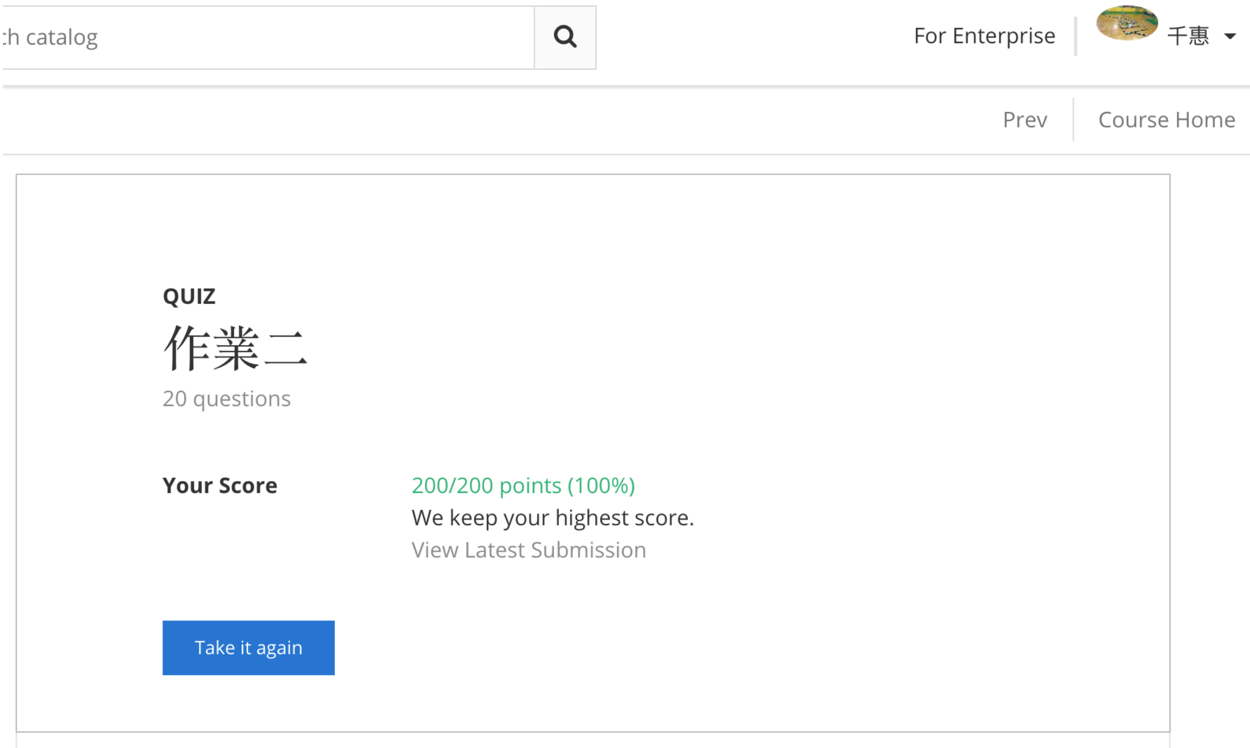
Homework #2

資工三 B04902009 蕭千惠

9th December, 2017

Problem 1

Score: 200 / 200

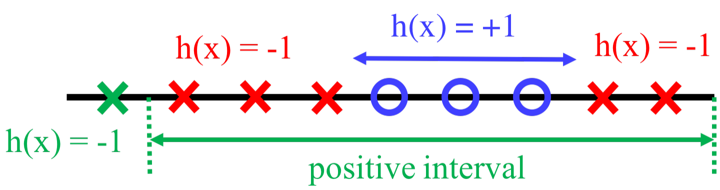


Problem 2

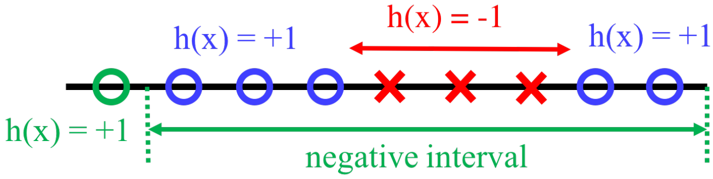
The growth function $m_{\mathcal{H}}(N)$ of positive or negative intervals on $\mathbb{R} = \binom{N+1}{2} + 1 = \frac{1}{2}N^2 + \frac{1}{2}N + 1$

We can devide positive-and-negative intervals into two categories:

1. The intervals with a leftmost 'x' and a positive interval $\Rightarrow m_{\mathcal{H}}(N) = \binom{N}{2} + 1$.



2. The intervals with a leftmost 'o' and a negative interval $\Rightarrow m_{\mathcal{H}}(N) = \binom{N}{2} + 1$.



The growth function $m_{\mathcal{H}}(N)$ of positive-and-negative intervals on $\mathbb{R} = 2*(\binom{N}{2}+1) = N^2 - N + 2$.

Problem 3

$\mathcal{H} = \left\{h_c \middle| h_c(x) = sign\left(\sum_{i=0}^{\mathcal{D}} c_i x^i\right)\right\}$

Let $g(x) = \sum_{i=0}^{\mathcal{D}} c_i x^i$ and $a_1, a_2, ..., a_{\mathcal{D}-1}, a_{\mathcal{D}}$ be the root of $g(x) \Rightarrow g(x) = \prod_{i=1}^{D}(x - a_i)$

g has $\mathcal{D} + 1$ intervals: $(-\infty, a_1], (a_1, a_2], ..., (a_{\mathcal{D}-1}, a_{\mathcal{D}}), (a_{\mathcal{D}}, \infty)$ and g alternates sign in the successive intervals.

If $\mathcal{D} = 1$, $\mathcal{H}$ is positive-and-negative rays $\Rightarrow d_{vc}(\mathcal{H}) = 2$

If $\mathcal{D} = 2$, $\mathcal{H}$ is positive-and-negative intervals $\Rightarrow d_{vc}(\mathcal{H}) = 3$

Let $\mathcal{H}_i$ be the $\mathcal{H}$ with $\mathcal{D} = i$

From the same way described in Problem 2, we conclude that $m_{\mathcal{H}_i}(N) = 2m_{\mathcal{H}_{i-1}}(N)$

$\Rightarrow d_{vc}(\mathcal{H}_i) = d_{vc}(\mathcal{H}_{i-1}) + 1$

$\Rightarrow d_{vc}(\mathcal{H}_{\mathcal{D}}) = d_{vc}(\mathcal{H}_1) + (\mathcal{D} - 1) = 2 + (\mathcal{D} - 1) = \mathcal{D} + 1$

$\Rightarrow$ The VC-demension of $\mathcal{H}$ is $\mathcal{D} + 1$.

Problem 4

$\mathcal{H} = \{h_\alpha | h_\alpha(x) = sign(|(\alpha x) \mod 4 - 2| - 1), \alpha \in \mathbb{R}\}$

Consider N point on $\mathbb{R}$, from 1st point to Nth point, i^{th} point on 4^i

$\Rightarrow \mathcal{X} = \{4^1, 4^2, ..., 4^{N-1}, 4^N\}$

Let $\mathcal{Y}$ be a set family containing all $\{+1, -1\}^N$ combinations.

When given an α, let's denote the set$\{h_\alpha(x_1), h_\alpha(x_2), ..., h_\alpha(N)\}$ by $y = \{y_1, y_2, ..., y_N\}$

$\forall y \in \mathcal{Y}$, we can find an α to construct it by the pseudocode below.

find_alpha(){

$\alpha = 1$

for($i = 1$; $i \leq N$; $i++$)

if($y_i == -1$)

$\alpha = \alpha + 2/\text{pow}(4, \text{i})$

return α

}

When $\alpha = 1$, $\forall x \in \mathcal{X}, h_\alpha(x) = +1$

When flipping x_i from +1 to -1, we add $\frac{2}{4^i}$ to α.

To proof that adding $\frac{2}{4^i}$ to α will only affect the point x_i, we divided all $x \in \mathcal{X}$ into three cases:

1. $\boldsymbol{x_j}$ **with** $\boldsymbol{1 \leq j < i}$

   We have to proof that flipping all points larger than x_j won't affect the value of y_j.

   To flip all points larger than x_j, $\alpha = 1 + \sum_{k=j+1}^{n} \frac{2}{4^k} = 1 + \frac{2}{3 * 4^j}[1 - (\frac{1}{4})^{n-j+1}]$

   $\alpha x_j = \{1 + \frac{2}{3 * 4^j}[1 - (\frac{1}{4})^{n-j+1}]\} * 4^j = 4^j + \frac{2}{3}[1 - (\frac{1}{4})^{n-j+1}] < 4^j + \frac{2}{3}$

   $\Rightarrow sign(|(\alpha x_j) \mod 4 - 2| - 1) = +1$

   $\Rightarrow y_j$ remains the same.

2. $\boldsymbol{x_i}$

   $\frac{2}{4^i} x_i = \frac{2}{4^i} * 4^i = 2$

   $\Rightarrow y_i = sign(|(\frac{2}{4^i} x_i) \mod 4 - 2| - 1) = sign(|2 \mod 4 - 2| - 1) = -1$

3. $\boldsymbol{x_j}$ **with** $\boldsymbol{i < j \leq N}$

   $x_j = x_i * 4^{j-i}$

   $(\frac{2}{4^i} * x_j) \mod 4 = (\frac{2}{4^i} * x_i * 4^{j-i}) \mod 4 = 0$

   $\Rightarrow y_j$ remains the same.

Since all $\{+1, -1\}^N$ combinations can be constructed, $\mathcal{H}$ can shatter any N.

$\Rightarrow$ The VC-demension of $\mathcal{H}$ is ∞.

Problem 5

1. $\forall N < d_{vc}(\mathcal{H}_1)$, any N inputs can be shattered by $\mathcal{H}_1$.

   Since $\mathcal{H}_1 \subseteq \mathcal{H}_2$, any N inputs can be shattered by $\mathcal{H}_2 \Rightarrow d_{vc}(\mathcal{H}_1) \not> d_{vc}(\mathcal{H}_2)$

2. When $\mathcal{H}_1 = \mathcal{H}_2, d_{vc}(\mathcal{H}_1) = d_{vc}(\mathcal{H}_2)$

3. Let $\mathcal{H}_1$ be positive interval hypothesis $\Rightarrow d_{vc}(\mathcal{H}_1) = 2$ (Known from the course slide.)

   Let $\mathcal{H}_2$ be positive-and-negative intervals hypothesis.

   $\Rightarrow m_{\mathcal{H}_2}(N) = N^2 - N + 2$ (Known from Problem 2) $\Rightarrow d_{vc}(\mathcal{H}_2) = 3$

   $\mathcal{H}_1 \subseteq \mathcal{H}_2$ and $d_{vc}(\mathcal{H}_1) < d_{vc}(\mathcal{H}_2)$

From 1 & 2 & 3 $\Rightarrow d_{vc}(\mathcal{H}_1) \leq d_{vc}(\mathcal{H}_2)$

Problem 6

$\mathcal{H}1 \cup \mathcal{H}2 =$ the positive-and-negative ray set.

The growth function $m_{\mathcal{H}1\cup\mathcal{H}2}(N)$ of positive-and-negative ray on $\mathbb{R} = 2(N+1) - 2 = 2N$.

$m_{\mathcal{H}1\cup\mathcal{H}2}(N) = 2N \neq 2^N$ when $N > 3$

$\Rightarrow$ The VC-demension of $\mathcal{H}1 \cup \mathcal{H}2$ is 2.

Problem 7

$$P(y|\mathbf{x}) = \begin{cases} 0.8, y = f(\mathbf{x}) \\ \\ 0.2, y \neq f(\mathbf{x}) \end{cases}$$

Let $h_{s,\theta}$ makes an error with probability μ. $(P(h_{s,\theta}(\mathbf{x}) \neq f(\mathbf{x})) = \mu)$

μ	$\theta > 0$	$\theta < 0$
s = 1	$\frac{\theta}{2}$	$\frac{\lvert\theta\rvert}{2}$
s = -1	$1 - \frac{\theta}{2}$	$1 - \frac{\theta}{2}$

$\Rightarrow$ When s = 1, $\mu = \frac{|\theta|}{2}$. When s = -1, $\mu = 1 - \frac{|\theta|}{2}$

$\mu = \frac{s+1}{2}\frac{|\theta|}{2} + \frac{1-s}{2}(1 - \frac{|\theta|}{2}) = \frac{1 - s + s|\theta|}{2}$

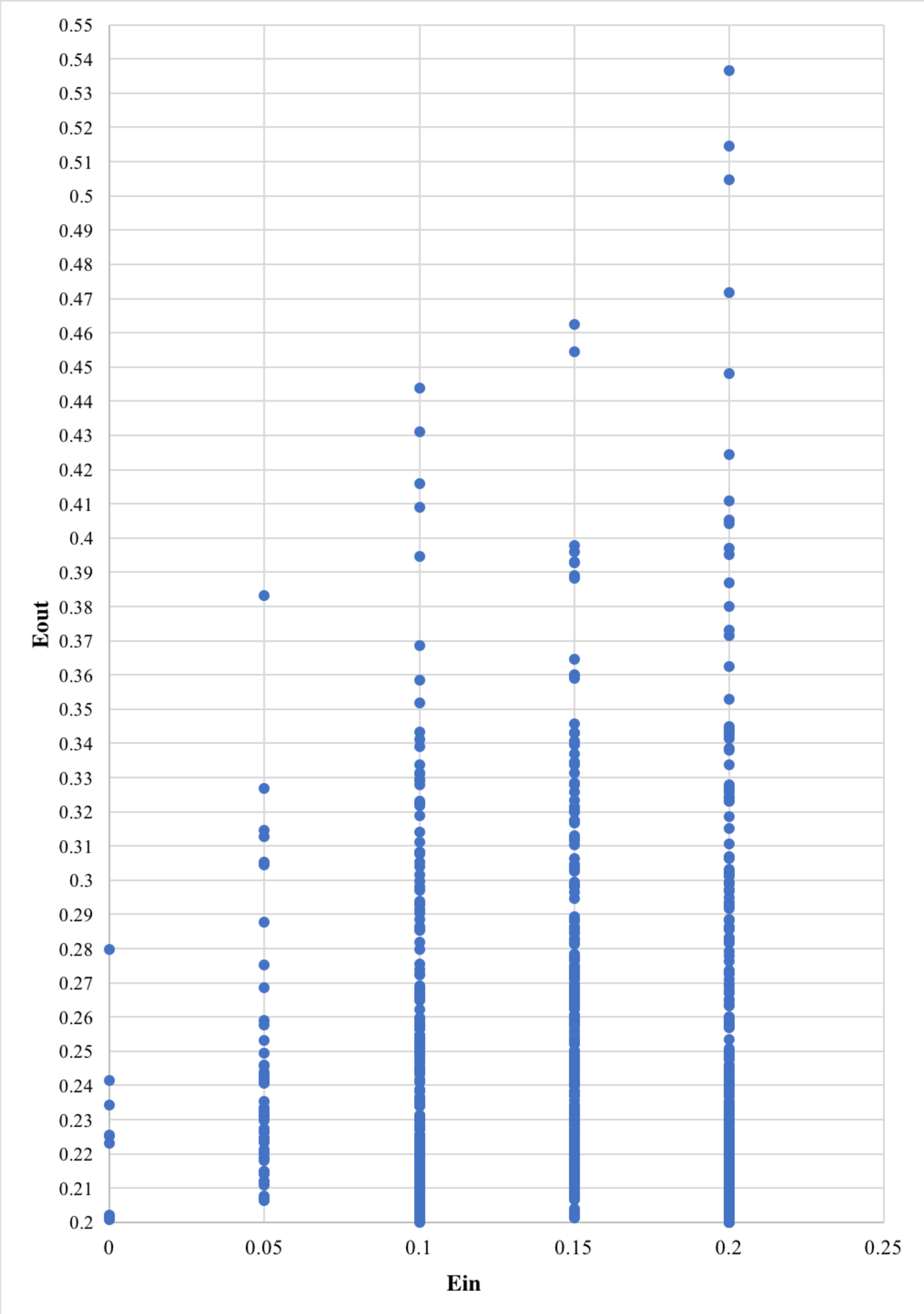
$E_{out}(h_{s,\theta}) = 0.8\mu + 0.2(1-\mu) = 0.5 + 0.3s(|\theta| - 1)$

Problem 8

Findings:

1. The smaller the E_{in} is, the smaller the maximum E_{out} is.
2. $P(E_{in} = 0.2) = 0.41$
   $\Rightarrow E_{in} = 0.2$ most frequently occured since $P(y|\mathbf{x}) = 0.2$ when $y \neq f(\mathbf{x})$.

Averaged $E_{in} = 0.150450$, Averaged $E_{out} = 0.243626$


Eout
0.55
0.54
0.53
0.52
0.51
0.5
0.49
0.48
0.47
0.46
0.45
0.44
0.43
0.42
0.41
0.4
0.39
0.38
0.37
0.36
0.35
0.34
0.33
0.32
0.31
0.3
0.29
0.28
0.27
0.26
0.25
0.24
0.23
0.22
0.21
0.2
0
0.05
0.1
0.15
0.2
0.25
Ein

Problem 9(Bonus)

Let's denote the growth function for perceptron learning model in d dimensions with N points by $m_{\mathcal{H},d}(N)$.

We will find an expression for $m_{\mathcal{H},d}(N)$ through induction.

First imaging having N-1 points and then we add one more point.

The linearly separable partitions of the previous $N-1$ points can be separated into two cases:

1. **There is a separating hyperplane for the previous $N-1$ points passing through the new point:** each such linearly separable partition of the previous $N-1$ points gives rise to two distinct linearly separable partitions when the new point is added as the hyperplane can be shifted infinitesimally to place the new point in either class. (Let's denotes the number of linearly separable partitions of the previous $N-1$ points in this case by P_1.)

2. **There is no separating hyperplane passing through the new point:** each such linearly separable partition gives rise to only one linearly separable partition when the new point is added. (Let's denotes the number of linearly separable partitions of the previous $N-1$ points in this case by P_2.)

$m_{\mathcal{H},d}(N) = 2P_1 + P_2 = (P_1 + P_2) + P_1 = m_{\mathcal{H},d}(N-1) + P_1$

Restricting the separating hyperplane to go through a particular point is the same as eliminating one degree of freedom and thus projecting the previous $N-1$ points to a $(d-1)$-dimensional space. Therefore, we can know that $P_1 = m_{\mathcal{H},d-1}(N-1)$

$\Rightarrow m_{\mathcal{H},d}(N) = m_{\mathcal{H},d}(N-1) + m_{\mathcal{H},d-1}(N-1)$

By recursive: $m_{\mathcal{H},d}(N) = \binom{N-1}{0} m_{\mathcal{H},d}(1) + \binom{N-1}{1} m_{\mathcal{H},d-1}(1) + \ldots + \binom{N-1}{N-1} m_{\mathcal{H},d-N+1}(1)$

$m_{\mathcal{H},d}(1) = 0$ when $d < 0$, otherwise $m_{\mathcal{H},d}(1) = 1$

$$\begin{cases} \forall d < N-1, m_{\mathcal{H},d}(N) = \binom{N-1}{0} m_{\mathcal{H},d}(1) + \binom{N-1}{1} m_{\mathcal{H},d-1}(1) + \ldots + \binom{N-1}{d} m_{\mathcal{H},0}(1) = 2\sum_{i=0}^{d}\binom{N-1}{i} \\ \forall d \geq N-1, m_{\mathcal{H},d}(N) = 2\sum_{i=0}^{N-1}\binom{N-1}{i} = 2\sum_{i=0}^{d}\binom{N-1}{i} \end{cases}$$

(Since $\binom{\mathbf{a}}{\mathbf{b}} = \mathbf{0}$ when $\mathbf{a} < \mathbf{b} \Rightarrow \binom{\mathbf{N-1}}{\mathbf{i}} = \mathbf{0}$ when $N \leq i \leq d$)

$\Rightarrow m_{\mathcal{H},d}(N) = 2\sum_{i=0}^{d}\binom{N-1}{i}$