**Assignment-3 Part-2(Report)**

**Team Members:** 1) Hardik Punjabi hvp160030

2) Aakash Shah axs165231

The implemented ANN algorithm does classification and regression of the data provided to it.

The implementation is divided into two parts. The first part is to perform pre-processing of data as mentioned in the assignment requirement. The second part is to read the post-processed data and perform classification or regression on the data.

1>Pre-processing:

A separate python programme is created for pre-processing. The programme takes the values of complete input path of the raw dataset, complete output path of pre-processed dataset as arguments. Among three datasets provided two data sets have values separated by comma whereas one has values separated by space. The program checks whether splitting should be performed on comma or space. For non-numeric data the programme check for number of unique values for the attribute and assigns numeric value accordingly. For example, the iris data will have final class attributes as 0,1 and 2 as it has three unique values. The numerical data is standardized according to the formula provided. The processed data is then written to a file, which will be stored at the output path specified.

2>Training Neural Net:

A separate python programme is created for training the neural net. The programme takes the complete path of the post-processed data set, training percentage, error tolerance, number of hidden layers and number of neurons in each layer as arguments. The programme creates the neural network according to the arguments provided. We have put a restriction on the maximum number of hidden layers to 5. The restriction has been set to keep the processing requirements low because having more hidden layers will require greater processing power. This restriction can easily be elevated if needed. The maximum number of back propagation is set to 100. The learning rate is set to 0.30. This can be easily change by changing the value of a variable. We have used the concept of cross validation to improve the test accuracy. So if 80-20 split is given the programme will generate 5 folds of training data. We experimented with different values of number of hidden layers and number of neurons in each layer. The best results obtained were as follows.

1>For iris dataset:

Number Of hidden layers: 2

Number of neurons in Hidden layer-1: 4

Number of neurons in Hidden layer-2: 3

Train accuracy: 98.833%

Test accuracy: 96.667%

2>For housing dataset:

Number Of hidden layers: 4

Number of neurons in Hidden layer-1: 2

Number of neurons in Hidden layer-2: 4

Number of neurons in Hidden layer-2: 3

Number of neurons in Hidden layer-2: 2

Train accuracy: 66.66%

Test accuracy: 63.667%

3>For adult dataset:

Number Of hidden layers:2

Number of neurons in Hidden layer-1: 3

Number of neurons in Hidden layer-2: 4

Train accuracy: 83.888%

Test accuracy: 83.729%