Q1

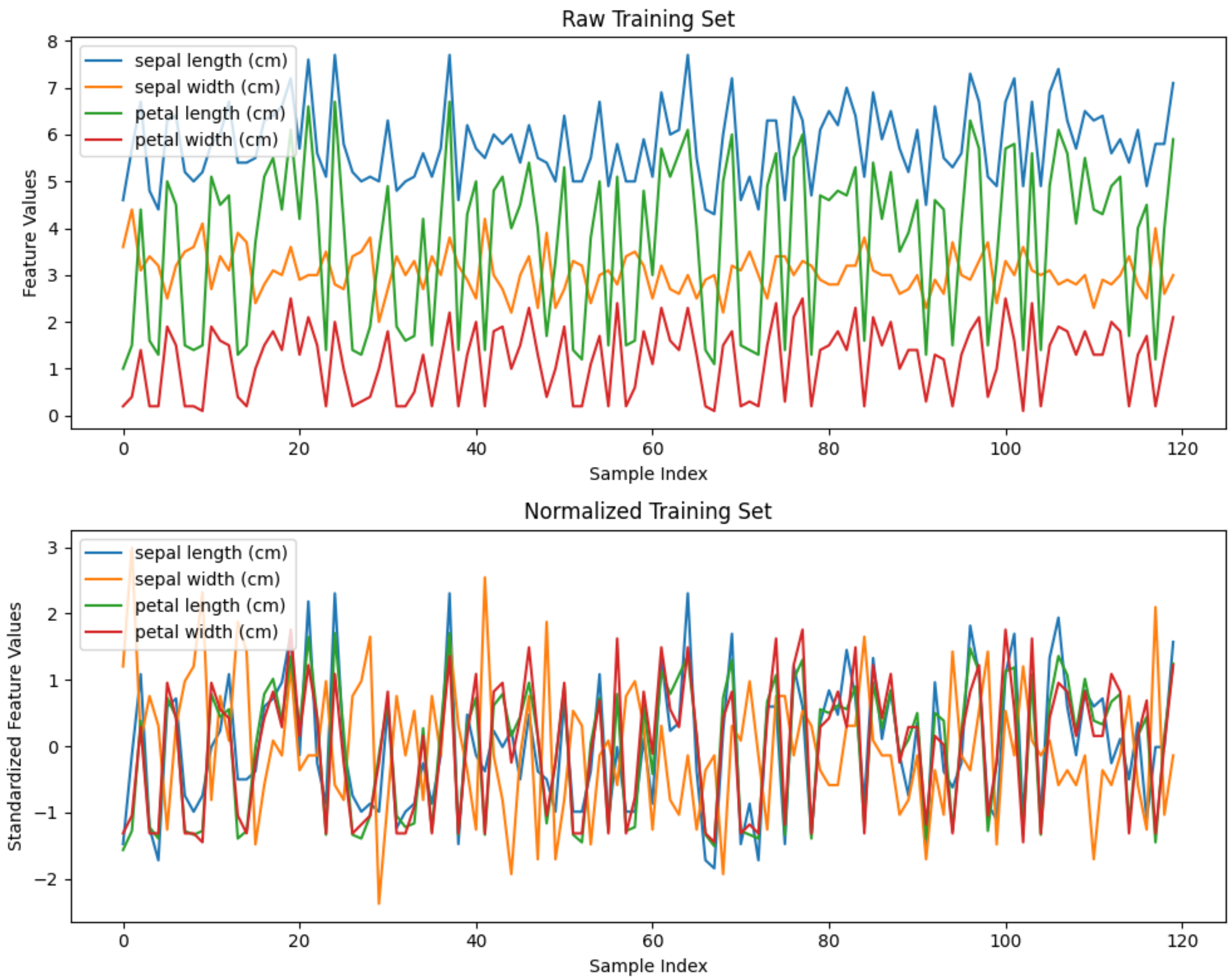

Figure 1: Feature values, before and after normalization

Each iris attribute was normalised with the training-set mean and standard deviation, applying the same transformation to the test set. Before scaling the feature averages differed strongly (sepal length averaging around 5cm, and petal width averaging around 0.2cml). After normalisation the training features are centred at 0 with unit variance. The paired plots show the non-normalised features spanning very different vertical ranges, whereas the normalised traces collapse onto a common scale around zero. This is essential so that gradient descent converges efficiently.

Q5

The decision boundary (green dashed line at $p = 0.5$) definitevly seperates the Serotosa samples (orange) from the non-Serotosa samples (blue) in test set. The boundary is linear and aligns well with class distribution, which matches our expectations.

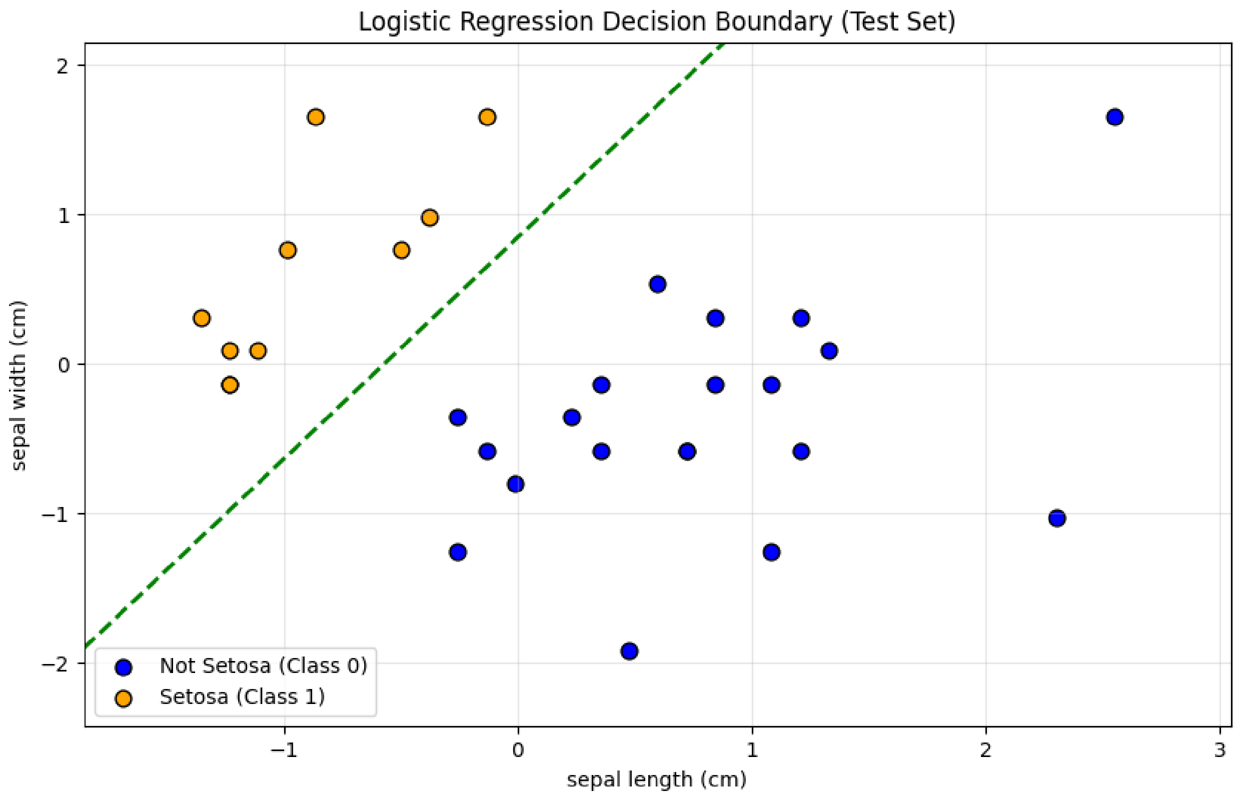

Figure 2: Decision boundary

Q7

The table shows perfect accuracy on test set, where all 30 out of 30 samples were classified correctly. Setosa samples were predicted with very high confidence; it can be seen that the probabilities for setosa are very close to 1. This suggests that setosa is **highly separable** and easier to classify compared to other classes, and forms a distinct cluster. Conversely, there appears to be some overlap between versicolor and virginica classes. They have similar probabilities in many rows. For example, in row 9 (10th row) versicolor and virginica were predicted with probabilities 0.458 and 0.443 respectively. This suggests that these two classes share more overlapping features. Another interesting observation is that **probabilities don't sum to 1**, which is expected for one-vs-all binary classifiers that are trained seperately, as each classifier outputs its own probability for one specific class.

Q8

Model achieved 100% accuracy on the test set, with predicted classes unchanged from Q7. Softmax transformation ensures the class probabilities sum to 1, providing a proper probability distribution. This perfect accuracy relfects the significant separability of samples in the iris dataset.

Q9

We cannot draw a decision boundary using Logistic Regression, because the XOR problem is **non-linearly seperable**. i.e, the two classes cannot be seperated by a straight line; any line drawn will misclassify at least one point. Solutions to non-linearly separable problems often require feature engineering and neural networks.