**Name: Sayed Sohail Pasha Peerzade**

**Student Number: 220541549**

**Assignment Number: Assignment 1 Part 2 (Classification and Neural Networks)**

**Module Code: ECS708U/ECS708P**

# Report for Assignment 1 Part 2 (Classification and Neural Networks)

## Classification

**Q1.** We again notice that the attributes are on different scales. Use the normalisation method from last lab, to standardize the scales of each attribute on both sets. Plot the normalized and raw training sets; what do you observe? [2 marks]

Un-Normalised Values

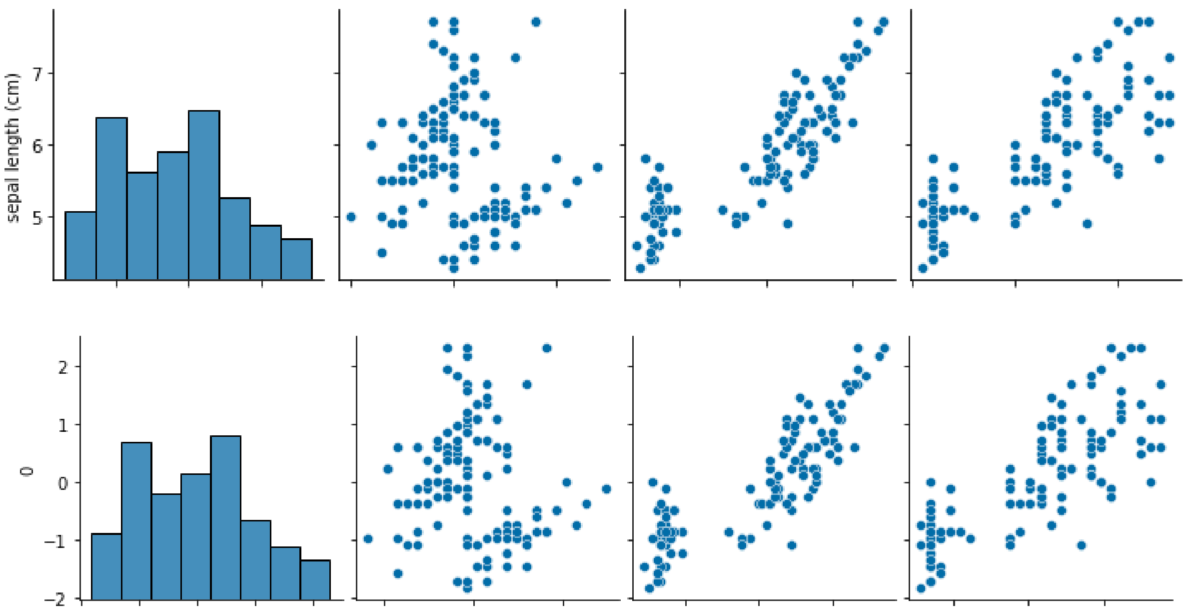
| | sepal length (cm) | sepal width (cm) | petal length (cm) | petal width (cm) |
|---|---|---|---|---|
| 22 | 4.6 | 3.6 | 1.0 | 0.2 |
| 15 | 5.7 | 4.4 | 1.5 | 0.4 |
| 65 | 6.7 | 3.1 | 4.4 | 1.4 |
| 11 | 4.8 | 3.4 | 1.6 | 0.2 |
| 42 | 4.4 | 3.2 | 1.3 | 0.2 |
| 146 | 6.3 | 2.5 | 5.0 | 1.9 |
| 51 | 6.4 | 3.2 | 4.5 | 1.5 |
| 27 | 5.2 | 3.5 | 1.5 | 0.2 |
| 4 | 5.0 | 3.6 | 1.4 | 0.2 |
| 32 | 5.2 | 4.1 | 1.5 | 0.1 |

Normalised Values

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | -1.467781 | 1.198629 | -1.556010 | -1.307120 |
| 1 | -0.132515 | 2.979875 | -1.270678 | -1.041265 |
| 2 | 1.081363 | 0.085351 | 0.384247 | 0.288010 |
| 3 | -1.225005 | 0.753318 | -1.213611 | -1.307120 |
| 4 | -1.710556 | 0.308007 | -1.384811 | -1.307120 |

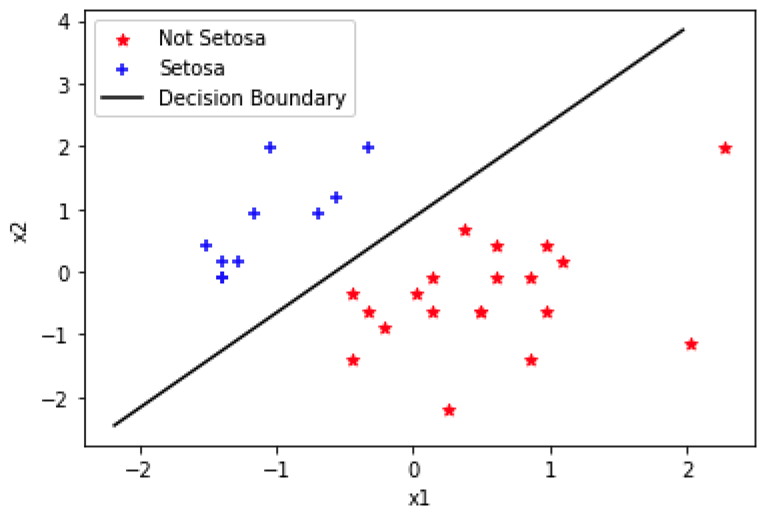
As we can see in the iris dataset the values of features sepal length, sepal width, petal length and petal width are all at different scales. Since features are in different scale, we need to bring these features to a similar scale so that performance and training stability of the model is improved.

If features are not normalised, this difference in ranges of features will cause different step sizes for each feature(In case of Gradient descent). The gradient descent will converge more quickly towards minima if the features are of similar scale.




By looking at the above plots of unnormalised sepal length and normalised sepal length we can concur that the range of values for unnormalised sepal length is between 4 to 7 cm, but after normalisation the range has been reduced to -2 to 1.

**Q5.** Draw the decision boundary on the test set using the learned parameters. Is this decision boundary separating the classes? Does this match our expectations? [2 marks].




The decision boundary is a line that separates data points of different class labels. For the above example the test inputs are separated into Setosa and Not Setosa (Class 0).

**Q6.** USING THE 3 CLASSIFIERS, PREDICT THE CLASSES OF THE SAMPLES IN THE TEST SET AND SHOW THE PREDICTIONS IN A TABLE. DO YOU OBSERVE ANYTHING INTERESTING?

| Sepal Length | Sepal Width | Setosa prediction | Versicolor prediction | Virginica prediction | Belongs to Setosa class | Belongs to Versicolor class | Belongs to Virginica class |
|---|---|---|---|---|---|---|---|
| 0.14200716 | -0.623848 | 0.06670781 | 0.7569088 | 0.42473173 | 0.0 | 1.0 | 0.0 |
| -0.3313506 | 1.9755213 | 0.99998736 | 0.11287928 | 0.14569163 | 1.0 | 0.0 | 0.0 |
| 2.0354376 | -1.143722 | 6.596948e-06 | 0.8565093 | 0.9086236 | 0.0 | 0.0 | 1.0 |
| 0.023667859 | -0.36391073 | 0.07814734 | 0.59129304 | 0.6471505 | 0.0 | 1.0 | 0.0 |
| 0.97038335 | -0.623848 | 0.014543642 | 0.76736563 | 0.46476084 | 0.0 | 1.0 | 0.0 |
| -0.68636847 | 0.9357739 | 0.9998324 | 0.24795699 | 0.08861963 | 1.0 | 0.0 | 0.0 |
| -0.4496899 | -0.36391073 | 0.4200083 | 0.5396897 | 0.46128887 | 0.0 | 1.0 | 0.0 |
| 1.0887227 | 0.15596263 | 0.0052441475 | 0.34047928 | 0.9454898 | 0.0 | 0.0 | 1.0 |
| 0.26034644 | -2.1834693 | 0.0004288236 | 0.93243825 | 0.2640918 | 0.0 | 1.0 | 0.0 |
| -0.21301073 | -0.8837847 | 0.11338656 | 0.7450932 | 0.31286508 | 0.0 | 1.0 | 0.0 |

We have created a One vs All multiclass classification by training the models for each of the 3 classes (Setosa Prediction, Versicolor Prediction, Virginica Prediction). The prediction values obtained from each of the classifiers depict the probability that it belongs to that class.
The above table displays the three class predictions Setosa Prediction, Versicolor Prediction, Virginica Prediction) along with actual classes being displayed in the following three columns (Belongs to Setosa Class, Belongs to Versicolor class, Belongs to Virginica Class).

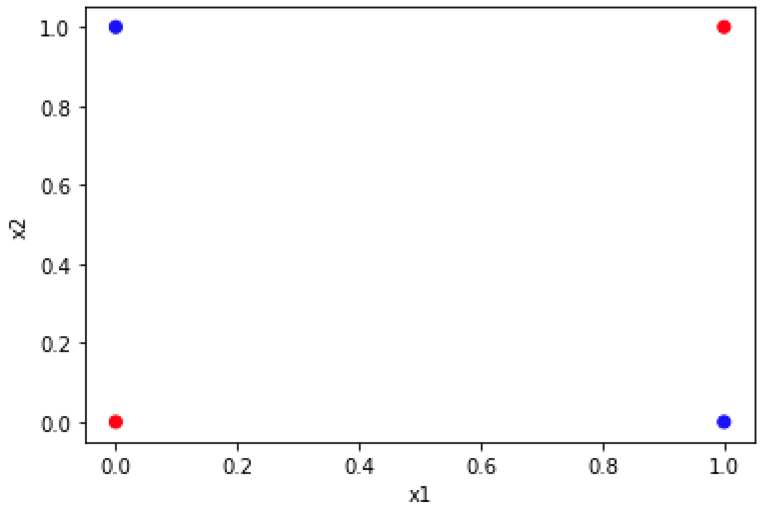
For each row we have three predictions, and the maximum value among these three predictions will be our predicted result. For example, in the first row the Setosa prediction is 0.0667(6.67%), Versicolor Prediction is 0.7569(75.69%) and Virginica Prediction is 0.4347(43.47%). The maximum of these three probabilities is 0.7569 (75.69%) Versicolor Prediction. This is corroborated by the column Belongs to Versicular class which has 1 (True) and all others have a 0.
Since the actual class and the maximum probability of 0.7569 (75.69%) Versicolor Prediction were the same, we can conclude that out classifier can accurately predict if a feature belongs to the specific class.

**Q7.** Calculate the accuracy of the classifier on the test set, by comparing the predicted values against the ground truth. Use a softmax for the classifier outputs. [1 mark]

```
No of predictions:  30.0
No Of lables predicted correctly:  tensor(27.)
Accuracy:  tensor(90.)
```

To calculate the accuracy of the classifier, the predictions from 3 classifiers were concatenated and were passed into the softmax function. Of 30 predictions made, the model accurately predicted 27 predictions. And the model has an accuracy rate of 90%.

**Q8.** Looking at the datapoints below, can we draw a decision boundary using Logistic Regression? Why? What are the specific issues or logistic regression with regards to XOR? [2 marks]




We cannot draw a decision boundary using Logistic Regression for the above given data plot, because logistic regression creates a linear boundary around the data points to classify them under one class or another. In the above plot we can see that there are two classes and 4 data points, the data points are so far apart that no straight line can provide the accurate classification in this case. To solve the XOR problem the input data must be recreated by adding an additional dimension, turning it from a 2 dimensional problem to a 3 dimensional problem. Then we can draw a straight line in the 3d vector space to classify the data points.

## Neural networks

**Q1.** Why is it important to use a random set of initial weights rather than initializing all weights as zero in a Neural Network? [2 marks]

If all weights were initialised to zeros, the neurons perform the same computations in each iteration and produce the same outputs. If all the weights were assigned to zero then the derivatives would remain the same for every weight. This problem is called a network failing to break symmetry and this will occur not only in case of zero initialisations but all constant initialisations. Zero Initialisations lead to models with lower accuracy.

Whereas, random initialisation prevents neurons from learning the same features of its inputs. This method is used to break the symmetry and the process gives better accuracy than zero/constant initialisations.

**Q2.** How does a NN solve the XOR problem? [1 marks]

In the XOR problem, the linear separability of data is not possible as we cannot draw a decision boundary around the data points. However Neural Networks are made up of multiple extra layers called hidden layers. The weights and bias are updated via backpropagation through a number of such layers.