Homework #3
RELEASE DATE: 10/11/2012

DUE DATE: 10/25/2012, 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.

3.1 Perceptron Dimension versus VC Dimension

(1) (10%)   Do Exercise 2.4(a) of LFD.

(2) (10%)   Do Exercise 2.4(b) of LFD.

3.2 Infinite VC Dimension

(1) (10%)   Do Problem 2.18 of LFD.

3.3 The Upper Bounds

(1) (10%)   Do Problem 2.5 of LFD.

(2) (10%)   Do Problem 2.6(a) of LFD.
(*For those who are motivated, you are encouraged to solve the whole Problem 2.6, of course. :-) *)

(3) (10%)   Do Problem 2.7 of LFD.

3.4 More on Growth Functions

(1) (20%)   There is a possible bug (or not?) in Problem 2.9 of LFD. Fix the bug by writing down the correct growth function for perceptrons in d dimensions and prove the correctness (or if you believe the original growth function is correct, prove it!).

(2) (10%)   Do Problem 2.10 of LFD.

3.5 Test Set

Note that we only briefly mentioned the test set in the class. So it is recommended that you read Section 2.2.3 of LFD in order to finish this problem.

(1) (10%)   Do Exercise 2.6(a) of LFD.

(2) (10%)   Do Exercise 2.6(b) of LFD.

3.6   Decision Stump Learning (*)

In class, we taught about the learning model of "positive and negative rays" (which is simply one-dimensional perceptron) for one-dimensional data. The model contains hypotheses of the form:

$$h_{s,\theta}(x) = s \cdot \text{sign}(x - \theta).$$

The model is frequently named the "decision stump" model and is one of the simplest learning models. As shown in class, for one-dimensional data, the VC dimension of the decision stump model is 2.

(1) (10%)   Write down the pseudo code of an $O(N^2)$-time deterministic algorithm that minimizes E_{in} for the decision stump model. Briefly explain why your algorithm runs in $O(N^2)$-time. If you can design a faster algorithm, write it down and the TAs would give you extra credit. *(Hint: Effectively, you only need to consider a finite number of choices for θ.)*

(2) (10%)   Define $P(x, y)$ as follows:

   (a) Generate x by a uniform distribution in $[-1, 1]$.
   (b) Generate y by $f(x)$ + noise where $f(x) = \text{sign}(x)$ and the noise flips the result with 20% probability.

   For any decision stump $h_{s,\theta}$ with $\theta \in [-1, 1]$, derive $E_{\text{out}}(h_{s,\theta})$ as a function of θ.

(3) (10%)   Implement the one-dimensional decision stump algorithm. Generate a data set of size 20 by the procedure above and run your algorithm on the data set. Record E_{in} and compute E_{out} with the formula above. Repeat the experiment 5000 times and plot a scatter plot

   `http://en.wikipedia.org/wiki/Scatter_plot`

   of E_{in} versus E_{out}.

(4) (20%)   Decision stumps can also work for multi-dimensional data. In particular, each decision stump now deals with a specific dimension i, as shown below.

$$h_{s,i,\theta}(\mathbf{x}) = s \cdot \text{sign}((\mathbf{x})_i - \theta).$$

   Implement the following decision stump algorithm for multi-dimensional data:

   (a) for each dimension $i = 1, 2, \cdots, d$, find the best decision stump $h_{s,i,\theta}$ using the one-dimensional decision stump algorithm that you have just implemented.
   (b) return the "best of best" decision stump in terms of E_{in}, with ties arbitrarily broken.

   Run the algorithm on the following set for training:

   `http://www.csie.ntu.edu.tw/~htlin/course/ml12fall/hw3/hw3_train.dat`

   Then, use the returned decision stump to predict the label of each example within the following test set:

   `http://www.csie.ntu.edu.tw/~htlin/course/ml12fall/hw3/hw3_test.dat`

   Report the best decision stump $h_{s,i,\theta}$ (i.e. list the s, i, θ found by the algorithm) and its E_{in} and E_{out} (actually, E_{test}).

3.7   Pocket Algorithm (*)

Do Exercise 3.2 of LFD.

(1) (10%)   Please generate a data set of size 100 as directed by the exercise, and generate a test set of size 10000 by the same distribution. Then, implement the pocket algorithm and run it on the data set for 2000 updates. Record $E_{\text{in}}(\mathbf{w}(t))$, $E_{\text{in}}(\mathbf{w}^*(t))$, $E_{\text{out}}(\mathbf{w}(t))$, and $E_{\text{out}}(\mathbf{w}^*(t))$ as functions of t (where E_{out} is estimated by E_{test}). Repeat the experiment for 20 times.

Plot the average $E_{\text{in}}(\mathbf{w}(t))$ and $E_{\text{in}}(\mathbf{w}^*(t))$ as functions of t and briefly state your findings.

(2) (10%)   For the same process above, plot the average $E_{\text{out}}(\mathbf{w}(t))$ and $E_{\text{out}}(\mathbf{w}^*(t))$ as functions of t and briefly state your findings.

(*Hint: For the two sub-problems above, checking Figure 3.2 of LFD may help you understand the intention.*)

(3) (20%)   Run the pocket algorithm on the following set for training:

$$\texttt{http://www.csie.ntu.edu.tw/\~htlin/course/ml12fall/hw3/hw3\_train.dat}$$

Then, use the returned perceptron to predict the label of each example within the following test set:

$$\texttt{http://www.csie.ntu.edu.tw/\~htlin/course/ml12fall/hw3/hw3\_test.dat}$$

Report the returned perceptron $h_{\mathbf{w}}$ (i.e. list the $\mathbf{w}$ found by the algorithm) and its E_{in} and E_{out} (actually, E_{test}). Compare the results with what you get in Problem 3.6(4) and briefly state your findings.

3.8   Mysterious B-function Leads to Bonus

Recall that we proved $B(N,n) \le \sum_{i=0}^{n-1} \binom{N}{i}$ in class.

(1) (Bonus 10%) Do Problem 2.4 of LFD.

Thus, $B(N,n) = \sum_{i=0}^{n-1} \binom{N}{i}$.