Machine Learning Foundation Hw2

資工四 B05902006 蕭縈瀅

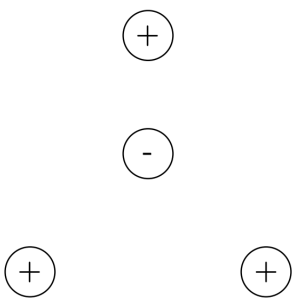
Problem 1

Score：100 %

Problem 2

The VC-Dimension of "positive rectangle" hypothesis set is 4.

Not all sets containing 4 points can be shattered. For example, Fig.1 cannot be shattered. But this doesn't mean the VC-Dimension is at most 3. As shown in Fig.2, the set of 4 points does shatter. Therefore, the VC-Dimension is at least 4.



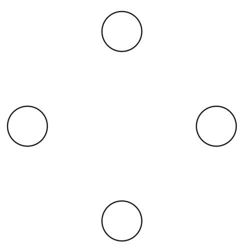


Fig. 1 Fig.2

Consider any 5 distinct points $\{v_1, v_2, v_3, v_4, v_5\} \subseteq R^2$. If there is a rectangle containing the points with maximum $x-$coordinate, minimum $x-$coordinate, maximum $y-$coordinat, minimum $y-$coordinat. Then, these 5 points may not be distinct. As a consequence, the VC-Dimension is 4.

Problem 3

The VC-Dimension of H is ∞

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

Assume $x_i = 2^i, 0 \leq i \leq N$, then we can use $(1 + \frac{k}{2^N}) \leq \alpha \leq (1 + \frac{k+1}{2^N})$ to indicate all 2^N dichotomies.

Therefore, VC-Dimension of H is ∞

Problem 4

Consider $H_1 \cap H_2$ can shatter $d_{vc}(H_1 \cap H_2)$ input

$\Rightarrow H_1$ can shatter **at least** $d_{vc}(H_1 \cap H_2)$ input ($\because H_1 \cap H_2 \leq H_1$)

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

Problem 5

The question is similar to "Fun Time" in Lecture 5 on Coursera (slide: p22/27).

H_1 is the positive ray, H_2 is the negative ray.

$\Rightarrow m_{H_1 \cup H_2} = 2(N+1) - 2 = 2N$

$\Rightarrow d_{vc}(H_1 \cup H_2) = 2$

Problem 6

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

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

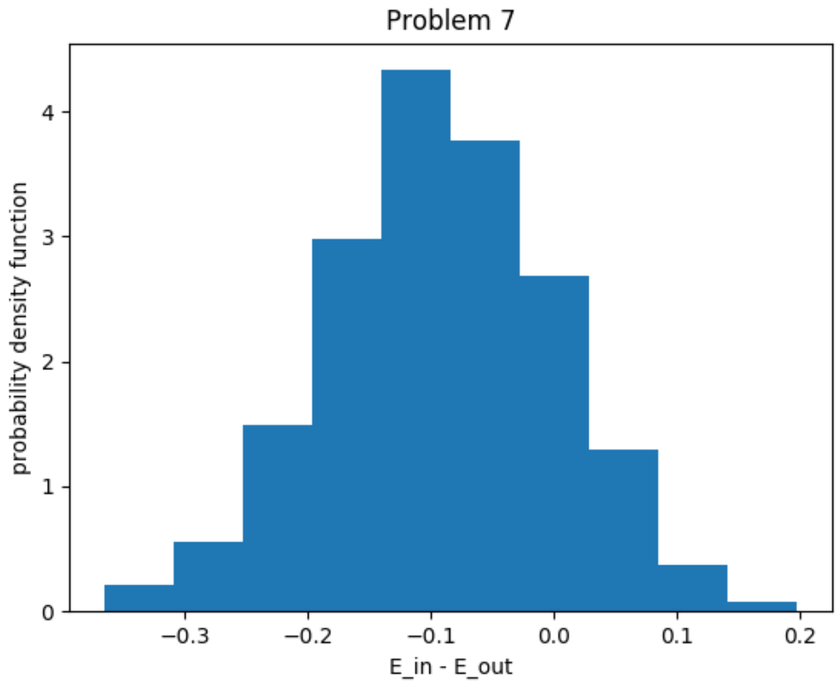
$\mu = \frac{s+1}{2} \times \frac{|\theta|}{2} + \frac{1-s}{2} \times (1 - \frac{|\theta|}{2}) = \frac{1-s+s|\theta|}{2}$

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

Problem 7

Averaged E_{in} : 0.16735, Averaged E_{out} : 0.25802

Findings：Most of the data falls in [-0.2, 0.1], which indicate $h_{s,\theta}$ found by decision stump algorithm has generalization capability.

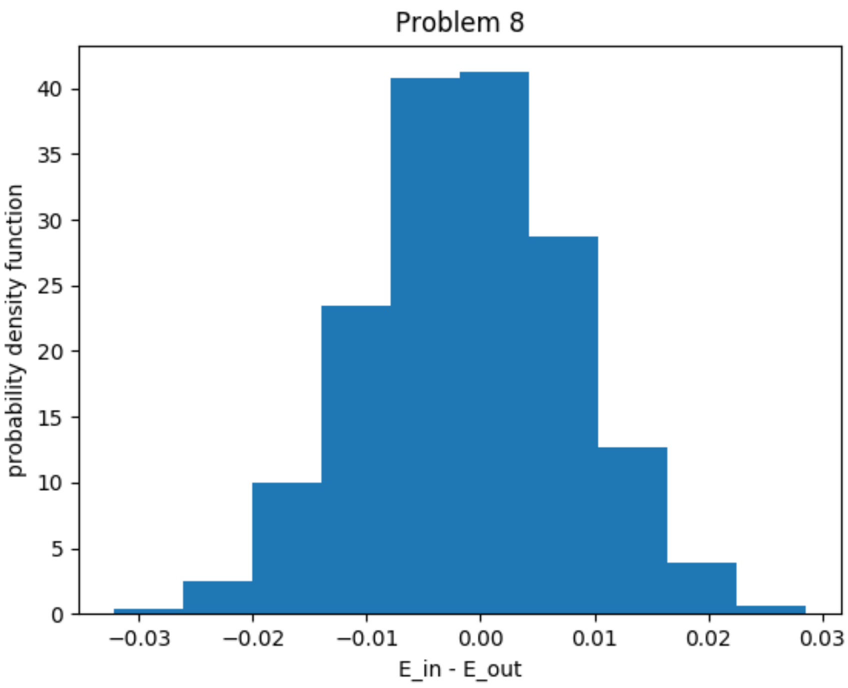


Problem 8

Averaged E_{in} ：0.19965, Averaged E_{out} ：0.20072

Findings：Most of the data still falls in [-0.2, 0.1], which indicate $h_{s,\theta}$ found by decision stump algorithm has generalization capability.

Compare：Averaged $E_{in} -$ Averaged E_{out} became colser, the difference between E_{in} and E_{out} is approaching 0.



Bonus

The VC-Dimension of the "simplified decision trees" hypothesis set is 2^d.

$H = \left\{ h_{t,S} | h_{t,S}(x) = 2[[v \in S]] - 1, where\ v_i = [[x_i > t_i]], S\ a\ collection\ of\ vectors\ in\ \{0,1\}^d, t \in R \right\}$

在 R^2 中, 我們可以將其劃分為 4 個 hyper-rectangular regions, 如 Fig.3 所示；每個 region 內的 decision 都是獨立的， 即這 4 個 region 所代表的類別為 (o,o,o,o)、(o,o,o,x)、(o,o,x,x,)......(x,x,x,x) 共 $2^4 = 16$ 種，故最多能 shatter 4 個點。

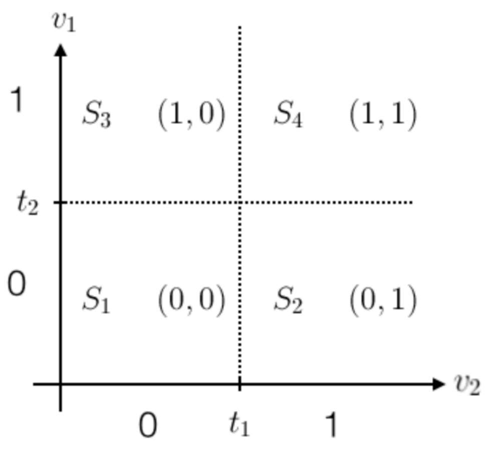


Fig.3

我們由二維推廣至高維，simplified decision tree 的 VC-Dimension 即 hyper-rectangular regions 的個數，而 d 維空間 R^d 中，最多可由 d 條直線分割出 2^d 個相互獨立的 hyper-rectangular regions，亦即最多可 shatter 2^d 個點， 故，VC-Dimension of the "simplified decision trees" hypothesis set is 2^d.