CaltechX: Learning From Data: Homework #3

Due on Monday, October 16, 2017

Yaser Abu-Mostafa

Andrew Watson

Contents

Generalization Error

Problem 1

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

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

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

To solve this, we can just plug in the values we are given, and then solve for N:

$$\begin{aligned}\mathbb{P}[|E_{\text{in}}(g) - E_{\text{out}}(g)| > \epsilon] &\leq 2Me^{-2\epsilon^2 N} \\ &= 2e^{-2(0.05)^2 N} = 2e^{-0.005N} \leq 0.03 \\ &\implies N \geq 840.\end{aligned}$$

Therefore [**b**] is the correct answer.

Problem 2

Repeat for the case $M = 10$.

[**a**] 500

[**b**] 1000

[**c**] 1500

[**d**] 2000

[**e**] More examples are needed

For any value of M (but still with $\epsilon = 0.05$), we require that $N \geq -\frac{\ln(0.015/M)}{0.005}$. By plugging $M = 10$ into this formula, we see that now the requirement is $N \geq 1301$, and therefore the answer is [**c**].

Problem 3

Repeat for the case $M = 100$.

[**a**] 500

[**b**] 1000

[**c**] 1500

[**d**] 2000

[**e**] More examples are needed

Finally, plugging $M = 100$ into the formula above, we find that $N \geq 1761$. Therefore the answer is [**d**].

Break Point

Problem 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

Problem 5

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

i) $1 + N$

ii) $\binom{1+N+N}{2}$

iii) $\binom{\sum_{i=1}^{\lfloor\sqrt{N}\rfloor} N}{i}$

iv) $2^{\lfloor N/2 \rfloor}$

v) 2^N

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

Problem 6

Consider the "2-intervals" learning model, where $h : \mathbb{R} \to \{-1, +1\}$ 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

Problem 7

Which of the following is the growth function $m_{\mathcal{H}}(N)$ 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

Problem 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

Problem 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

Problem 10

Compute the growth function $m_{\mathcal{H}}(N)$ for the learning model made up of two concentric circles around the origin 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, where a and b are the model parameters. 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