Homework #1

RELEASE DATE: 09/13/2012

DUE DATE: 9/27/2012, BEFORE THE END OF CLASS

QUESTIONS ABOUT HOMEWORK MATERIALS ARE WELCOMED ON THE FORUM.

Unless granted by the instructor in advance, you must turn in a printed/written copy of your solutions (without the source code) for all problems. For problems marked with (*), please follow the guidelines on the course website and upload your source code to designated places.

Any form of cheating, lying, or plagiarism will not be tolerated. Students can get zero scores and/or fail the class and/or be kicked out of school and/or receive other punishments for those kinds of misconducts.

Discussions on course materials and homework solutions are encouraged. But you should write the final solutions alone and understand them fully. Books, notes, and Internet resources can be consulted, but not copied from.

Since everyone needs to write the final solutions alone, there is absolutely no need to lend your homework solutions and/or source codes to your classmates at any time. In order to maximize the level of fairness in this class, lending and borrowing homework solutions are both regarded as dishonest behaviors and will be punished according to the honesty policy.

You should write your solutions in English with the common math notations introduced in class or in the problems. We do not accept solutions written in any other languages.

1.1 Learning Cannot Always Help

Argue that Machine Learning may not be able help (much), or may not be needed, in the following four prediction tasks. You need to make your arguments *convincing* to get the points.

(1) (5%) Predicting the lottery-winning numbers for tomorrow.

(2) (5%) Predicting whether a given graph contains a cycle.

(3) (5%) Predicting whether the earth will be destroyed by the misuse of nuclear power in the next ten years.

(4) (5%) Predicting whether Hsuan-Tien Lin's baby daughter will cry in the next thirty minutes.

1.2 Learning Exercises

In the following exercises, you need to make your arguments *convincing* to get the points.

(1) (5%) Do Exercise 1.1 of LFD on "detection of breast cancer from X-ray images of the breast." (Search KDD Cup 2008 if you want to know more.)

(2) (5%) Do Exercise 1.1 of LFD on "predicting whether a cellphone customer would switch her/his service provider." (Search KDD Cup 2009 if you want to know more.)

(3) (5%) Do Exercise 1.1 of LFD on "predicting movie rating" as described in Section 1.1 of LFD. (Search Netflix Challenge if you want to know more.)

(4) (5%) Do Exercise 1.1 of LFD on "predicting whether an Internet advertisement will be clicked or not". (Search KDD Cup 2012 if you want to know more.)

(5) (10%) Do Exercise 1.1(e) of LFD.

1.3 Types of Learning

In the following exercises, you need to make your arguments *convincing* to get the points.

(1) (5%) Do Exercise 1.6(a) of LFD.

(2) (5%) Do Exercise 1.6(b) of LFD.

(3) (5%) Do Exercise 1.6(c) of LFD.

(4) (5%) Do Exercise 1.6(d) of LFD.

1.4 Perceptron Learning Algorithm

(1) (5%) Do Exercise 1.2(a) of LFD.

(2) (5%) Do Exercise 1.2(c) of LFD.

1.5 Proof of Perceptron Learning Algorithm

(1) (5%) Do Exercise 1.3(a) of LFD.

(2) (5%) Do Exercise 1.3(b) of LFD.

(3) (10%) Do Problem 1.3(a) of LFD.

(4) (10%) Do Problem 1.3(b) of LFD.

(5) (10%) Do Problem 1.3(c) of LFD.

(6) (10%) Do Problem 1.3(d) of LFD.

(7) (10%) Do Problem 1.3(e) of LFD.

1.6 Experiments with Perceptron Learning Algorithm (*)

(1) (10%) Do Problem 1.4(a) of LFD.

(2) (10%) Do Problem 1.4(b) of LFD.

(3) (10%) Do Problem 1.4(c) of LFD.

(4) (10%) Do Problem 1.4(d) of LFD.

(5) (10%) Do Problem 1.4(e) of LFD.

(6) (10%) Do Problem 1.4(g) of LFD.

1.7 Another Perceptron Learning Algorithm

The original perceptron learning algorithm does not take the "seriousness" of the prediction error into account. That is, regardless of whether $y(t)\mathbf{w}^T(t)\mathbf{x}(t)$ is very negative or just slightly negative, the update rule is always

$$\mathbf{w}(t+1) \leftarrow \mathbf{w}(t) + y(t)\mathbf{x}(t).$$

Dr. Learn decides to use a different update rule. Namely, if $y(t) \neq \text{sign}(\mathbf{w}^T(t)\mathbf{x}(t))$, the doctor will use

$$\mathbf{w}(t+1) \leftarrow \mathbf{w}(t) + y(t)\mathbf{x}(t) \cdot \left\lfloor \frac{-y(t)\mathbf{w}^T(t)\mathbf{x}(t)}{\|\mathbf{x}(t)\|^2} + 1 \right\rfloor.$$

(1) (Bonus 5%) Prove that with the new update rule, $y(t)\mathbf{w}^T(t+1)\mathbf{x}(t) > 0$. That is, $\mathbf{w}(t+1)$ always classifies $(\mathbf{x}(t), y(t))$ correctly.

(2) (Bonus 5%) When the data set is linearly separable, does this new update rule still ensure convergence to a linear separator? Why or why not?