

ENGR-UH 4560

Selected Topics in Information and Computational Systems

Mini Project - SVM for classification

Due Date: Refer to NYU Class

Introduction

Support vector machine is another simple algorithm that every machine learning expert should have in the arsenal. Support vector machine is highly preferred by many as it produces significant accuracy with less computation power. Support Vector Machine abbreviated as SVM can be used for both regression and classification tasks. But it is widely used in classification objectives. The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin.

Dataset

In this project, we are going to use synthetic data. We provided two files “X.csv” and “y.csv” where the “X.csv” contains the information for all the instances and “y.csv” has labels corresponding to each instance in “X.csv”. For each instance, there are two attributes and there are total 100 instances.

Requirements

1. Complete the SVM Primal problem Submit your program and answer the following questions.
2. The data comes from the file X.csv and the labels come from y.csv. Load the data and target from X.csv and y.csv
3. Plot the dataset with the labels.
4. Compute the SVM primal problem and return the weights
5. Print the decision boundary.
6. Which of the training examples are closest to the decision boundary in the SVM primal problem?
7. What is the decision function $f_{\text{primal}}(x)$
8. Using this decision function, what are the labels of (3.0, 1.5) and (1.2, 3.0)

Deliverables

A zip file containing the following:

1. a working project (source code, makefiles if needed, etc)
2. a report for the detailed description of the project
 - a. Instructions on how to run your project
 - b. Answers to the programming questions.

Before submitting your project, please make sure to test your program on the given dataset.

Notes

*You may discuss the general concepts in this project with other students, but you must implement the program on your own. **No sharing of code or report is allowed.** Violation of this policy can result in a grade penalty.*

Late submission is acceptable with the following penalty policy:

10 points deduction for every day after the deadline