Learning From Data
Yaser Abu-Mostafa, *Caltech*
http://work.caltech.edu/telecourse
Self-paced version

Homework # 3

*All questions have multiple-choice answers (***[a]***,* **[b]***,* **[c]***,* **...***). You can collaborate with others, but do not discuss the selected or excluded choices in the answers. You can consult books and notes, but not other people's solutions. Your solutions should be based on your own work. Definitions and notation follow the lectures.*

Note about the homework

- The goal of the homework is to facilitate a deeper understanding of the course material. The questions are not designed to be puzzles with catchy answers. They are meant to make you roll up your sleeves, face uncertainties, and approach the problem from different angles.
- The problems range from easy to difficult, and from practical to theoretical. Some problems require running a full experiment to arrive at the answer.
- The answer may not be obvious or numerically close to one of the choices, but one (and only one) choice will be correct if you follow the instructions precisely in each problem. You are encouraged to explore the problem further by experimenting with variations on these instructions, for the learning benefit.
- You are also encouraged to take part in the forum

  http://book.caltech.edu/bookforum

  where there are many threads about each homework set. We hope that you will contribute to the discussion as well. Please follow the forum guidelines for posting answers (see the "BEFORE posting answers" announcement at the top there).

- **Generalization Error**

1. The modified Hoeffding Inequality provides a way to characterize the generalization error with a probabilistic bound

$$\mathbb{P}\left[|E_{in}(g) - E_{out}(g)| > \epsilon\right] \leq 2Me^{-2\epsilon^2 N}$$

for any $\epsilon > 0$. If we set $\epsilon = 0.05$ and want the probability bound $2Me^{-2\epsilon^2 N}$ to be at most 0.03, what is the least number of examples N (among the given choices) needed for the case $M = 1$?

[**a**] 500

[**b**] 1000

[**c**] 1500

[**d**] 2000

[**e**] More examples are needed.

2. Repeat for the case $M = 10$.

[**a**] 500

[**b**] 1000

[**c**] 1500

[**d**] 2000

[**e**] More examples are needed.

3. Repeat for the case $M = 100$.

[**a**] 500

[**b**] 1000

[**c**] 1500

[**d**] 2000

[**e**] More examples are needed.

- **Break Point**

4. As shown in class, the (smallest) break point for the Perceptron Model in the two-dimensional case ($\mathbb{R}^2$) is 4 points. What is the smallest break point for the Perceptron Model in $\mathbb{R}^3$? (i.e., instead of the hypothesis set consisting of separating lines, it consists of separating planes.)

[**a**] 4

[**b**] 5

[**c**] 6

[**d**] 7

[**e**] 8

- **Growth Function**

5. Which of the following are possible formulas for a growth function $m_{\mathcal{H}}(N)$:

$$\begin{array}{ll} \text{i) } 1+N & \text{iv) } 2^{\lfloor N/2 \rfloor} \\ \text{ii) } 1+N+\binom{N}{2} & \text{v) } 2^N \\ \text{iii) } \sum_{i=1}^{\lfloor \sqrt{N} \rfloor} \binom{N}{i} & \end{array}$$

where $\lfloor u \rfloor$ is the biggest integer $\leq u$, and $\binom{M}{m} = 0$ when $m > M$.

[**a**] i, v

[**b**] i, ii, v

[**c**] i, iv, v

[**d**] i, ii, iii, v

[**e**] i, ii, iii, iv, v

- **Fun with Intervals**

6. Consider the "2-intervals" learning model, where h: $\mathbb{R} \rightarrow \{-1, +1\}$ and $h(x) = +1$ if the point is within either of two arbitrarily chosen intervals and -1 otherwise. What is the (smallest) break point for this hypothesis set?

[**a**] 3

[**b**] 4

[**c**] 5

[**d**] 6

[**e**] 7

7. Which of the following is the growth function $m_H(N)$ for the "2-intervals" hypothesis set?

[a] $\binom{N+1}{4}$

[b] $\binom{N+1}{2}+1$

[c] $\binom{N+1}{4}+\binom{N+1}{2}+1$

[d] $\binom{N+1}{4}+\binom{N+1}{3}+\binom{N+1}{2}+\binom{N+1}{1}+1$

[e] None of the above

8. Now, consider the general case: the "M-intervals" learning model. Again $h : \mathbb{R} \to \{-1,+1\}$, where $h(x) = +1$ if the point falls inside any of M arbitrarily chosen intervals, otherwise $h(x) = -1$. What is the (smallest) break point of this hypothesis set?

[a] M

[b] $M+1$

[c] M^2

[d] $2M+1$

[e] $2M-1$

• Convex Sets: The Triangle

9. Consider the "triangle" learning model, where $h : \mathbb{R}^2 \to \{-1,+1\}$ and $h(\mathbf{x}) = +1$ if $\mathbf{x}$ lies within an arbitrarily chosen triangle in the plane and -1 otherwise. Which is the largest number of points in $\mathbb{R}^2$ (among the given choices) that can be shattered by this hypothesis set?

[a] 1

[b] 3

[c] 5

[d] 7

[e] 9

• Non-Convex Sets: Concentric Circles

10. Compute the growth function $m_{\mathcal{H}}(N)$ for the learning model made up of two concentric circles in $\mathbb{R}^2$. Specifically, $\mathcal{H}$ contains the functions which are $+1$ for

$$a^2 \leq x_1^2 + x_2^2 \leq b^2$$

and -1 otherwise. The growth function is

[a] $N + 1$

[b] $\binom{N+1}{2} + 1$

[c] $\binom{N+1}{3} + 1$

[d] $2N^2 + 1$

[e] None of the above