and 5. The accuracy vs L graph varies a lot with each trial (loading different training sets). The training accuracy takes values between 77 and 80 while test accuracy takes values between 72.5 and 74.9. This may change with each iteration due to different data. The training time would however increase with L since we would calculate value for each layer over the iterations. The accuracy decreases when L is higher showing some signs of overfitting, it also depends on the complexity of the data we are trying to train our model with.

Optimal Value: L =2

Reason: We got the highest test accuracy (74.9%) at L=2. This suggests that the model produces the best results with the test dataset when the number of hidden layers is around this point with units in each layer is 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.**

For M=100 and L=2 below were the findings for different activation function:

- sigmoid:
  Training completed in 91.59 seconds.
  Training set Accuracy: 80.20%
  Testing set Accuracy: 76.34%
- relu:
  Training completed in 92.46 seconds.
  Training set Accuracy: 88.27%
  Testing set Accuracy: 73.63%
- tanh:
  Training completed in 94.70 seconds.
  Training set Accuracy: 79.73%
  Testing set Accuracy: 75.61%

As we can see the test accuracy for relu is the least and tanh and sigmoid are almost the same with sigmoid being slightly better, hence sigmoid activation function should be the best choice. The range for sigmoid is (0, 1) whereas for tanh is (-1, 1), so when we are trying to train our data the model would take more time to converge with tanh activation function.