

Homework #2

RELEASE DATE: 09/27/2012

DUE DATE: 10/11/2012, 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.

2.1 Simplified No-Free-Lunch Theorem

- (1) (8%) Do Problem 1.10(a) of LFD.
- (2) (8%) Do Problem 1.10(b) of LFD.
- (3) (8%) Do Problem 1.10(c) of LFD.
- (4) (8%) Do Problem 1.10(d) of LFD.
- (5) (8%) Do Problem 1.10(e) of LFD.

2.2 A Concrete Learning Case

In this homework, we will use a concrete case of learning to help understand the whole framework. We will take the following settings:

- $\mathcal{X} = [-1, 1] \times [-1, 1]$, $\mathcal{Y} = \{-1, +1\}$
 - $f(\mathbf{x}) = \text{sign}(x_1 - 2x_2)$, with $\text{sign}(0) = +1$ for simplicity
 - $\mathcal{H} = \{h_1, h_2\}$. That is, $M = 2$.
 - $h_1(\mathbf{x}) = \text{sign}(-x_2)$
 - $h_2(\mathbf{x}) = \text{sign}(x_1)$
 - $P(\mathbf{x})$ is a uniform distribution within \mathcal{X} .
- (1) (10%) Calculate $E_{\text{out}}(h_1)$ and $E_{\text{out}}(h_2)$. Which is worse? (*Hint: h_2 shall be worse!*)
 - (2) (10%) Consider generating $N = 10$ examples by sampling i.i.d. from $P(\mathbf{x})$ and evaluating through $f(\mathbf{x})$. Numerically calculate the exact probability that $E_{\text{in}}(h_1) = 0$.
 - (3) (10%) Let $\epsilon = 0.99E_{\text{out}}(h_1)$. Calculate the Hoeffding's bound $2 \exp(-2\epsilon^2 N)$ for $N = 10$. Argue that the choice of ϵ allows the event $[E_{\text{in}}(h_1) = 0]$ to be included in the BAD event $[|E_{\text{in}}(h_1) - E_{\text{out}}(h_1)| > \epsilon]$ in Hoeffding's inequality. Compare with the probability you just calculated and briefly state your findings.

- (4) (10%) Consider a learning algorithm that always returns the hypothesis $h \in \mathcal{H}$ with a smaller E_{in} as the final hypothesis g . When there is a tie, the hypothesis h_1 will be returned. Illustrate a data set in which h_2 (*Hint: the worse hypothesis*) will be returned instead of h_1 (*Hint: the better hypothesis*) by the algorithm.
- (5) (10%) Numerically calculate the exact probability of generating a size-10 data set (with P) such that h_2 will be returned instead of h_1 by the algorithm.
- (6) (10%) Let $\epsilon = 0.5(E_{\text{out}}(h_2) - E_{\text{out}}(h_1))$. Calculate the extended Hoeffding's bound $2 \cdot 2 \exp(-2\epsilon^2 N)$ for $N = 10$. Argue that the choice of ϵ allows the event [the algorithm returns h_2] to be included in the BAD event $[\max_{h \in \{h_1, h_2\}} |E_{\text{in}}(h) - E_{\text{out}}(h)| > \epsilon]$ in extended Hoeffding's inequality. Compare with the probability you just calculated and briefly state your findings.

2.3 Multiple Coins and Multiple Bins

- (1) (7%) Do Exercise 1.10(a) of LFD.
- (2) (7%) Do Exercise 1.10(b) of LFD.
- (3) (7%) Do Exercise 1.10(c) of LFD.
- (4) (7%) Do Exercise 1.10(d) of LFD.
- (5) (7%) Do Exercise 1.10(e) of LFD.

2.4 Probably Approximately Correct

Read the derivation that links Equation (1.6) to Equation (2.1) on Page 40 of LFD. Then,

- (1) (5%) Do Problem 2.1(a) of LFD.
- (2) (5%) Do Problem 2.1(b) of LFD.
- (3) (5%) Do Problem 2.1(c) of LFD.

The title of this problem, *Probably Approximately Correct*, states what we can interpret from (2.1) if we have enough training examples. “Probably” means the statement is true with a high probability ($\geq 1 - \delta$). “Approximately” means that every $E_{\text{out}}(g)$ is close to $E_{\text{in}}(g)$ (within ϵ). “Correct” means that we can guarantee $E_{\text{out}}(g)$ to be small (by getting some hypothesis g with small $E_{\text{in}}(g)$).

2.5 Experiments with Another Perceptron Algorithm (*)

- (1) (10%) Do Problem 1.5(a) of LFD.
- (2) (10%) Do Problem 1.5(b) of LFD.
- (3) (10%) Do Problem 1.5(c) of LFD.
- (4) (10%) Do Problem 1.5(d) of LFD.
- (5) (10%) Do Problem 1.5(e) of LFD.

2.6 Friend or Foe

- (1) (Bonus 5%) Do Exercise 1.12 of LFD. You need to provide a convincing reason to get the points.
- (2) (Bonus 5%) If we replace your “friend” in the exercise with your “foe”, you need to be extremely careful in the guarantee you can provide. Do you see any traps in the setup and/or the choices? Write down your own version of the guarantee that you'd provide to your foe, and state the convincing reasons behind your own version.