Systems and Biomedical Engineering Department
Faculty of Engineering
Cairo University

Fourth Year / AI in Medicine

Due Date 26/12/2020

Assignment #3 Support Vector Machines

Problem# 1(5 pts):

Employ the Credit dataset Inside the compressed file, you will find two files:

1- data.txt: provides the examples with the last column for class category (1 or 2).
2- readme.data.txt: describes the information regarding this dataset

You are required to train and understand SVM classification model, by randomly select 60% of the examples from the data.txt file as training examples, and use the remaining examples for testing.

1- Train the SVM classifier while employing a linear kernel. Repeat the experiment ten times. In the submitted homework, you need to include the classification accuracy of SVM on the test examples averaged over ten trials.
2- Train SVM with the scaled dataset Instead of using the raw data from the data.txt file for training a SVM classification model. You need to conduct preprocessing for SVM , which is initialized by subtracting each attribute by its mean computed from the training data, and then scale each attribute to the range between —1 and 1. Rerun the above experiment, and calculate the classification accuracy of trained SVM model on the test examples. Discuss the difference in the results between scaled datasets and non-scaled datasets. In the submitted homework, you need to include ((2) the classification accuracy of SVM on the test examples averaged over ten runs, and (3) discuss the difference in the results between using the scaled data sets and not-scaled data sets. Try to explain why.

Problem# 2(3 pts):

Implement, from scratch, linear SVM model using Gradient descent as an optimization function. Your model should have a function called fit for training the data and predict for the prediction and any other needed function. use Iris Data for training and testing your binary classifier model.
Hint: Use only two features. and plot them together.

General Instructions

1- This is an individual task.
2- The source code as well as the report describing your functions and output should be submitted through Google Classroom.
3- The due date for the submission of this phase is Saturday, December, 26, 2020 at 12:00 am.
4- Please Review the definition of cheating in the first presentation.

Best Regards,