**Machine Learning**

**Fall 2022**

**HW6**

**Due: xxxxxxxxxxx, via Blackboard**

**Problem 1:** In this exercise. you are required to write a Matlab code of a SVM model to classify the dataset ‘**dataset3forSVM.csv**’.

**Table 1 (0:class 1, 1: class 2)**

|  |  |  |
| --- | --- | --- |
| **Exam 1** | **Exam 2** | **Status** |
| -0.1589 | 0.42397 | 1 |
| -0.3479 | 0.4707 | 1 |

**.**

**.**

**.**

|  |  |  |
| --- | --- | --- |
| -0.31694 | -0.22895 | 1 |
| -0.29412 | -0.1348 | 0 |
| -0.15311 | 0.1845 | 0 |

Please divide the entire dataset into two subsets (training and test sets) in a ratio of 7:3, which means 70% of the dataset is the training set (148 samples) and the remaining 30% is the test set (63 samples). Under different kernel functions, you are required to train different SVM models with the training set and test the classification performance of the trained SVM models by the test set.

1. Plot the original dataset.
2. With different kernel functions (‘gaussian’, ’linear’, and ’polynomial’), you are required to train different SVM models and compute the accuracy on the training set and test set. (please try to tune KernelScale/PolynomialOrder/other parameters in ‘fitcsvm’ function to make sure that the training and test set accuracy both is greater than 85% ).
3. With different kernel functions (‘gaussian’, ’linear’, and ’polynomial’), you are required to train different SVM models. Include the decision boundary and training set points in a figure, and, in another figure, include the decision boundary and test set points. In the figures(training set points and decision boundary), please indicate the support vectors.

**Note**: When you submit the homework, you should ensure that the homework includes following items:

**a matlab file:**

maincode.m / maincode.mlx

**A report** that briefly addresses 1) accuracy on training and test sets, 2) decision boundary figures, 3) a flow chart of your codes, 4) all other results which you think are necessary.

If you make changes to the dataset file, such as changing and adding the column name, please upload your dataset to BB.

**Hints:**

1. You do not need to preprocess the dataset.
2. Make sure all the accuracy on the training set and test set>=85%.