Lab 4 report (Classification)

- **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?**
    - The plots show that the raw training set has features on different scales, with Feature 1 ranging roughly from 4.5 to 7.5 and Feature 2 from 2.0 to 4.5. After normalization, both features are scaled to a similar range, typically centered around zero.
- **Draw the decision boundary on the test set using the learned parameters. Is this decision boundary separating the classes? Does this match our expectations?**

    - No, for my model I couldn't make prediction since the number of features is not the same as my model is trained to.

- **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?**

    - The predictions table reveals that the model often outputs high probabilities for multiple classes, indicating some overlap or ambiguity between classes, especially **Virginica**. It appears that the model never predicts the **Virginica** class (the **Predicted_Class** column does not show the value 2). **Setosa** predictions, in particular, are consistently confident, suggesting clearer separation in the feature space. Misclassifications tend to have high confidence, indicating that certain samples might be challenging for the model to distinguish.

- **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?**

    - Accuracy on the test set: 63.33%
      (**See notebook)**

- **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?**

  Logistic regression works by creating a linear decision boundary (a line in 2D) between two classes. However, the XOR problem requires a **non-linear decision boundary** to separate the classes effectively, which logistic regression cannot provide.

  Specifically:

  1. **Linear Separability**: Logistic regression assumes that the two classes can be separated by a single straight line. In XOR, this is not possible because each class forms a diagonal pattern. No straight line can separate the points of one class from the other.
  2. **Decision Boundary Requirements**: The XOR pattern needs a decision boundary that can form two separate regions to capture each class. Linear models, like logistic regression, cannot capture this structure because they can only separate data with a single, continuous decision boundary.

  -