# Artificial Neural Networks

Ryan Plante

March 2019

## 1 Introduction

In this assignment two simple neural networks were constructed in an attempt to classify species of iris using the common "iris" data set. The data set consists of four characteristic parameters (sepal-width, sepal-height, petal-length, petal-width) and three boolean output parameters corresponding to the species of iris the characteristic parameters belong to. Each neural network was trained on a subset of the iris data set and tested on a different subset of which it had never been exposed to before. Performance metrics of training accuracy, test accuracy, and average test confidence were obtained for each network. Section 2 describes the construction of each network while Section 3 presents the results and discusses possible explanations for the observed differences.

## 2 Network Construction

Each neural network consists of three different layers, an input layer, a hidden layer, and an output layer. The input and hidden layers utilize a rectifier activation function while the output layer utilizes a softmax activation function in order to return a probability for each of the potential outputs that sums to one. Additionally, all three layers utilize a standard dense connection topology where each input node is connected to each output node.

The first layer in each network is a 4 neuron input layer, with each neuron corresponding to either sepal-width, sepal-height, petal-length, or petal width. Furthermore the output layer in each network consists of 3 neurons, one for each different iris species. Where the networks differ is in regards to the hidden layer and the amount of training done with the training data. The first neural network has a hidden layer consisting of 8 distinct neurons and was trained for 150 epochs. The second neural networks hidden layer size was increased threefold to 24 and its training time was increased by an order of magnitude to 1500 epochs. The intention with the second network was for the increased number of nodes and training to allow for a more accurate classification function to be learned which would in turn increase the overall accuracy of the results. The outputs of these two networks for the most part showed little variation, however key differences are discussed below.

## 3 Results & Discussion

The output of the first neural network is depicted below. The image shows that the accuracy of the network on the training data after 150 epochs was 98.99% which is quite good, however begs the question of whether or not the network is over fitted. Fortunately when running the network on the test data it is clear to see that it also preforms very well, reaching 100% accuracy with an average confidence of 92.58%.

```
132/132 [==============================] - 0s 76us/step - loss: 0.0669 - acc: 0.9747
Epoch 150/150
132/132 [==============================] - 0s 128us/step - loss: 0.0667 - acc: 0.9747
132/132 [==============================] - 0s 2ms/step

acc: 98.99%
Neural network predictions (for test data):
Prediction: ['0.00595395', '0.97994494', '0.01410105'] ||| Actual: [0. 1. 0.] ||| Correct?: True

Prediction: ['0.98366815', '0.01632089', '0.00001089'] ||| Actual: [1. 0. 0.] ||| Correct?: True

Prediction: ['0.00119699', '0.92404276', '0.07476018'] ||| Actual: [0. 1. 0.] ||| Correct?: True

Prediction: ['0.00338680', '0.97738373', '0.01922951'] ||| Actual: [0. 1. 0.] ||| Correct?: True

Prediction: ['0.00624010', '0.95858371', '0.03517621'] ||| Actual: [0. 1. 0.] ||| Correct?: True

Prediction: ['0.00002901', '0.10582116', '0.89414984'] ||| Actual: [0. 0. 1.] ||| Correct?: True

Prediction: ['0.00142281', '0.99016482', '0.00841232'] ||| Actual: [0. 1. 0.] ||| Correct?: True

Prediction: ['0.00000104', '0.00936655', '0.99063241'] ||| Actual: [0. 0. 1.] ||| Correct?: True

Prediction: ['0.99555236', '0.00444629', '0.00000136'] ||| Actual: [1. 0. 0.] ||| Correct?: True

Prediction: ['0.00001103', '0.08227041', '0.91771853'] ||| Actual: [0. 0. 1.] ||| Correct?: True

Prediction: ['0.00002927', '0.15817335', '0.84179741'] ||| Actual: [0. 0. 1.] ||| Correct?: True

Prediction: ['0.00493341', '0.97454673', '0.02051984'] ||| Actual: [0. 1. 0.] ||| Correct?: True

Prediction: ['0.00030171', '0.45827678', '0.54142147'] ||| Actual: [0. 0. 1.] ||| Correct?: True

Prediction: ['0.00836622', '0.97523016', '0.01640369'] ||| Actual: [0. 1. 0.] ||| Correct?: True

Prediction: ['0.00001950', '0.05805394', '0.94192660'] ||| Actual: [0. 0. 1.] ||| Correct?: True

Accuracy: 100.00%
Average Confidence: 92.58%
```

Figure 1: Output of the first neural network

Each line of the test data output displays a 1 by 3 array of prediction values the network gave to each iris species out to 8 decimal places, a 1 by 3 array of the correct output, and a boolean value denoting whether the neural network was correct or not. Although the first neural network reached 100% accuracy and did not leave much to be improved upon the second network did technically preform better. Its output is show below.

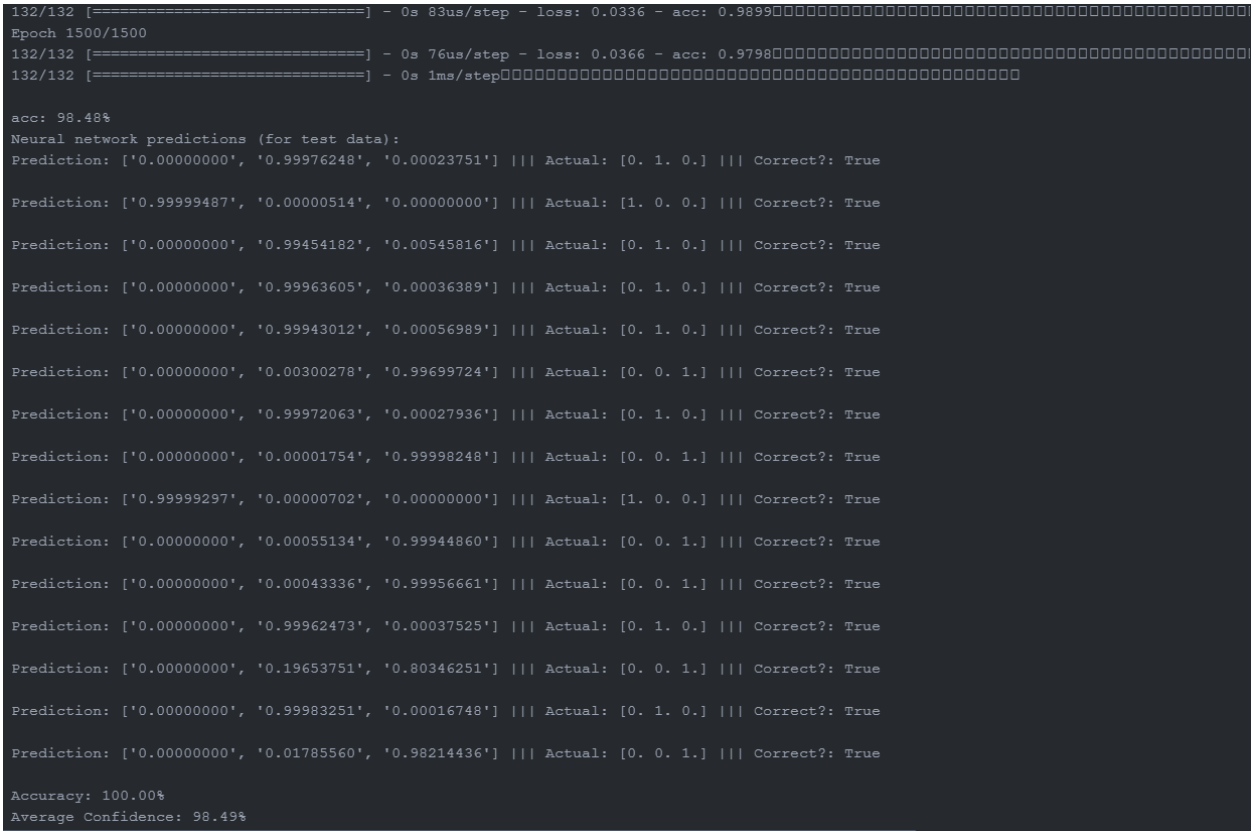


Figure 2: Output of the second neural network

As depicted above the training data accuracy was 98.48% which is marginally less accurate than that of the first network. However the test data performance still achieves an accuracy of 100% while also having a noticeably higher average confidence of 98.49%. This means that the second network was a little bit better at determining what species of iris each data point corresponded to, even thought both networks achieved 100% accuracy. The first cause of this difference is likely due to the second network having three times as many hidden neurons as the first. This in theory allows the second network to learn a more complex classification function that has the potential to better fit the data. In conjunction with this the significantly larger training time of 1500 epochs allows the second network much more time to refine each neurons weights, in the end resulting in a more complex, more refined network. That being said the test data set, and to some extent the training data set provided were quite small which makes it hard to definitively say the results are purely based on the performance of each network and not random chance.