BIRZEIT UNIVERSITY

Electrical and Computer Engineering Department

Machine Learning and Data Science - ENCS5341

Homework #2

Submission deadline: 8.1.2023 (***Late submissions will not be accepted***)

The `train.csv` file contains a set of training examples for a binary classification problem, and the testing examples are provided in the `test.csv` file. The following figures show these examples.

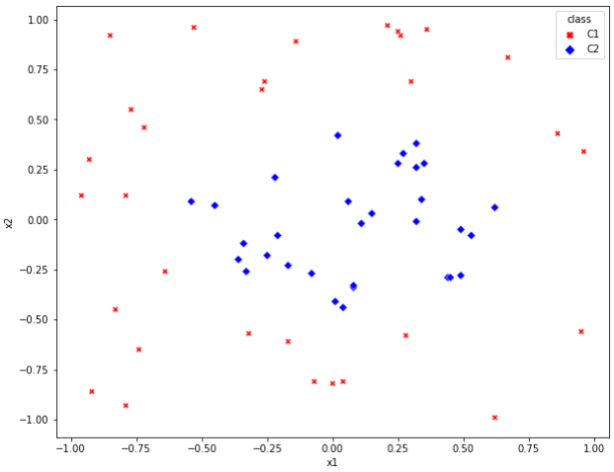

Figure 1 Training Set

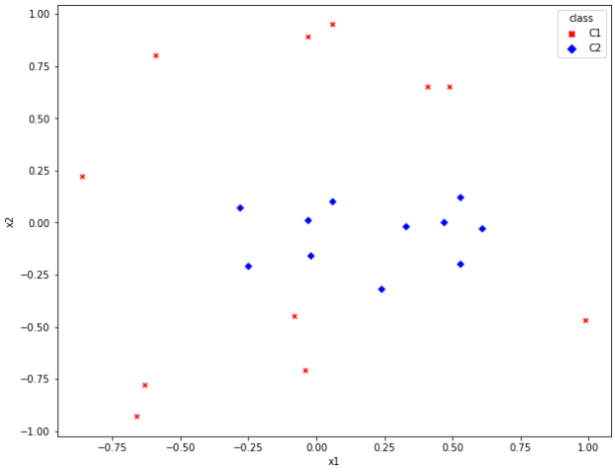

Figure 2 Testing Set

1- Learn a logistic regression model with a linear decision boundary by implementing the gradient descent algorithm. Draw the decision boundary of the learned model on a scatterplot of the training set (similar to Figure 1). Compute the training and testing accuracy of the learned model.

2- Repeat part 1 but now learn a logistic regression model with a decision boundary of the form $w_0 + w_1 x_1^2 + w_2 x_2^2$.

3- Comment on the learned models in 1 and 2 in terms of overfitting/underfitting.

4- Repeat parts 1 and 2 but now using the logistic regression implementation of `scikit-learn` python library.

You have to submit both the code (using python) and the results of running your code. To better structure your submission, it is recommended to use `Jupyter Notebook`.

To read csv files you can use the `pandas` library.