CS 4340/5340

Project #1

Points 100, Due: 11:59 pm, Sep 14

Implement the Perceptron Learning Algorithm (PLA) for binary classification. Any one variant of the PLA will do. You are free to choose your programming language. Please do not use an off-the-shelf implementation of PLA from any package. The goal is for students to develop a solid understanding of the working of the PLA as a vehicle for demonstration of the connection between training error and test error. We will use the PLA as a running example in illustrating many of the recurring themes in machine learning, namely hypothesis set complexity, VC dimension, regularization, etc. Choose appropriate training and test data sets of two dimensions (points on a plane). Use 50 data points for training (25 for each class) and 30 for testing. Submit a **single** doc/docx or pdf file (**no other file type, please**) containing the source code, the training and test data, and brief notes on:

1. whether the training data points are linearly separable,
2. whether the test points are linearly separable,
3. your initial choice of the weights and constants,
4. the final solution equation of the line (decision boundary),
5. the total number of weight vector updates that your algorithm made,
6. the total number of iterations made over the training set, and
7. the final misclassification error, if any, on the training data and also on the test data.

Also, re-run your code by varying the following and describe the effect, if any, that the variation had on the final solution:

8. the initial choice of the weights,
9. the initial choice of the step size constant (c),
10. the order in which you consider the points in the training set.

Please submit (on Blackboard) a single pdf or Word file (no other format, please). Add notes for any other special issues/techniques that you think might be important in your implementation. Feel free to discuss your project with your classmates but your final submission should be your own, independent work.

Evaluation of your work will be based roughly on the following criteria:

- code correctness and completeness 60%,
- addressing the above-mentioned ten issues in your code and in the report (please number your written responses 1 to 10 in the report) 30%,
- program readability, maintainability, and software engineering issues 10%

You will find numerous implementations of PLA on the Internet; feel free to check them, but write your own code.