

Lab Assignment 1

CSCI 5992 - Neural Networks and Deep Learning

Rajeev R Menon - 110581437

November 3, 2022

Problem 1

Part b

The dataset used for this experiment consists of data from forests in the United States. The samples in the dataset are of 30m x 30m patches of forests. Each patch is classified into one of seven different forest cover types ie. the species of trees which cover most of the patch. The dataset was imported from real world datasets in the scikit-learn machine learning library.

- Experiment setup
 - MacBook Pro 14 inch
 - M1 Pro Apple Silicon
- Hyperparameters
 - Optimizer: Adam algorithm
 - Learning Rate: 0.001 (Default value in sklearn)
 - Number of hidden layers: 2,3 and 4
 - Neurons per layer: 20
 - Activation functions: Sigmoid, ReLU and tanh
 - Number of epochs: 20

The dataset was split into two, 70% training and 30% test data respectively. Nine different neural network models were trained using all combinations of the three activation functions and three different numbers of hidden layers mentioned above. The accuracy and confusion matrices corresponding to all nine models are shown in figures 1-8.

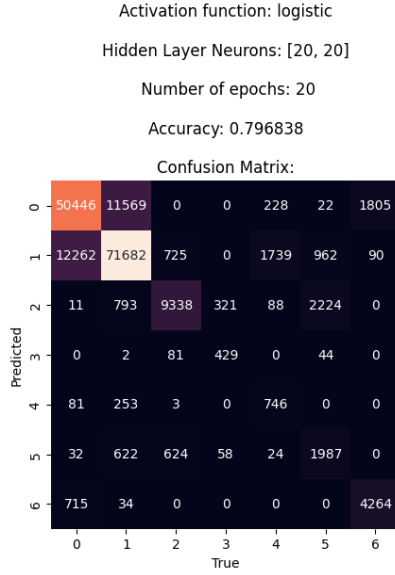


Figure 1: Sigmoid, [20, 20]

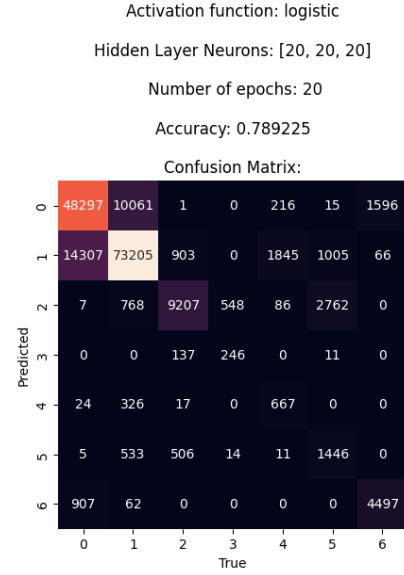


Figure 2: Sigmoid, [20,20,20]

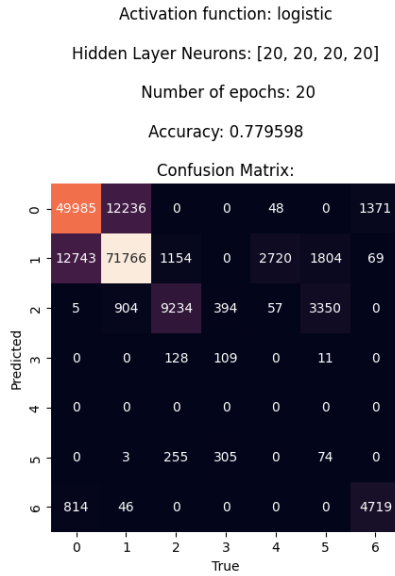


Figure 3: Sigmoid, [20, 20, 20, 20]

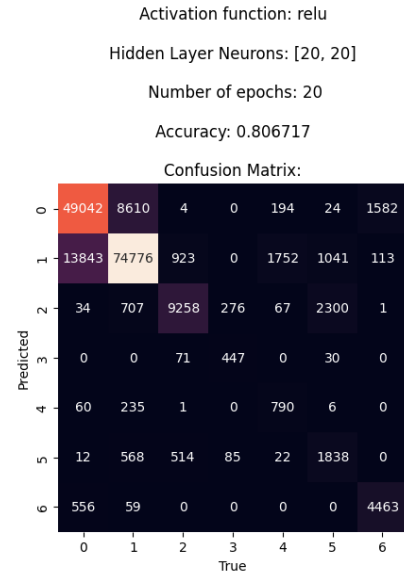


Figure 4: ReLU, [20,20]

In the three models trained using Sigmoid activation function, it is noted that the accuracy reduces with increase in the number of hidden layers. For both ReLU and Tanh activation functions, the accuracy is seen to be increasing along with the number of hidden layers. Models trained using the Tanh activation function had the better accuracy over both ReLU and Sigmoid. Models trained using ReLU functions performed slightly better than the ones using the Sigmoid function.

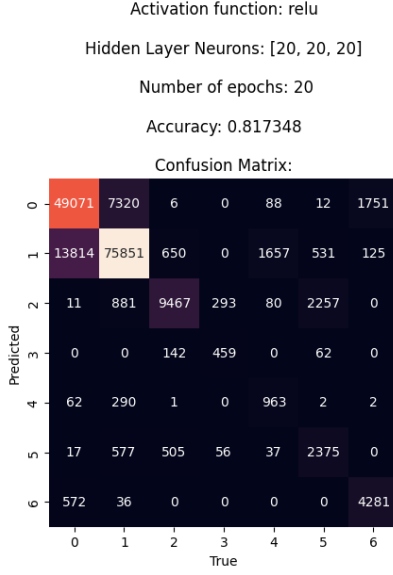


Figure 5: ReLU, [20, 20, 20]

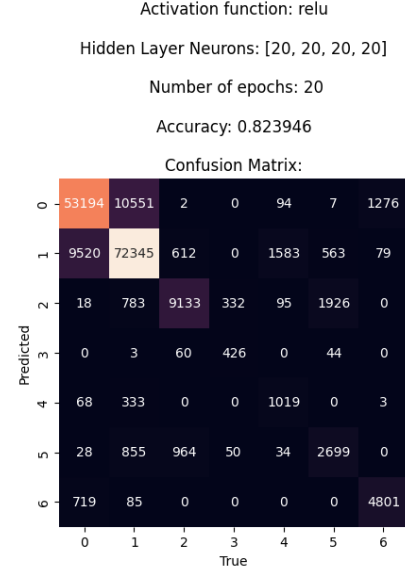


Figure 6: ReLU, [20,20,20,20]

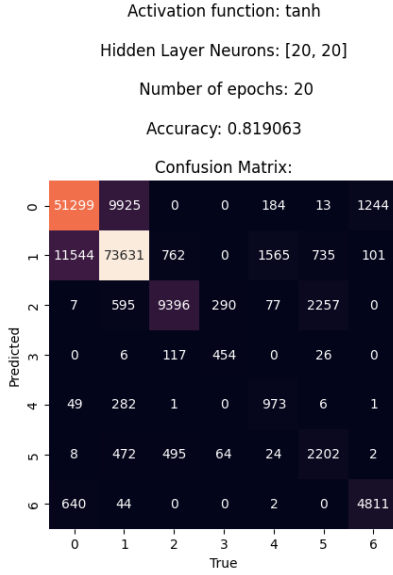


Figure 7: tanh, [20, 20]

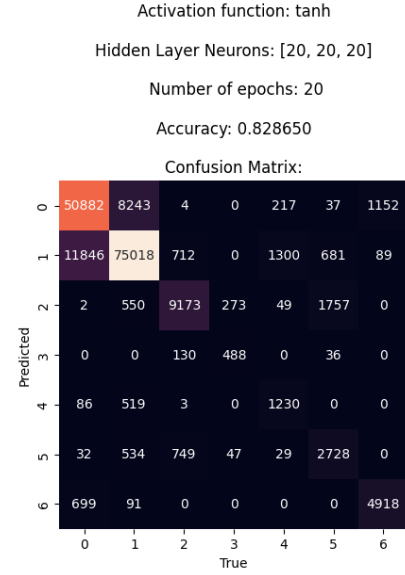


Figure 8: tanh, [20, 20, 20]

The training dataset contains more samples from classes 0 and 1, which skews the heat map highlighting only those two options more. Tanh data is known to converge quicker with lesser number of epochs when the dataset is big enough. Standardizing the data helped in stronger gradients which helps in reducing the number of steps to convergence. Thus Tanh produced the best result among all three functions.

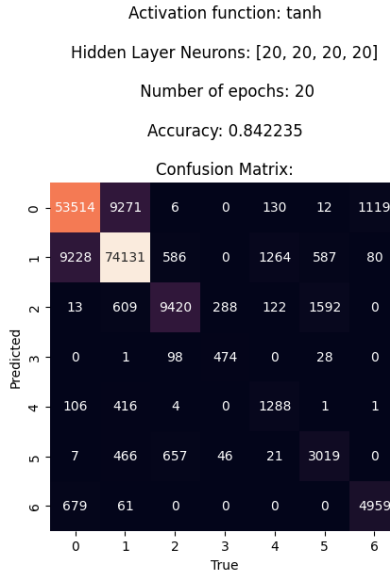


Figure 9: tanh, [20, 20, 20, 20]

Problem 2

Part b

The dataset used for this experiment consists for data from forests in the United States. The samples in the dataset are of 30m x 30m patches of forests. Each patch is classified into one of seven different forest cover types ie. the species of trees which cover most of the patch. The dataset was imported from real world datasets in the scikit-learn machine learning library.

- Experiment setup
 - MacBook Pro 14 inch
 - M1 Pro Apple Silicon
- Hyperparameters
 - Optimizer: Adam algorithm
 - Learning Rate: 0.001 (Default value in sklearn)
 - Number of hidden layers: 3
 - Neurons per layer: 20
 - Activation functions: Tanh
 - Number of epochs: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100

The dataset was split into two, 70% training and 30% test data respectively. The training data was further split into 5 different training datasets containing 20, 40, 60, 80 and 100% of the initial training data. These five datasets were trained for various number of epochs given above. The performance of those five datasets over the number of epochs was plotted for evaluation in Fig 10.

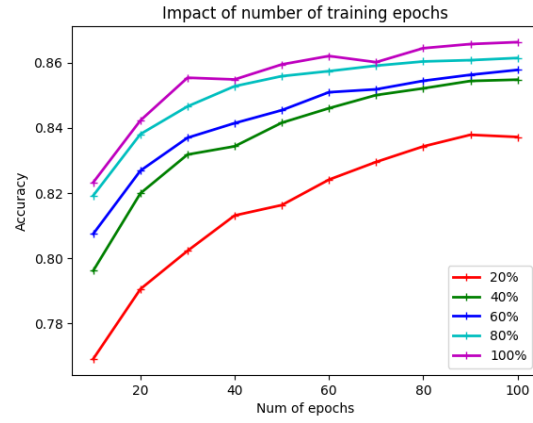


Figure 10: tanh, [20, 20, 20, 20]

The accuracy of the model increased with the number of epochs for which the model was trained. More data also was helpful in producing better accuracy.