Homework #3
TA in charge: Yao-Nan Chen

RELEASE DATE: 10/11/2010

DUE DATE: 10/25/2010, 4:00 pm IN CLASS

TA SESSION: 10/21/2010, 6:00 pm IN R110

Unless granted by the instructor in advance, you must turn in a hard 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   Growth Function and VC Dimension

By "verify" in Exercise 2.3 of LFD, we mean "mathematically verify." That is, you need to take one n that satisfy the condition in Theorem 2.3 and convincingly tell us why equation (2.6) is true for all N.

(1) (10%)   Do Exercise 2.2-2, 2.3-2, and 2.4-2 of LFD.

(2) (10%)   Do Exercise 2.2-3, 2.3-3, and 2.4-3 of LFD.

(3) (5%)   Do Exercise 2.5-1 of LFD.

(4) (5%)   Do Exercise 2.5-2 of LFD.

(5) (5%)   Do Exercise 2.5-3 of LFD.

(6) (5%)   Do Exercise 2.5-4 of LFD.

(7) (10%)   In class, we mentioned that the VC dimension of the perceptron hypothesis set corresponds to the number of parameters $(w_0, w_1, \cdots, w_d)$ of the set, and the idea is "usually" true for other hypothesis sets. On the other hand, we will present a counter-example here. Prove that the following hypothesis set for $x \in \mathbb{R}$ is of an infinite VC dimension:

$$\mathcal{H} = \left\{ h_\alpha \colon h_\alpha(x) = (-1)^{\lfloor \alpha x \rfloor}, \text{ where } \alpha \in \mathbb{R} \text{ and } \lfloor A \rfloor = \max\{n \in \mathbb{Z}, n \le A\} \text{ is the floor function.} \right\}$$

This hypothesis set comes with only one parameter α but "enjoys" an infinite VC dimension.

3.2   Proof of the Upper Bound

In this problem, we will prove that $m_{\mathcal{H}}(N) \le \left(\frac{eN}{d_{\mathsf{VC}}} \right)^{d_{\mathsf{VC}}}$ for $N \ge d_{\mathsf{VC}} \ge 1$.

(1) (5%)   Prove that for $N \ge d \ge 1$, $\left(\frac{d}{N} \right)^d \sum_{i=0}^{d} \binom{N}{i} \le \left(1 + \frac{d}{N} \right)^N$.

(2) (5%)   Using the result above, prove that for $N \geq d \geq 1$, $\sum_{i=0}^{d} \binom{N}{i} \leq \left(\frac{eN}{d} \right)^{d}$.

(3) (5%)   Using the result above, argue that you have proved that $m_{\mathcal{H}}(N) \leq \left(\frac{eN}{d_{\mathsf{VC}}} \right)^{d_{\mathsf{VC}}}$
for $N \geq d_{\mathsf{VC}} \geq 1$.

3.3   Polynomial Hypotheses

In this problem, we will consider $\mathcal{X} = \mathbb{R}$. That is, $\mathbf{x} = x$ is a one-dimensional variable. For a hypothesis set $\mathcal{H} = \left\{ h_{\mathbf{c}} : h_{\mathbf{c}}(\mathbf{x}) = \text{sign} \left(\sum_{i=0}^{D} c_i x^i \right) \right\}$. We will prove that the VC dimension of $\mathcal{H}$ is exactly $(D+1)$ by showing that

(1) (10%)   There are $(D+1)$ points on which all 2^{D+1} label patterns can be produced from $\mathcal{H}$.

(2) (10%)   There are no $(D+2)$ points on which all 2^{D+2} label patterns can be produced from $\mathcal{H}$.

3.4   Pocket Algorithm (*)

Do exercise 3.2 of LFD.

(1) (5%)   Generate a data set of size 100 as directed by the exercise, and plot the examples $\{(\mathbf{x}_n, y_n)\}$ as well as the target function f on a plane. Be sure to mark the examples from different classes differently, and add labels to the axes of the plot. Generate a test set of size 1000 of the same nature.

Next, implement the pocket algorithm and run it on the data set for 1000 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 the test set). Repeat the experiment for 20 times.

(2) (5%)   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.

(3) (5%)   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.

3.5   Mysterious B-function Leads to Bonus

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

(1) (Bonus 5%)   Prove the following inequality: $B(N, n) \geq \sum_{i=0}^{n-1} \binom{N}{i}$.

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