Homework 3 Simulation Exercises

Exercise 1

The ECG dataset is attached in the HW's folder. This dataset contains four labels: normal samples (shown by N), samples having vascular diseases (A), samples having other diseases (O), samples whose experiment's results are noisy (~).

Part A: Display the number of samples in each class .

Part B: Display the histogram of each class. Regarding the histogram, explain how the unbalanced distribution of classes can cause problems for classification.

Part C: Design a Multi Layered Perceptron (MLP) network with four layers and adequate number of neurons. Split the dataset into train and test subsets with ratios of 0.9 and 0.1 respectively. Calculate precision, recall, F1-score, and also plot the confusion matrix. Explain which class has the highest precision and why.

Part D: Now standardize the features of the dataset and repeat part C. Explain why the results of this section are better than those of part C.

Part E: Now omit the noisy class labeled ~, and also merge the least two populated classes into one class. Hence, there will be two classes of normal (N) and abnormal (A) samples. Repeat part D.

Exercise 2

Generate the following classes of points:

1. 100 points in a circular section with (0, 0) as its origin and 0 and 1 as its lower and upper radius.

2. 100 points in a circular section with (0, 0) as its origin and 1.3 and 1.6 as its lower and upper radius.

Part A: Separate the classes using a MADALINE network having three neurons. Report the accuracy of model. Do not use predefined modules.

Part B: Repeat the previous section but with a MADALINE having six neurons. Explain the difference between the accuracy of model in this part and the previous part.

Exercise 3

In this question, we are going to use a Convolutional Neural Network (CNN) for classifying images from the cifar10 dataset. This dataset contains 60000 images with 10 different

labels.

Part A: Download the dataset and display 5 random images. Split the dataset into train, test, and validation sets with sizes of 40000, 10000, 10000 respectively.

Part B: Design a CNN with the following architecture:

Layer1: Conv2D, 32 neurons, 3x3 windows, Relu activation function.

Layer2: Conv2D, 32 neurons, 3x3 windows, Relu activation function.

Layer3: Conv2D, 32 neurons, 3x3 windows, Relu activation function.

Maxpooling: 2x2 windows.

Dropout: 0.25 probability.

Layer4: Conv2D, 64 neurons, 3x3 windows, Relu activation function.

Layer5: Conv2D, 64 neurons, 3x3 windows, Relu activation function.

Layer6: Conv2D, 64 neurons, 3x3 windows, Relu activation function.

Maxpooling: 2x2 windows.

Dropout: 0.25 probability.

Dense: 512 neurons, Relu activation function.

Dropout: 0.5 probability.

Dense: 10 neurons, Softmax activation function.

Part C: Use Adam, Stochastic Gradient Descent, and Residual Mean Square activation functions and compare their outputs.

Exercise 4

In this exercise, decision trees are used to classify a dataset containing samples having diabetes.

Part A: Split the dataset into train and test subsets with ratios of 0.3 and 0.7 respectively. Then, apply a decision tree on them and report the accuracy of model on both sets.

Part B: As you have been taught, pruning techniques are used to prevent decision trees from getting complicated. Now design a tree whose height is lower than or equal to 2. Display the decision trees using library functions.

Part C: Compare the accuracy of trees in part A and B. Explain how a decision tree becomes more interpretable when pruning is used.