**Dataset Writeup**

Susung Choi

**Dataset**

The dataset that I used for testing my neural network is Iris classification dataset. The dataset can be found at : <https://archive.ics.uci.edu/ml/datasets/Iris>, a machine learning dataset repository. The purpose of this dataset is to predict the class of Iris plant based on its attributes. There are 4 attributes which include (in cm) : sepal length, sepal width, petal length, and petal width. There are three possible outputs including “Iris Setosa,” “Iris Versicolour,” and “Iris Virginica.” The dataset has 150 examples, all provided with 4 inputs and 3 different variations of output. I’ve taken 100 examples to be training set and the rest 50 to be the testing set.

Since the values of attributes were in centimeters, I normalized all the values, based on min and max of each attribute, to be between 0 and 1 with 3 digits after the decimal point. For output, I set “Iris Setosa” to “0 0 1,” “Iris Versicolour” to “0 1 0,” and “Iris Virginica” to “1 0 0.”

**Neural Network**

The learning rate for this dataset was determined by different trials. I tried to use information available online to predict the best learning rate, but I realized that trying different learning rate would be quicker and simpler. The learning rate that I’ve found to be working best is **0.25**. I did the same thing with number of epochs, and found that **200** works best. I decided to have **4** hidden nodes to keep the number of node between number of inputs and number of different outputs. Lastly, the initial weights were generated randomly, between 0 and 1 with 3 digits after the decimal point.

**Files**

init - initial network

trained - trained neural network

train - training set

test - testing set

result - results file