Pre-processing strategy

- The Iris dataset contains 4 numerical attributes and the task is to classify the sample into its corresponding class.
- There are no missing values in the dataset.
- As the attributes are numerical, they have been standardised.
- The formula used to scale is:

$$Xi - Mean(X) / Std\ Dev(X)$$

- The class labels are encoded using simple strategy:
- All unique labels are given labels.
- Iris Setosa : 0
- Iris Versicolour: 1
- Iris Virginica : 2
- And they have also been scaled to the range of $0 - 1$
- Iris Setosa : 0
- Iris Versicolour: 0.5
- Iris Virginica: 1

Best Parameters

Training Size = 80

Max Iterations = 100

Learning Rate = 0.9

No. of Hidden Layers = 2

No. of Hidden Nodes in each layer = 3, 2

Best Results

Training Accuracy = 97.5

Testing Accuracy = 96.67

Log of Experiments

Training Size = 80

Max Iterations = 100

Learning Rate = 0.9

No. of Hidden Layers = 2

No. of Hidden Nodes in each layer = 3, 2

Training Accuracy = 97.5

Testing Accuracy = 96.67

Training Size = 70

Max Iterations = 50

Learning Rate = 0.7

No. of Hidden Layers = 3

No. of Hidden Nodes in each layer = 3, 2, 2

Training Accuracy = 89.52

Testing Accuracy = 84.44

Training Size = 70

Max Iterations = 150

Learning Rate = 0.95

No. of Hidden Layers = 3

No. of Hidden Nodes in each layer = 3, 2, 2

Training Accuracy = 95.23

Testing Accuracy = 91.11

Training Size = 50

Max Iterations = 50

Learning Rate = 0.8

No. of Hidden Layers = 3

No. of Hidden Nodes in each layer = 3, 2, 1

Training Accuracy = 92

Testing Accuracy = 78.67