**CSE574 PA2, Spring 2021**

**Group # 36**

**Names: Sai Lakshmi Navya Maddu**

**Sreya Dhar**

**Vaishnavi Vukku**

**Report 1:**

1. The results when Part2 is evaluated in default setting with number of units in hidden layers (M) = 50 and lambda = 0.5 are as follows:

Training completed in 18.95 seconds.

Training set Accuracy: 67.66%

Test set Accuracy: 64.34%

1. The number of units in hidden layers (M) is increased from 10 to 100, with increments of 10 keeping the same value of lambda = 0. We can observe that the training and test accuracies are increasing as the value of M is increasing until M reaches 100. Hence, **the optimal value of M is chosen as 100,** based on both train and test sets' performance (Fig. 1(a)). In Fig. 1(b) training time (in seconds) vs. No. of hidden units (M). At M = 100 train and test accuracies are 72.02% and 67.73%, respectively.

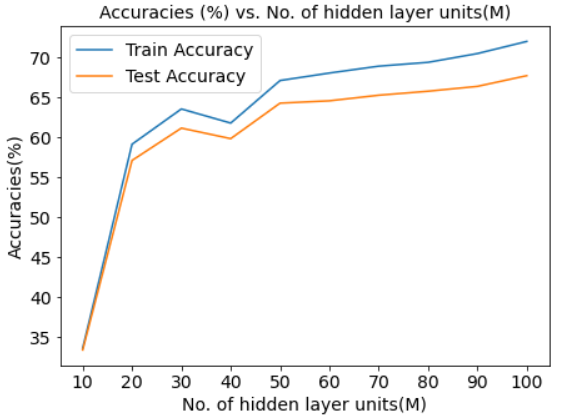
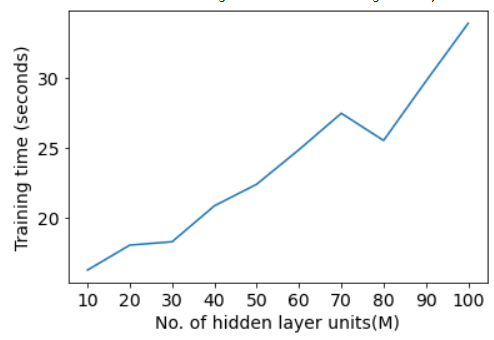
 (a)  (b)

Fig. 1 (a) Accuracies (%) vs. No. of hidden layer units (M), and (b) Training Time (Seconds) vs. No of hidden units (M)

1. Now, with the optimal value of M (units in hidden layer) = 100, the value of lambda (the regularization coefficient) is varied from 2 to 20, in steps of 2. The variation of train and test accuracies along with the training time are plotted in Fig. 2(a) and (b), respectively. Hence, **the optimal value of Regularization coefficient (lambda) is chosen as 20,** based on both train and test sets' performance (Fig. 2(a)). At lambda = 20 train and test accuracies are 72.33% and 64.84%, respectively.

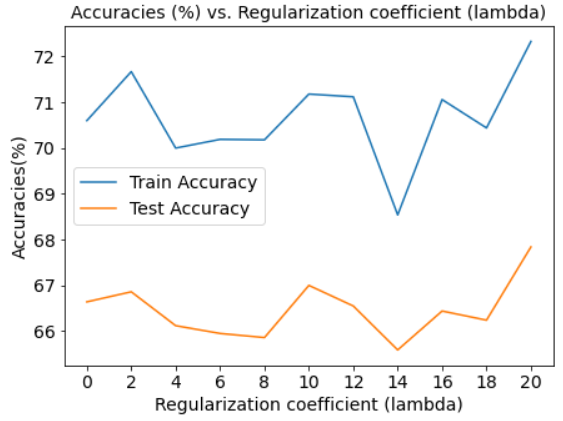
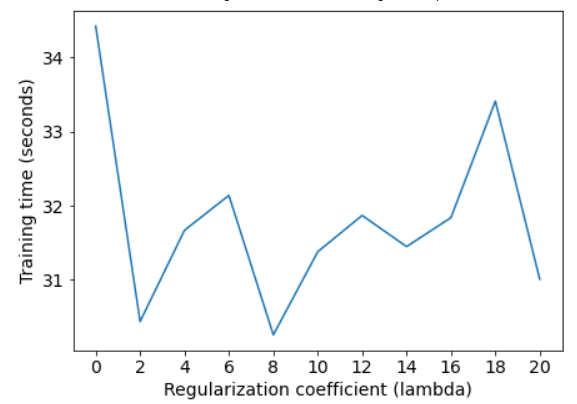
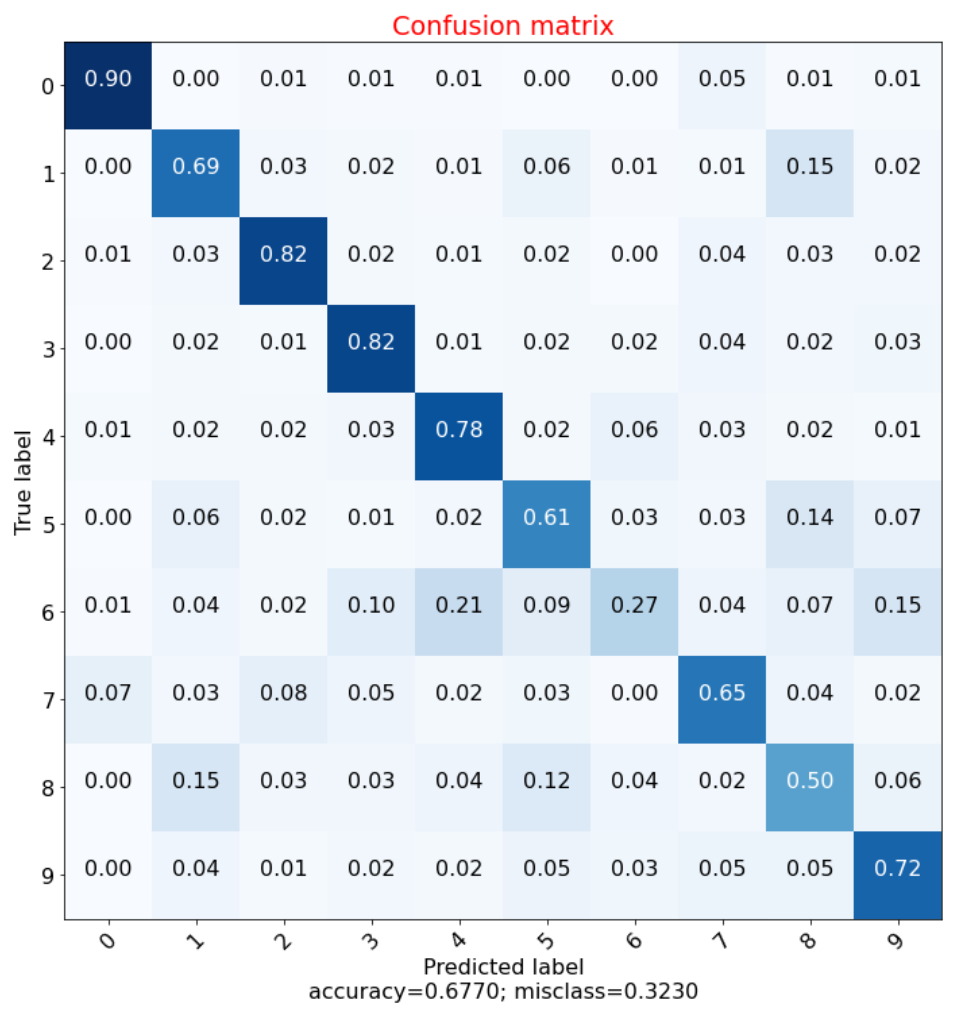
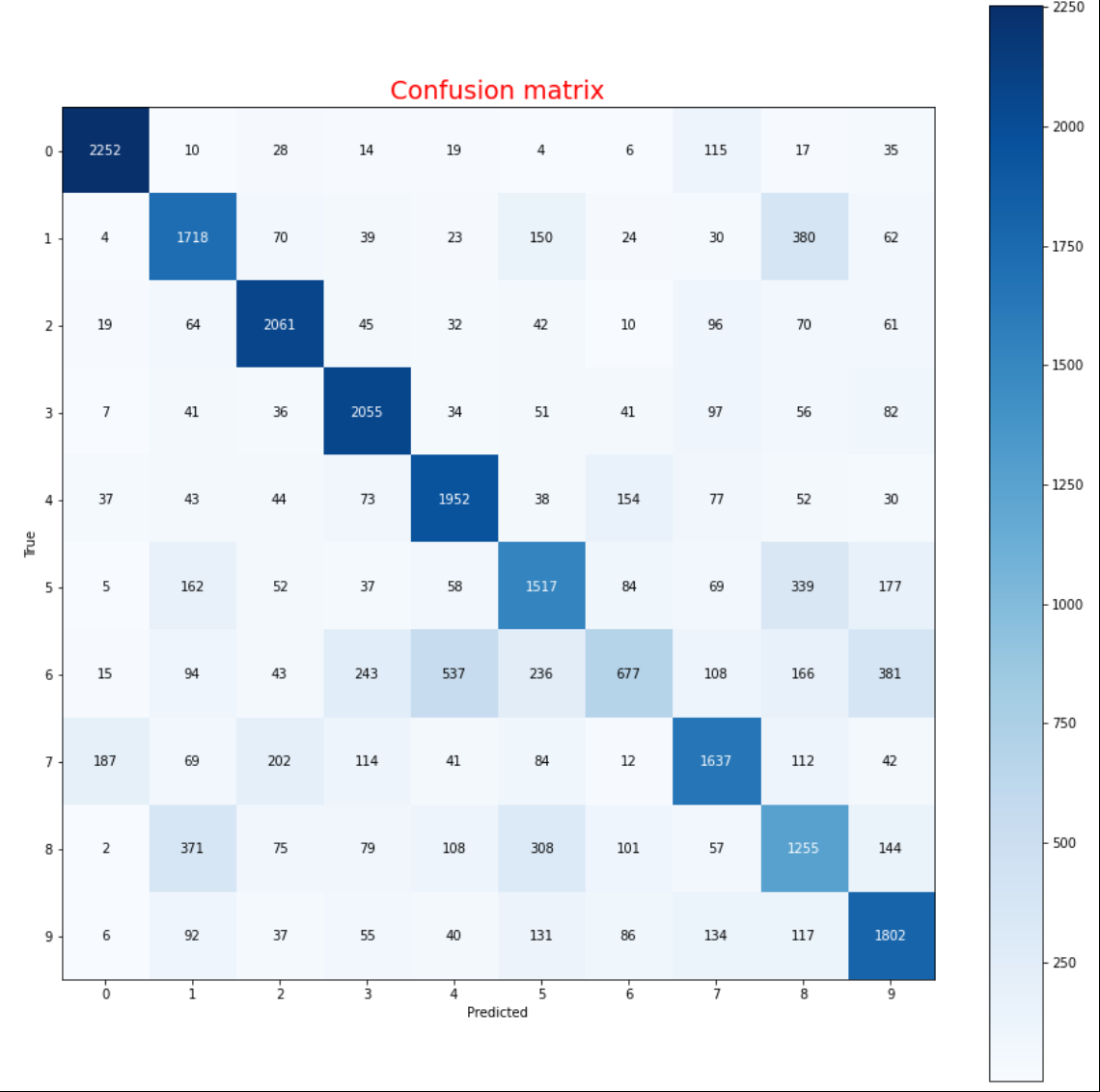
 (a)  (b)

Fig. 2 (a) Accuracies (%) vs. Regularization coefficient (lambda), and (b) Training Time (Seconds) vs. Regularization coefficient (lambda).

1. For the optimal values of M=100 and lambda=20, the performance on the test data has been investigated and confusion matrices (CMs) have been plotted in Fig. 3. In Fig. 3(a), the deterministic CM has been plotted where for a particular class how many samples have been correctly labeled is shown by the diagonal elements of the 10 X 10 matrix. In Fig. 3(b), the probabilistic performance of each classes could be investigated on the test set. Similarly, the correctly predicted labels would be the diagonal elements for each class. For each class there are 2500 observations to be predicted in the test set. The true prediction accuracy for 'Class 6' is the lowest among all classes, i.e., 27% only. However, 'Class 1' has the highest prediction accuracy of 90% on the test set. The overall accuracy on the test set is 67.7% from considering M=100 and lambda = 20.



1. (b)

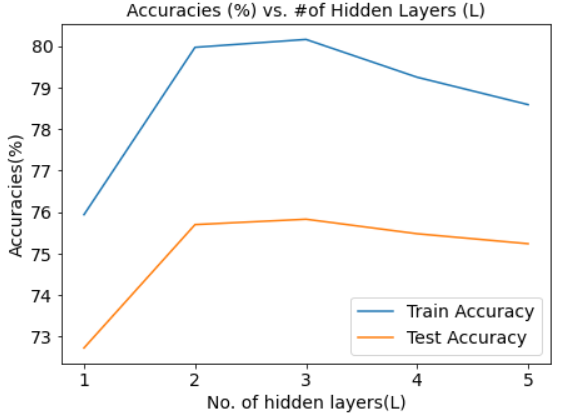
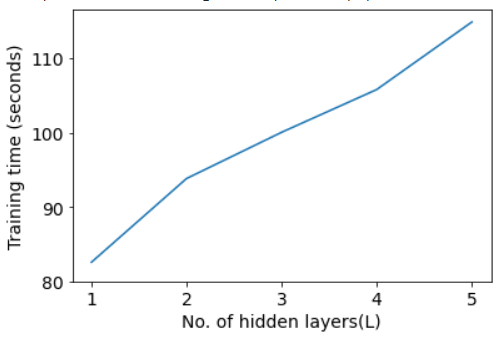
Fig. 3 (a) Deterministic CM and (b) Probabilistic CM from test set where optimal M = 100 and lambda = 20.

**Briefly discuss how your model's performance can be improved further:** The following ways could improve the ANN model's performance:

1. Modifying the model's architecture, adding more hidden layers in the network.
2. Exploring other activation functions than sigmoid.
3. Applying some efficient weight initialization procedure other than random weight initialization.

**Report 2:**

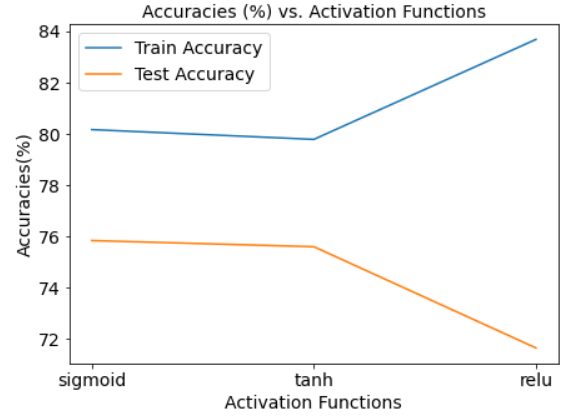
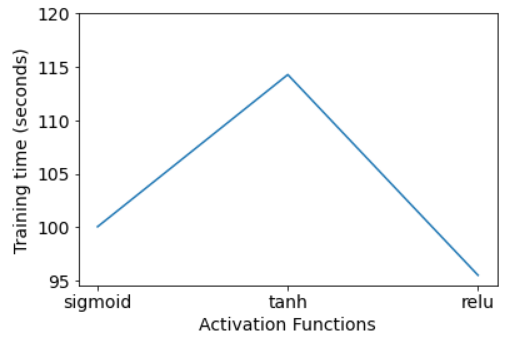
1. Train and test set accuracies vs. #of layers (L) plot has been given in Fig. 4(a). It has been seen from the figure that for L=3, the train and test accuracies are optimal i.e., 80.16% and 75.83%, respectively. Thus the model is performing at the best with 3 hidden layers of 100 hidden units (M), each. The training time in second has been plotted for different values of L = [1,5] in Fig. 4(b). the training time is increasing as the L increases.

1. (b)

Fig. 4 (a) Accuracies (%) vs. No. of hidden layers (L), and (b) Training Time (Seconds) vs. No of hidden layers (L)

1. Using the optimal M=100 and L=3 from the previous part (Report 2.1), the performance of the model (in terms of training and testing accuracies and the training time) for different choices of the activation function (*sigmoid*, *tanh*, and *relu*) has been compared in Fig. 5 (a) and (b), respectively. It has been shown in Fig. 5(a), optimal accuracies for train and test sets are for *relu* and *sigmoid* activation function, i.e., 83.68% and 75.83%, respectively. So, the best choice would be *sigmoid* activation function as its performing best on the test set. In Fig. 5(b), the best(least) training time (s) is for *relu* activation function, i.e, 95.51s. So the best choice for the image classification would be M=100, lambda = 20, L=3 and activation function = *sigmoid.*

1. (b)

Fig. 5 (a) Accuracies (%) vs Activation Functions and (b) Training Time (Seconds) vs. Activation Functions