Machine Learning Homework 2

B04705003 Tzu-Chuan, Lin

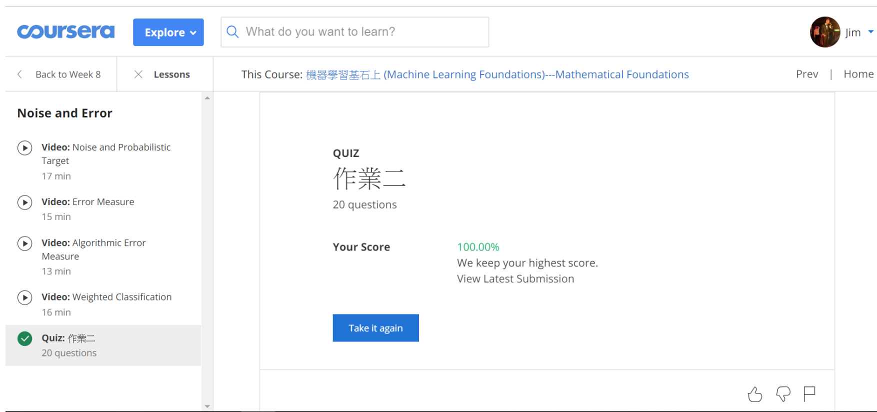
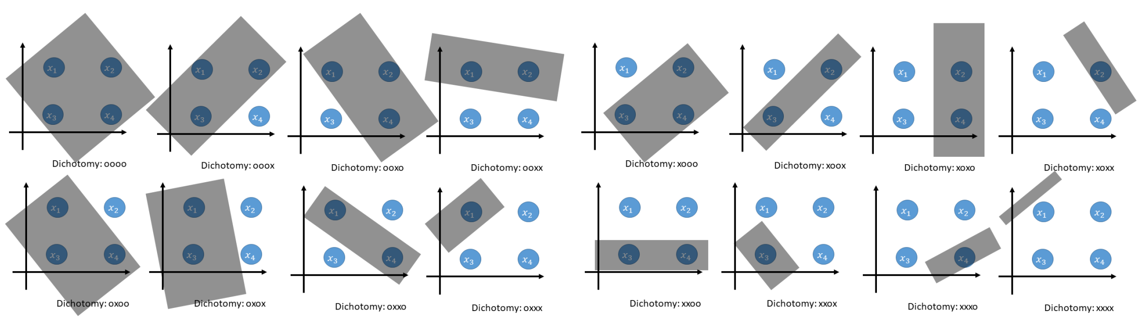


Figure 1: Coursera Quiz 2

Problem 1. I get 100. See Figure 1.



(a) Dichotomies Part 1 (b) Dichotomy Part 2

Figure 2: Dichotomy Illustration

Problem 2. See Figure 2. You can find that this negative thick line hypothesis set can shatter 4 data points. Therefore, its d_{vc} should be no less than 4.

Problem 3. The VC-Dimension of this set $\mathcal{H}$ is ∞.

Proof. **Method 1**: Set the data points to be:

$$X = \{x_1 = 2^1, x_2 = 2^2, x_3 = 2^3, ..., x_N = 2^N\}$$

Let $\lambda = \frac{1}{\alpha}$. Intuitively, if we change λ such that $\frac{x_1}{\lambda}$ move one step while $\frac{x_N}{\lambda}$ will move 2^{N-1} steps. We can construct a method so that every dichotomy of X will be produced by setting:

$$\frac{2}{4+\frac{i-1}{2^{N-1}}} > \lambda_i > \frac{2}{4+\frac{i}{2^{N-1}}}$$

where $i = 1, 2, ..., 2^N$. To confirm our construction is right, we will use mathematical induction to prove it.

Base case: $N = 1$: Let $i = 1$, we get:

$$\lambda_1 \in (\frac{2}{4+1}, \frac{2}{4}), \lambda_2 \in (\frac{2}{4+2}, \frac{2}{4+1})$$

and

$$i = 1 \Rightarrow h_\alpha(x_1 = 2^1) = sign(|\frac{2^1}{\lambda_1} \mod 4 - 2| - 1) > 0$$
$$i = 2 \Rightarrow h_\alpha(x_1 = 2^1) = sign(|\frac{2^1}{\lambda_2} \mod 4 - 2| - 1) < 0$$

Thus, $N = 1$ can be shattered.

Induction hypothesis: Assume $N = p$, our construction can shatter p data points, which can be formally written as:

$$\forall 1 \le i \le 2^p, \frac{2}{4+\frac{i-1}{2^{p-1}}} > \lambda_i > \frac{2}{4+\frac{i}{2^{p-1}}}$$

and

$$2^{k-1} \cdot 4 + \frac{i-1}{2^{p-k}} < \frac{x_k = 2^k}{\lambda_i} < 2^{k-1} \cdot 4 + \frac{i}{2^{p-k}} \tag{1}$$

Induction step: Let $N = p + 1$, we want to verify if

$$\forall 1 \le i \le 2^{p+1}, \frac{2}{4+\frac{i-1}{2^p}} > \lambda_i > \frac{2}{4+\frac{i}{2^p}} \tag{2}$$

can shatter $p + 1$ data points by enumerating i from 1 to 2^{p+1}.

When $i = 2, 4, ..., 2^{p+1}$, we can rewrite above equation (2)(by setting $i = 2i'$) to be:

$$\forall 1 \le i' \le 2^p, \frac{2}{4+\frac{2i'-1}{2^p}} > \lambda_{i'} > \frac{2}{4+\frac{2i'}{2^p}}$$

and

$$2^{k-1} \cdot 4 + \frac{i' - \frac{1}{2}}{2^{p-k}} < \frac{x_k = 2^k}{\lambda_{i'}} < 2^{k-1} \cdot 4 + \frac{i'}{2^{p-k}} \tag{3}$$

We can easily find out that Inequality (3) is in the range of Inequality (1) so after enumerating index i' from 1 to 2^p, we can shatter $\{x_1, x_2, ..., x_p\}$(by Induction Hypothesis).

When $i = 1, 3, ..., 2^{p+1} - 1$, we can rewrite Equation (2)(by setting $i = 2i' - 1$) into:

$$\forall 1 \le i' \le 2^p, \frac{2}{4+\frac{2i'-2}{2^p}} > \lambda_{i'} > \frac{2}{4+\frac{2i'-1}{2^p}}$$

and

$$2^{k-1} \cdot 4 + \frac{i'-1}{2^{p-k}} < \frac{x_k = 2^k}{\lambda_{i'}} < 2^{k-1} \cdot 4 + \frac{i' - \frac{1}{2}}{2^{p-k}} \tag{4}$$

We can also easily find out that Inequality (4) is in the range of Inequality (1) too. Thus, after enumerating index i' from 1 to 2^p, we can shatter $\{x_1, x_2, ..., x_p\}$(by Induction Hypothesis). **Moreover**, its dichotomy order will be same as $i = 2, 4, ..., 2^{p+1}$. Rigorously, it means dichotomy of $\{x_1, x_2, ..., x_p\}$ are equal between $i = 2k - 1$ and $i = 2k$ where $1 \leq k \leq 2^p$.

The last step is to check whether for each dichotomy of $\{x_1, x_2, ..., x_p\}$, we can produce two dichotomies of x_{p+1}. Intuitively, we can check whether the dichotomy on x_{p+1} of even $i = 2, 4, ..., 2^{p+1}$ and $i - 1$ are different. If we can verify they are different, then we are done.

Consider four cases for x_{p+1}:

(1) $i \mod 4 = 0$:

$$2^p \cdot 4 + \frac{i-1}{2^0} < \frac{x_{p+1} = 2^{p+1}}{\lambda_i} < 2^p \cdot 4 + \frac{i}{2^0}$$
$$\Rightarrow h_\alpha(x_{p+1}) = sign(|\bullet - 2| - 1) = +1$$

(2) $i \mod 4 = 1$:

$$2^p \cdot 4 + \frac{i-1}{2^0} < \frac{x_{p+1} = 2^{p+1}}{\lambda_i} < 2^p \cdot 4 + \frac{i}{2^0}$$
$$\Rightarrow h_\alpha(x_{p+1}) = sign(|\bullet - 2| - 1) = +1$$

(3) $i \mod 4 = 2$:

$$2^p \cdot 4 + \frac{i-1}{2^0} < \frac{x_{p+1} = 2^{p+1}}{\lambda_i} < 2^p \cdot 4 + \frac{i}{2^0}$$
$$\Rightarrow h_\alpha(x_{p+1}) = sign(|\bullet - 2| - 1) = -1$$

(4) $i \mod 4 = 3$:

$$2^p \cdot 4 + \frac{i-1}{2^0} < \frac{x_{p+1} = 2^{p+1}}{\lambda_i} < 2^p \cdot 4 + \frac{i}{2^0}$$
$$\Rightarrow h_\alpha(x_{p+1}) = sign(|\bullet - 2| - 1) = -1$$

By above 4 cases, we can find out that given an **even** index i, dichotomy on x_{p+1} are different between λ_i and λ_{i-1}. For example, if $i = 2$, λ_{2-1} will have +1 while λ_2 will have -1; if $i = 4$, λ_{4-1} will have -1 while λ_4 will have +1.

Method 2: Set the data points to be:

$$X = \{x_1 = 2^2, x_2 = 2^4, x_3 = 2^6, ..., x_N = 2^{2N}\}$$

Represent $\alpha > 0$ by binary number $\alpha_A ... \alpha_1 \alpha_0 . \beta_1 . \beta_2 ... \beta_B$. We can find out that $\alpha \cdot x_k$ is simply a right-shift operation by $2k$. Furthermore, a modulo operation will simply extract last two bits out. Therefore, if we want to set x_k to be +1, we can simply adjust β_{2k-1} and β_{2k} bits to 00. Note that adjust each pair of two bits will not intervene other x_k corresponding two bits - β_{2k-1} and β_{2k}. Consequently, we can shatter any data points with

$$X = \{x_1 = 2^2, x_2 = 2^4, x_3 = 2^6, ..., x_N = 2^{2N}\}$$

by adjusting α's bits.

$\square$

Problem 4.

Proof. Consider $\mathcal{H}_1 \cap \mathcal{H}_2$ can shatter $d_{vc}(\mathcal{H}_1 \cap \mathcal{H}_2)$ data points, we can extract $\mathcal{H}_1 \cap \mathcal{H}_2$ from $\mathcal{H}_2$ to shatter $d_{vc}(\mathcal{H}_1 \cap \mathcal{H}_2)$ data points. Because we are sure to shatter at least $d_{vc}(\mathcal{H}_1 \cap \mathcal{H}_2)$ data points by $\mathcal{H}_2$, we can conclude that $d_{vc}(\mathcal{H}_1 \cap \mathcal{H}_2) \leq d_{vc}(\mathcal{H}_2)$ $\square$

Problem 5. $m_{\mathcal{H}_1 \cup \mathcal{H}_2}(N) = 2N$
Let $N = 2$, $m_{\mathcal{H}_1 \cup \mathcal{H}_2}(2) = 4 = 2^2$.
Let $N = 3$, $m_{\mathcal{H}_1 \cup \mathcal{H}_2}(3) = 6 < 2^3$.
Thus, $d_{vc}(\mathcal{H}_1 \cup \mathcal{H}_2) = 2$

Problem 6.

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

$$\begin{aligned}
\mathbb{E}_{out}(h_{s,\theta}) &= \int_{-1}^{1} p(x) \cdot \mathbb{E}_{y \sim P(y|x)}[[\![h_{s,\theta}(x) \neq y]\!]]dx \\
&= \frac{1}{2}[\int_{0}^{1} \mathbb{E}_{y \sim P(y|x)}[[\![h_{s,\theta}(x) \neq y]\!]]dx + \int_{-1}^{0} \mathbb{E}_{y \sim P(y|x)}[[\![h_{s,\theta}(x) \neq y]\!]]dx] \\
&= \frac{1}{2}[\int_{0}^{1} 0.8[\![h_{s,\theta}(x) \neq 1]\!] + 0.2[\![h_{s,\theta}(x) \neq -1]\!]dx + \int_{-1}^{0} 0.2[\![h_{s,\theta}(x) \neq 1]\!] + 0.8[\![h_{s,\theta}(x) \neq -1]\!]dx]
\end{aligned}$$

- Assume $s = +$: $h_{s,\theta}(x) = sign(x - \theta)$
  - If $\theta \geq 0$:

$$\begin{aligned}
\mathbb{E}_{out}(h_{s,\theta}) &= \frac{1}{2}[\int_{0}^{\theta} 0.8dx + \int_{\theta}^{1} 0.2dx + \int_{-1}^{0} 0.2dx + 0] \\
&= \frac{1}{2}[0.8\theta + (1 - \theta)0.2 + 0.2] \\
&= 0.3\theta + 0.2
\end{aligned}$$

  - If $\theta < 0$:

$$\begin{aligned}
\mathbb{E}_{out}(h_{s,\theta}) &= \frac{1}{2}[0 + \int_{0}^{1} 0.2dx + \int_{-1}^{\theta} 0.2dx + \int_{\theta}^{0} 0.8dx] \\
&= \frac{1}{2}[0.2 + (\theta + 1)0.2 + (-\theta)0.8] \\
&= 0.2 - 0.3\theta
\end{aligned}$$

- Assume $s = -$: $h_{s,\theta}(x) = -sign(x - \theta)$
  - $\theta \geq 0$

$$\begin{aligned}\mathbb{E}_{out}(h_{s,\theta}) &= \frac{1}{2}[\int_{\theta}^{1} 0.8dx + \int_{0}^{\theta} 0.2dx + \int_{-1}^{0} 0.8dx] \\ &= \frac{1}{2}[(1-\theta)0.8 + \theta 0.2 + 0.8] \\ &= \frac{1}{2}[1.6 - 0.6\theta] \\ &= 0.8 - 0.3\theta\end{aligned}$$

  - $\theta \leq 0$:

$$\begin{aligned}\mathbb{E}_{out}(h_{s,\theta}) &= \frac{1}{2}[\int_{0}^{1} 0.8dx + + \int_{\theta}^{0} 0.2dx + \int_{-1}^{\theta} 0.8dx] \\ &= \frac{1}{2}[1.6 + 0.6\theta] \\ &= 0.3\theta + 0.8\end{aligned}$$

After gathering all situations, we can get:

$$\mathbb{E}_{out}(h_{s,\theta}) = 0.5 + 0.3s(|\theta| - 1)$$

Problem 7.

- Histogram: See Figure 3.

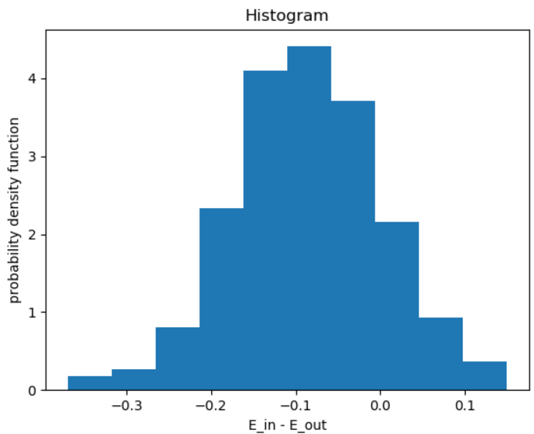


Figure 3: Histogram

- Findings: In most of time, you can see differences between E_{in} and E_{out} fall in the range of $[-0.2, 0.1]$. Therefore, we have high confidence that our $h_{s,\theta}$ found by decision stump algorithm have generalization capability.

Problem 8(Bonus).

Proof. Intuition: If we can construct a example so that this example's number of dichotomies is $\sum_{i=0}^{k-1}\binom{N}{i}$, it can imply $B(N,k) \geq \sum_{i=0}^{k-1}\binom{N}{i}$.

Construct a set S of length-N vector dichotomies which contains:

$$S = \{\text{dichotomy which contains } i \text{ o, where } 0 \leq i \leq k-1\}$$

You can easily find that in S, every length-k subvector is not shattered. Because every length-k subvector, "all o" dichotomy is missed. Furthermore, the number of S's elements is $\sum_{i=0}^{k-1}\binom{N}{i}$

In conclusion, we can say:

$$\sum_{i=0}^{k-1}\binom{N}{i} \leq B(N,k) \leq \sum_{i=0}^{k-1}\binom{N}{i}$$

which implies

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

$\square$