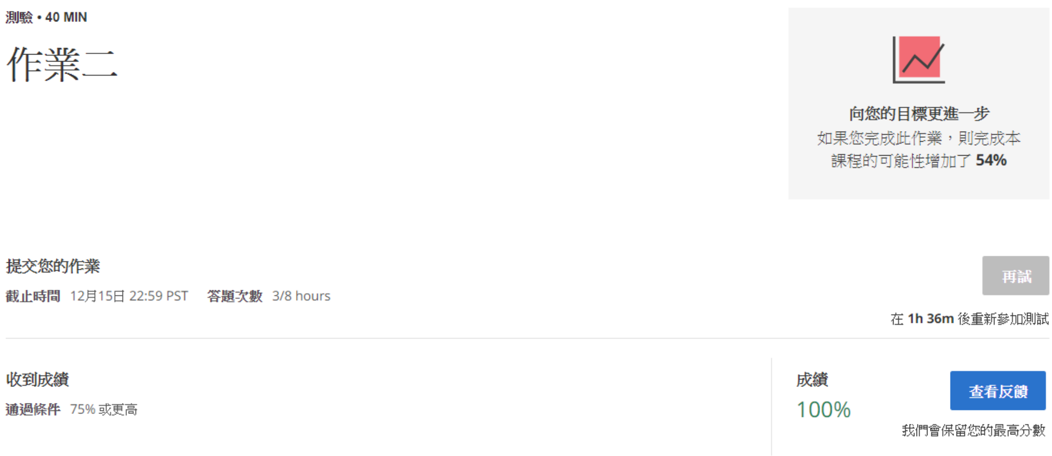
Machine Learning Foundation Homework #2
R08942062 林志皓

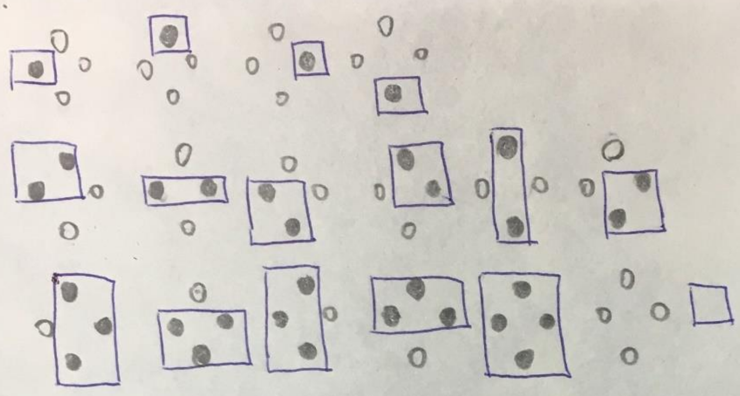
Problem 1



Problem 2

Problem 2

Consider 4 points (● = +1 , O = −1)

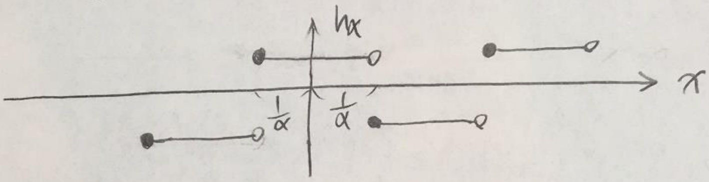


all possible dichotomies ($\# = 2^4 = 16$)
are shattered by "positive rectangle"
$\Rightarrow$ VC-dimension ≥ 4

Problem 3

Problem 3

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



$$h_\alpha(x) = \begin{cases} -1, & \text{if } (\alpha x) \bmod 4 \in [1, 3) \\ +1, & \text{otherwise} \end{cases}$$

Given $Y = (y_1, y_2 \cdots y_N) \in \{-1, +1\}^N$,

Let $Z = (z_1, z_2 \cdots z_N)$, $z_i = \begin{cases} 0, & \text{if } y_i = 1 \\ 2, & \text{if } y_i = -1 \end{cases}$,

$X = (x_1, x_2 \cdots x_N)$ $x_i = 10^i$

assign $\alpha = 0.z_1 z_2 \cdots z_N$

$$\begin{aligned}(\alpha x_i) \bmod 4 &= (z_1 z_2 \cdots z_i . z_{i+1} \cdots z_N) \bmod 4 \\ &= (z_{i-1} z_i . z_{i+1} \cdots z_N) \bmod 4 \quad (\because 100 \bmod 4 = 0). \\ &= (z_i . z_{i+1} \cdots z_N) \bmod 4 \quad (\because 0 \bmod 4 = 20 \bmod 4 = 0) \\ &= z_i . z_{i+1} \cdots z_N \quad (\text{no matter } z_i = 0 \text{ or } 2)\end{aligned}$$

$\therefore (\alpha x_i) \bmod 4 \begin{cases} 0.z_{i+1} \cdots z_N & \text{if } z_i = 0.\ y_i = +1 \\ 2.z_{i+1} \cdots z_N & \text{if } z_i = 2,\ y_i = -1 \end{cases}$

$\therefore h_\alpha(x_i) = \begin{cases} +1, & \text{if } y_i = +1 \\ -1, & \text{if } y_i = -1 \end{cases}$

$\Rightarrow$ Any $X = (x_1, x_2 \cdots x_N)$ can be shattered by $\mathcal{H}$.

N could be any number

$\therefore$ VC-dimension of $\mathcal{H} = \infty$

Problem 4

Problem 4

Consider H_A, H_B s.t. $H_A \subseteq H_B$. $d_{vc}(H_B) = k$

if some $h \subseteq H_A$ can shatter $k+1$ dichotomies.

$\because h \subseteq H_B$. $\therefore d_{vc}(H_B) \geq k+1$, contradiction.

$\therefore d_{vc}(H_A) \leq d_{vc}(H_B)$ if $H_A \subseteq H_B$.

$\because H_1 \cap H_2 \subseteq H_1$ by the result above,

$d_{vc}(H_1 \cap H_2) \leq d_{vc}(H_1)$.

Problem 5

Problem 5

$$\underbrace{\circ \;\; \circ \;\; \cdots\cdots \;\; \circ}_{N \text{ points}}$$

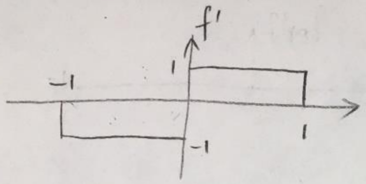
$$m_{H_1 \cup H_2}(N) = 2N$$

$$\left.\begin{aligned} m_{H_1 \cup H_2}(1) &= 2 = 2^1 \\ m_{H_1 \cup H_2}(2) &= 4 = 2^2 \\ m_{H_1 \cup H_2}(3) &= 6 < 2^3 \end{aligned}\right\} d_{vc}(H_1 \cup H_2) = 2$$

Problem 6

Problem 6

Consider case $f'(x) = \text{sign}(x)$ without any noise



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

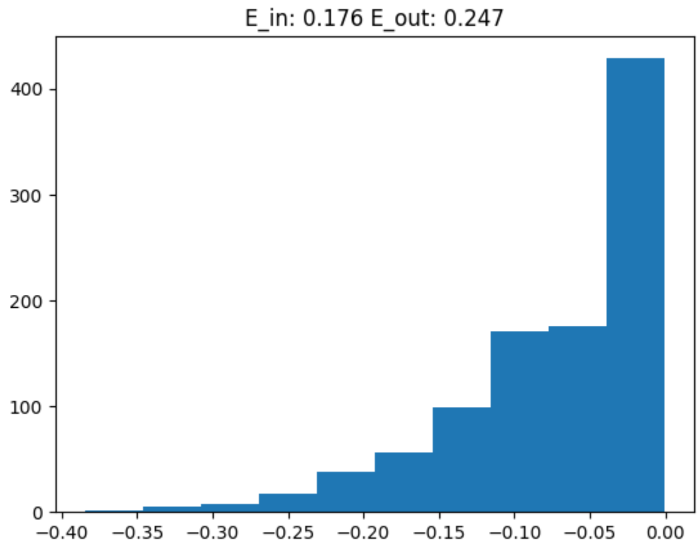
Now add the noise that flips the result with 20 % probability

$$E_{out}(h_{s,\theta}) = 0.8E'_{out} + 0.2(1 - E'_{out})$$

$$= 0.8(0.5 + 0.5s(|\theta| - 1)) + 0.2(0.5 - 0.5s(|\theta| - 1))$$

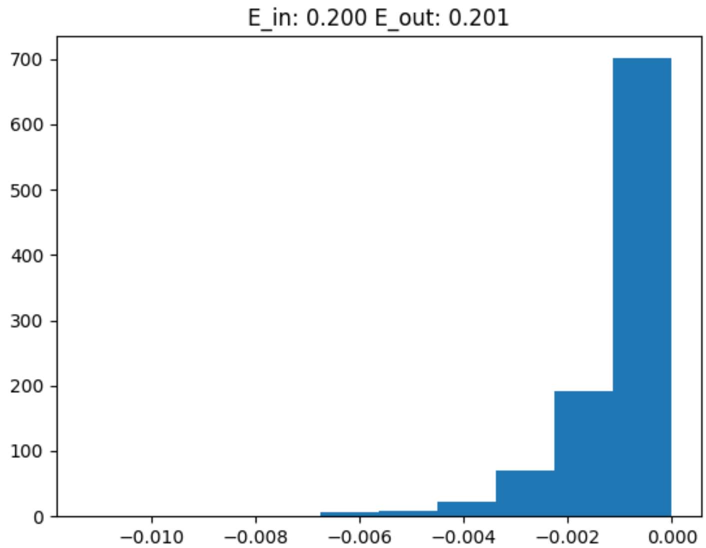
$$= 0.5 + 0.3s(|\theta| - 1)$$

Problem 7



On average E_in = 0.176, E_out = 0.247. From result of problem#6, minimum value of E_out calculated from the formulation is 0.2 (s=1, theta= 0). In this problem, we achieved E_in in training data which is slightly smaller than 0.2, and E_out is slightly bigger than 0.2.

Problem 8



The observations in problem 6 still hold, and in this problem, I found that the difference between E_in and E_out is smaller on average, I think it's because more training data lead to better generalization and less overfitting.

Problem 9

Problem 9 Let $\Theta_t(x)=v,\ v_i=[\![x_i>t_i]\!]$

Let 2^d points X located on vertices of hyper-cube in d-dimension ($x=\{-1,+1\}^d$) set $t=\{0\}^d$ then $v=\Theta_t(x)=\{0,1\}^d$

$V_\alpha=\Theta_t(x_\alpha)\neq V_\beta=\Theta_t(x_\beta), \forall x_\alpha\neq x_\beta$. For any $X^+\subseteq X$, $f(x)=\begin{cases}+1, \forall x\in X^+\\ -1, \text{otherwise}\end{cases}$.

set $S=\{v=\Theta_t(x) \mid \forall x\in X^+\}$

then $h_{t,s}(x)=2[\![v\in S]\!]-1=\begin{cases}+1, \text{ if } f(x)=+1\\ -1, \text{ if } f(x)=-1\end{cases}$

$\Rightarrow$ 2^d points are shattered by $\mathcal{H}$, $d_{vc}(\mathcal{H})\geq 2^d$ —— ①

Consider 2^d+1 points X' in d-dimension. while $V=\{0,1\}^d$ has total 2^d possible values,

$\Rightarrow$ for some $x_\alpha, x_\beta\in X'$, $x_\alpha\neq x_\beta$. $\Theta_t(x_\alpha)=V_\alpha=V_\beta=\Theta_t(x_\beta)$

$\Rightarrow$ $h_{t,s}(x_\alpha)=h_{t,s}(x_\beta)$, can't satisfy case $f(x_\alpha)\neq f(x_\beta)$

$\Rightarrow$ 2^d+1 points can't be shattered by H

$\Rightarrow$ $d_{vc}(H)<2^d+1$ —— ②

by ①&②, $d_{vc}(H)=2^d$ #