CSIE 5432 — Machine Learning Foundations

Name: 李吉昌 **Homework 2**

Student Number: r08922a27 **Due Date:** November 6 2020, 13:00

Perceptrons

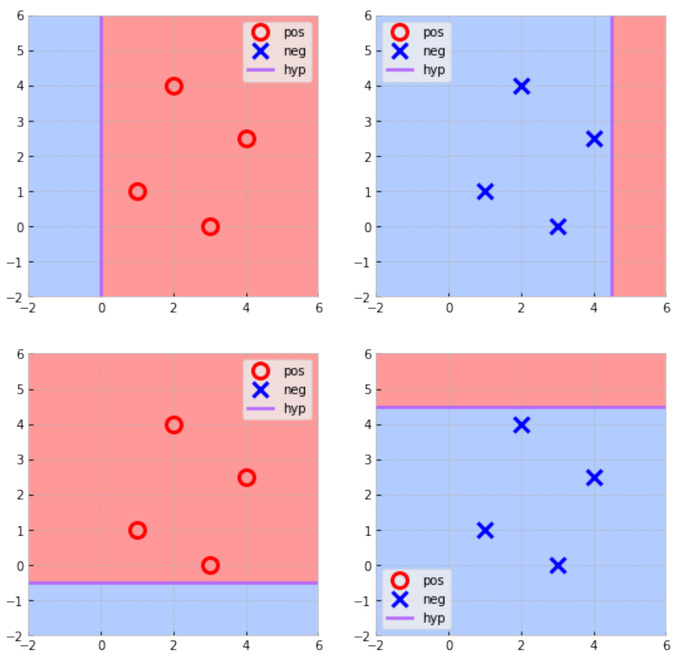
1. Answer: [c]

由 Lecture 7 slides 的第 12 頁可得, 若 data sample 和 bias 項 $(x_0 = 1)$ 構成的矩陣為 X。令 X 為 d 筆資料, 給定任一 label $y \in \{-1, +1\}^d$。若 X 可逆, 則一定找得到一組參數 w 使得 $Xw = y$。 $Xw = y$ 成立則 $\text{sign}(Xw) = y$ 必成立, 即該組 data sample 必定可以被 shatter。選項中僅有 [c] 為可逆方陣, 得 [c] 的 data sample 可以被 shatter。

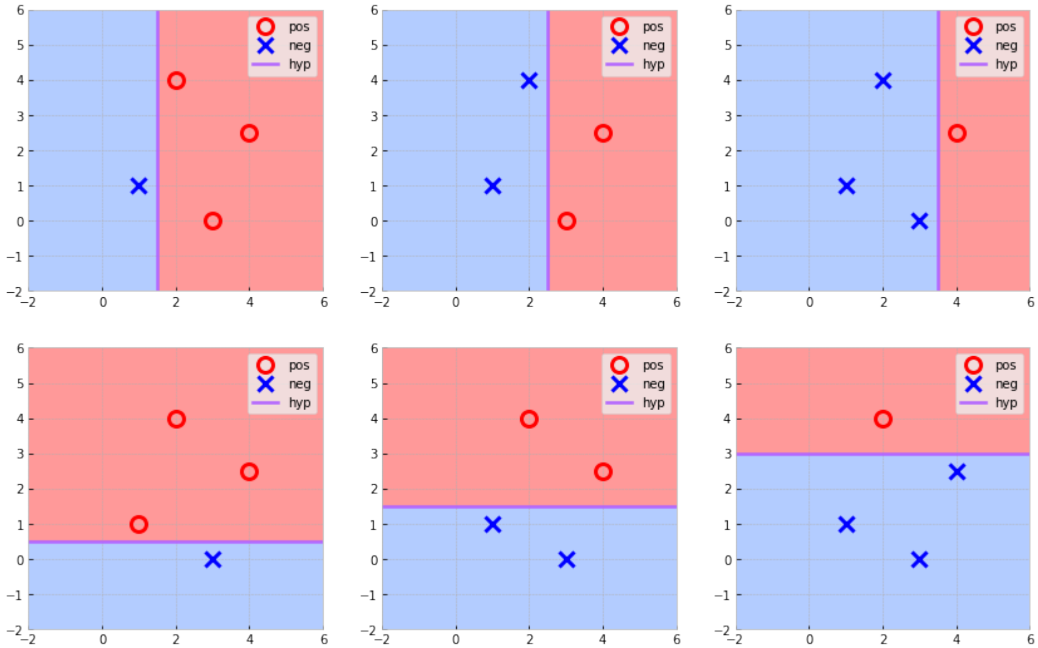
$$X = \begin{pmatrix} 1 & 1 & 1 & 3 \\ 1 & 7 & 8 & 9 \\ 1 & 15 & 16 & 17 \\ 1 & 21 & 23 & 25 \end{pmatrix}, \quad X^{-1} = \begin{pmatrix} 0 & \frac{9}{8} & \frac{7}{8} & -1 \\ \frac{1}{2} & \frac{-11}{8} & \frac{11}{8} & \frac{-1}{2} \\ -1 & \frac{7}{4} & \frac{-3}{4} & 0 \\ \frac{1}{2} & \frac{-1}{2} & \frac{-1}{2} & \frac{1}{2} \end{pmatrix}$$

2. Answer: [d]

令平行 y 軸的 dichotomy set 為 $\mathcal{H}_\mathcal{X}$, 平行 x 軸的 dichotomy set 為 $\mathcal{H}_\mathcal{Y}$, 任兩集合聯集元素的數量可寫成 $|\mathcal{H}_\mathcal{X} \cup \mathcal{H}_\mathcal{Y}| = |\mathcal{H}_\mathcal{X}| + |\mathcal{H}_\mathcal{Y}| - |\mathcal{H}_\mathcal{X} \cap \mathcal{H}_\mathcal{Y}|$, x 和 y 軸的 dichotomy 可以各自視為一樣的 decision stump 的 dichotomy。 在 N 筆 sample 彼此的 x, y 的數值不重複的情況, $|\mathcal{H}_\mathcal{X}| = |\mathcal{H}_\mathcal{Y}| = 2N$。 由下圖所示, 可以確定 $\mathcal{H}_\mathcal{X}$ 和 $\mathcal{H}_\mathcal{Y}$ 一定包含將全部 sample 分成 positive 或 negative 的兩種 dichotomy , 因此 $\mathcal{H}_\mathcal{X} \cap \mathcal{H}_\mathcal{Y}$ 至少會包含這兩種情況, 則 $|\mathcal{H}_\mathcal{X} \cap \mathcal{H}_\mathcal{Y}| \geq 2$:



綜上所述, 得 $|\mathcal{H}_{\mathcal{X}} \cup \mathcal{H}_{\mathcal{Y}}| \leq 2N+2N-2 = 4N-2$。在交集 $\mathcal{H}_{\mathcal{X}} \cap \mathcal{H}_{\mathcal{Y}}$ 只有把全部 sample 分成 positive 或 negative 的兩種 dichotomy 的時候, $|\mathcal{H}_{\mathcal{X}} \cup \mathcal{H}_{\mathcal{Y}}|$ 才有可能是 $4N-2$。在 $N=4$ 時, 可以找到一例符合 $|\mathcal{H}_{\mathcal{X}} \cup \mathcal{H}_{\mathcal{Y}}| = 4 \cdot 4 - 2 = 14$ 的情況, 如下圖所示(每個 dichotomy 因為對稱性可以標籤變號, 僅列出兩種的其中一種):



令 $\mathcal{H}_{\mathcal{X}}' = \mathcal{H}_{\mathcal{X}}/\{◎◎◎◎, ××××\}, \mathcal{H}_{\mathcal{Y}}' = \mathcal{H}_{\mathcal{Y}}/\{◎◎◎◎, ××××\}$, $|\mathcal{H}_{\mathcal{X}} \cup \mathcal{H}_{\mathcal{Y}}| = 4N-2$ 等價於 $\mathcal{H}_{\mathcal{X}}' \cap \mathcal{H}_{\mathcal{Y}}'$ 為空集合。撇除全部 sample 分成 positive 或 negative 的兩種 dichotomy 的情況, 由 x 軸座標數值大小由左而右排序可以得到下面的表格:

$\mathcal{H}_{\mathcal{X}}'$		$\mathcal{H}_{\mathcal{Y}}'$	
×◎◎◎	◎×××	◎◎×◎	××◎×
××◎◎	◎◎××	×◎×◎	◎×◎×
×××◎	◎◎◎×	×◎××	◎×◎◎

由上圖可知, 因為 $\mathcal{H}_{\mathcal{X}}'$ 的 dichotomy 只能將兩邊劃分為不同的類別, 等價所有在 $\mathcal{H}_{\mathcal{X}}'$ 的 dichotomy 標籤只能變號一次, 假使任一 dichotomy 發生兩次以上變號, 垂直線的其中一邊必存在兩種標籤類別, 則一定沒辦法用垂直線分開, 因此只要 $\mathcal{H}_{\mathcal{Y}}'$ 的所有 dichotomy 都發生兩次變號, 則 $\mathcal{H}_{\mathcal{X}}'$ 和 $\mathcal{H}_{\mathcal{Y}}'$ 的交集必為空集合, 得 $|\mathcal{H}_{\mathcal{X}} \cup \mathcal{H}_{\mathcal{Y}}| = 4 \cdot 3 + 2 = 14$。基於上述情況, 若加入一個 sample 在最右邊的第二高的位置, 可得下列表格(新增的部分標為紅色):

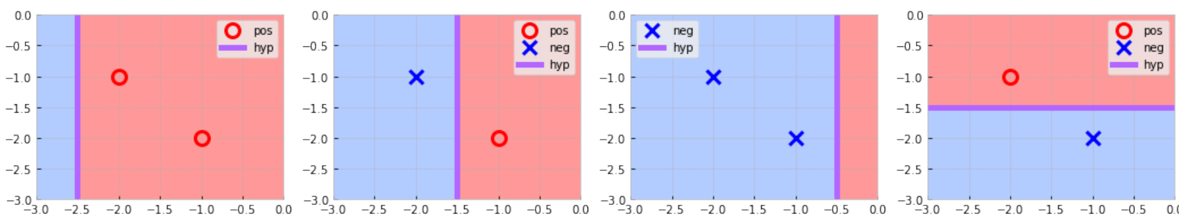
$\mathcal{H}_{\mathcal{X}}'$		$\mathcal{H}_{\mathcal{Y}}'$	
×◎◎◎◎	◎××××	◎◎×◎◎	××◎××
××◎◎◎	◎◎×××	×◎×◎◎	◎×◎××
×××◎◎	◎◎◎××	×◎××◎	◎×◎◎×
××××◎	◎◎◎◎×	×◎×××	◎×◎◎◎

上述表格的 $\mathcal{H}_{\mathcal{Y}}'$ 原 dichotomy 已經變號兩次, 任何新增在右邊的紅色 sample 不論任何情況都只有可能讓標籤變號的次數變高, 一定不可能被包含於 $\mathcal{H}_{\mathcal{X}}'$, 因此在加入新的 sample 時, 只有最後一層全紅的 dichotomy 有可能被包含於 $\mathcal{H}_{\mathcal{X}}'$, 使 $|\mathcal{H}_{\mathcal{X}} \cup \mathcal{H}_{\mathcal{Y}}| < 4N-2$。假設「新增的 sample 置於最右邊, y軸數小於左邊最高的 sample、大於左邊最低的 sample 」的規則存在能夠使 $\mathcal{H}_{\mathcal{Y}}'$ 新增全紅的 dichotomy 被包含於 $\mathcal{H}_{\mathcal{X}}'$, 等價於全紅的 dichotomy 存在一只變號一次的情況, 水平線依據最高或最低切分兩類

時必定先切到最右邊新增的 sample, 代表新增的 sample 必須為最高或是最低的 sample, 則與原新增 sample 的規則矛盾, 表示這一新增 sample 的規則絕對不會出現標籤只變號一次的情況, 新增全紅的 dichotomy 亦標籤變號兩次以上, 依據「新增的 sample 置於最右邊, y軸數小於左邊最高的 sample、大於左邊最低的 sample」的規則產生在 $\mathcal{H}_{\mathcal{Y}}'$ 的 dichotomy 絕對不會被包含於 $\mathcal{H}_{\mathcal{X}}'$, 得依據該規則在 $N=4$ 具備 $|\mathcal{H}_{\mathcal{X}} \cup \mathcal{H}_{\mathcal{Y}}| = 4N-2$ 的條件下, 創造出來的 $N=5$ 的 case 亦會成立 $|\mathcal{H}_{\mathcal{X}} \cup \mathcal{H}_{\mathcal{Y}}| = 4N-2$ 的條件, 依此類推依據同樣條件以及規則創造 $N=6$ 的 case 亦會成立 $|\mathcal{H}_{\mathcal{X}} \cup \mathcal{H}_{\mathcal{Y}}| = 4N-2$。由歸納法得, 在 $N=4$ 具備 $|\mathcal{H}_{\mathcal{X}} \cup \mathcal{H}_{\mathcal{Y}}| = 4N-2$ 的條件下, 用「新增的 sample 置於最右邊, y軸數小於左邊最高的 sample、大於左邊最低的 sample」的規則可以持續使之後的情況皆符合 $|\mathcal{H}_{\mathcal{X}} \cup \mathcal{H}_{\mathcal{Y}}| = 4N-2$。因 $4N-2$ 為 $|\mathcal{H}_{\mathcal{X}} \cup \mathcal{H}_{\mathcal{Y}}|$ 最大值, 並且確定在不同 N 的情況都能確定 找到符合 $|\mathcal{H}_{\mathcal{X}} \cup \mathcal{H}_{\mathcal{Y}}| = 4N-2$ 條件的 case, 得 growth function 為 $4N-2$。

3. Answer: [c]

Stage 1:
由下圖所示, 在 $N=2$, 可以找到一組被 shatter 的 sample 組合, 得 $d_{vc} \geq 2$。



Stage 2:
令三個 sample 為 $\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3} \in \mathbb{R}^2$, 三個 sample 包含 bias 項構成矩陣為 $X = \begin{pmatrix} 1, & -\mathbf{x_1}^T- \\ 1, & -\mathbf{x_2}^T- \\ 1, & -\mathbf{x_3}^T- \end{pmatrix}_{3\times 4}$, X 的 row vector 線性獨立; 令 $\mathbf{w}' \in \mathbb{R}^2$ 為 2D-perceptron 的 weight, bias 為一 scalar w_0, 構成參數向量 $w = \begin{pmatrix} w_0 \\ \mathbf{w}' \end{pmatrix} \in \mathbb{R}^3$, 則 hypothesis 預測結果分別為 $\text{sign}(w_0 + \mathbf{w'}^T \mathbf{x_1})$, $\text{sign}(w_0 + \mathbf{w'}^T \mathbf{x_2})$, $\text{sign}(w_0 + \mathbf{w'}^T \mathbf{x_3})$。若 $\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}$ 中存在零向量, 零向量的預測結果為 $\text{sign}(w_0 + 0) = +1$, label 已經被決定, 一定不能被 shatter; 若 $\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}$ 中存在任兩向量 $\mathbf{x_i}, \mathbf{x_j}$ 為 dependent($\mathbf{i}, \mathbf{j}$ 為屬於 $\{1,2,3\}$ 中兩個任意不同的 index), 可將 $\mathbf{x_i}$ 寫成 $\mathbf{x_i} = c \cdot \mathbf{x_j}$($c$ 為任意實數常數), 則 $\mathbf{x_i}$ 預測結果為 $\text{sign}(w_0 + c \cdot \mathbf{w'}^T \mathbf{x_j})$, 表 $\mathbf{x_i}$ 結果已被 $\mathbf{x_j}$ 決定, 必定沒辦法舉出其他預測結果, 故無法被 shatter ; 若 $\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}$ 彼此間為 pairwise independent, 則由於 $\mathbb{R}^2$ 空間維度為 2, 三個向量必為線性相依, 一定可以找到一組系數 α, β 使得 $\mathbf{x_3} = \alpha \cdot \mathbf{x_1} + \beta \cdot \mathbf{x_2}$。在 $\alpha \cdot \mathbf{w'}^T \mathbf{x_1} > 0, \beta \cdot \mathbf{w'}^T \mathbf{x_2} > 0$ 的情況下, 則 $\mathbf{w'}^T \mathbf{x_3} = \alpha \cdot \mathbf{w'}^T \mathbf{x_1} + \beta \cdot \mathbf{w'}^T \mathbf{x_2} > 0$, 配合題意 $w_0 > 0$, 得 $\mathbf{w'}^T \mathbf{x_3} + w_0 > 0$, 可推得 $\text{sign}(\mathbf{w'}^T \mathbf{x_3} + w_0) > 0$, 綜上述可得知 $\mathbf{x_3}$ 在 $\mathbf{w'}^T \mathbf{x_1}, \mathbf{w'}^T \mathbf{x_2}$ 和 α, β 同號時已被決定結果, 必定不能被 shatter, 故 $d_{vc} \leq 2$。

由 Stage 1 和 Stage 2 結果得 $d_{vc} = 2$。

Ring Hypothesis Set

4. Answer: [b]

令 $\begin{cases} x_1 = \rho \sin\theta \cos\phi \\ x_2 = \rho \sin\theta \sin\phi \\ x_3 = \rho \cos\theta \end{cases}$, $x_1^2 + x_2^2 + x_3^2 = \rho^2$, 得 $a \leq \rho^2 \leq b$, hypothesis 的結果與 θ, ϕ 無關, 即 $h(\mathbf{x})|_{\rho,\theta,\phi} = h(\mathbf{x})|_{\rho,\theta',\phi'}$。在 N 個不同的 sample 如果存在相同的 ρ 會有相同 label, 則 dichotomy 的

數量小於 N 個不同 ρ 的 sample 得出的 dichotomy。 因為 growth function 為 dichotomy set 的最大維度, 僅考慮 N 個不同 ρ 的情況。 與求取區間為 $\sqrt{a}$ 至 $\sqrt{b}$, 以 ρ 為數線軸的 Positive Intervals 方法一樣, N 個 sample 彼此有 $N+1$ 個區間, 共 $N+1$ 取 2 種區間選擇方式以及無區間一種, 共 $\binom{N+1}{2}+1$ 個 dichotomy, 假設存在 N 個 sample 使得 dichotomy set 的維度大於 $\binom{N+1}{2}+1$, 表示 N 個 sample 彼此的區間必須大於 $N+1$, 則 sample 必須大於 N 個, 與假設矛盾, 得 $\binom{N+1}{2}+1$ 為 dichotomy set 的最大維度, growth function 為 $\binom{N+1}{2}+1$。

5. Answer: [b]

growth function 為 $\frac{1}{2}N^2+\frac{1}{2}N+1$。 當 $N=2$ 時, $\frac{1}{2}\cdot 4+\frac{1}{2}\cdot 2+1=4=2^2$; 當 $N=3$ 時, $\frac{1}{2}\cdot 9+\frac{1}{2}\cdot 3+1=7<2^3$, 2 為最大 N 使得 $m_{\mathcal{H}}(2N)=2^N$, 據 Lecture 7 slides 第 6 頁定義, 得 $d_{vc}=2$。

Deviation from Optimal Hypothesis

6. Answer: [d]

令 $\Delta=\sqrt{\frac{8}{N}\ln\left(\frac{4m_{\mathcal{H}}(2N)}{\delta}\right)}$, 由 Lecture 7 slides 的第 24 頁可得, 對任意 $h\in\mathcal{H}$, 皆成立 $E_{in}(h)-\Delta\leq E_{out}(h)\leq E_{in}(h)+\Delta$, 將 g 和 g^* 代入:

$$\begin{cases} E_{in}(g)-\Delta\leq E_{out}(g)\leq E_{in}(g)+\Delta \\ E_{in}(g^*)-\Delta\leq E_{out}(g^*)\leq E_{in}(g^*)+\Delta \end{cases}$$

下式同乘負號可得, $-E_{in}(g^*)-\Delta\leq -E_{out}(g^*)\leq -E_{in}(g^*)+\Delta$, 兩式相加合併可得下面結果,

$$(E_{in}(g)-E_{in}(g^*))-2\cdot\Delta\leq E_{out}(g)-E_{out}(g^*)\leq(E_{in}(g)-E_{in}(g^*))+2\cdot\Delta, \quad (1)$$

因為 $E_{in}(g)\leq E_{in}(g^*)$, 得 $(E_{in}(g)-E_{in}(g^*))\leq 0$, 則 $(E_{in}(g)-E_{in}(g^*))+2\cdot\Delta\leq 2\cdot\Delta$,

得 $E_{out}(g)-E_{out}(g^*)\leq 2\Delta=2\sqrt{\frac{8}{N}\ln\left(\frac{4m_{\mathcal{H}}(2N)}{\delta}\right)}$

The VC Dimension

7. Answer: [d]

在二元分類中, VC dimension 為 d_{vc}, 表示一定存在 d_{vc} 個 sample 可以被 shatter, 則表示 $\mathcal{H}$ 至少具備 $2^{d_{vc}}$ 種 dichotomy, 則 $|\mathcal{H}|\geq 2^{d_{vc}}$。 因為 d_{vc} 為整數, 兩邊取對數後, 以 floor 求取符合不等式最大整數, 代入 $|\mathcal{H}|=M$, 可得 $d_{vc}\leq\lfloor\log_2 M\rfloor$。

8. Answer: [d]

令輸入數量為 N 種, 當 $N\leq k+1$ 時, 可以選擇 sample 出 0 至 N 個 1 的數量的不同輸入, 這個情況 hypothesis 可以一對一給出其對應的 label, 故一定可以被 shatter, 得 $d_{vc}\geq k+1$; 當 $N>k+1$ 時, 必然存在兩種以上具備一樣 1 數量的輸入的情況, 一樣 1 數量的輸入對應的 label 被限制只能有一種, 因此一定沒辦法 shatter 其同類型輸入不同 label 的情況, 得 $d_{vc}\leq k+1$, 綜上述可歸納出 $d_{vc}=k+1$。

9. Answer: [c]

根據 Lecture 7 slides 的第 5 頁的描述, d_{vc} 為 d, 表示在 $\mathcal{H}$ 內必存在 d 個 distinct input 可以被 shatter, 但不一定適用所有 d 個 distinct input 的情況, 且 d 為 distinct input 可以被 shatter 的最大數量, 亦即超過 d 個 distinct input 一定不能被 shatter。

(1) 「some set of d distinct inputs is shattered by $\mathcal{H}$」不成立, 等價 「any set of d distinct inputs is not shattered by $\mathcal{H}$」, 表示 d_{vc} 必小於 d, 則 $d_{vc} = d$ 不成立。
(2) 「some set of d distinct inputs is not shattered by $\mathcal{H}$」 不成立, 等價 「any set of d distinct inputs is shattered by $\mathcal{H}$」, 表示 d_{vc} 在 d 以上, 則 $d_{vc} = d$ 有可能成立。
(3) 「any set of d distinct inputs is shattered by $\mathcal{H}$」 不成立, 等價 「some set of d distinct inputs is not shattered by $\mathcal{H}$」, 不確定 d distinct inputs 能不能被 shatter, 則 $d_{vc} = d$ 有可能成立。
(4) 「any set of d distinct inputs is not shattered by $\mathcal{H}$」 不成立, 等價 「some set of d distinct inputs is shattered by $\mathcal{H}$」, 表示 d_{vc} 在 d 以上, 則 $d_{vc} = d$ 有可能成立。
(5) 「some set of $d+1$ distinct inputs is shattered by $\mathcal{H}$」不成立, 等價 「any set of $d+1$ distinct inputs is not shattered by $\mathcal{H}$」, 不確定 d distinct inputs 能不能被 shatter, 則 $d_{vc} = d$ 有可能成立。
(6) 「some set of $d+1$ distinct inputs is not shattered by $\mathcal{H}$」 不成立, 等價 「any set of $d+1$ distinct inputs is shattered by $\mathcal{H}$」, 表示 d_{vc} 在 $d+1$ 以上, 則 $d_{vc} = d$ 不成立。
(7) 「any set of $d+1$ distinct inputs is shattered by $\mathcal{H}$」 不成立, 等價 「some set of $d+1$ distinct inputs is not shattered by $\mathcal{H}$」, 不確定 d distinct inputs 能不能被 shatter, 則 $d_{vc} = d$ 有可能成立。
(8) 「any set of $d+1$ distinct inputs is not shattered by $\mathcal{H}$」 不成立, 等價 「some set of $d+1$ distinct inputs is shattered by $\mathcal{H}$」, 表示 d_{vc} 在 $d+1$ 以上, 則 $d_{vc} = d$ 不成立。
綜上所述, (1)(6)(8) 為「d_{vc} 為 d」的 necessary condition。

10. Answer: [c]

答案為 [c], 附上參考文章連結, 證明如下:

令待 shatter 的 data sample pair 為 $\{(\mathbf{x_i}, y_i)|i = 1, 2, ..., N\}$, 其中 $\mathbf{x_i} = 2\pi 10^{-i}$, $y_i \in \{-1, +1\}$ 為 $\mathbf{x_i}$ 對應的 label, 並將 α 令為 $\alpha = \frac{1}{2}(1 + \sum_{i=1}^{N} \frac{1-y_i}{2} 10^i) = \frac{1}{2}(1 + \sum_{i:y_i=-1} 10^i)$。 證明方法為分別討論 label y_i 分別為 -1 或 $+1$ 的時候, 當 y_i 為 -1 而 $\sin(\alpha \cdot \mathbf{x_i})$ 永遠為負且 y_i 為 $+1$ 而 $\sin(\alpha \cdot \mathbf{x_i})$ 永遠為正時, 表示 h_α 永遠可以 shatter 所有情況, 則其 VC dimension 為 ∞。

Stage 1:
任一 sample x_j 的 $y_j = -1$, 則:

$$
\begin{aligned}
\alpha \cdot x_j &= \frac{1}{2}(1 + \sum_{i:y_i=-1} 10^i) \cdot 2\pi 10^{-j} \\
&= \pi(10^{-j} + \sum_{i:y_i=-1} 10^{i-j}) \\
&= \pi(10^{-j} + 1 + \sum_{\substack{i:y_i=-1 \\ i>j}} 10^{i-j} + \sum_{\substack{i:y_i=-1 \\ i<j}} 10^{i-j}), \sum_{\substack{i:y_i=-1 \\ i>j}} 10^{i-j} \text{ 為 2 倍數的正整數, 令為 } 2k, \\
&= \pi(10^{-j} + 1 + \sum_{\substack{i:y_i=-1 \\ i<j}} 10^{i-j}) + 2k\pi
\end{aligned} \tag{2}
$$

其中 $\sum_{\substack{i:y_i=-1 \\ i<j}} 10^{i-j} < \sum_{i=1}^{\infty} 10^{-i} = \sum_{i=0}^{\infty} 10^{-i} - 1 = \frac{1}{1-0.1} - 1 = \frac{1}{9}$ 且 $10^{-j} \leq 10^{-1} = \frac{1}{10}$,
令 $10^{-j} + \sum_{\substack{i:y_i=-1 \\ i<j}} 10^{i-j}$ 為 ϵ, 由上述可得 ϵ 範圍, $0 < \epsilon < \frac{1}{9} + \frac{1}{10} = \frac{19}{90} < 1$,
由於 $\sin(\alpha \cdot x_j) = \sin(\pi(1 + \epsilon) + 2k\pi) = \sin(\pi(1 + \epsilon))$, 僅需討論 $\pi(1 + \epsilon)$ 的範圍, 由上述可得, $\pi < \pi(1 + \epsilon) < 2\pi$, 在該區間內 $\sin(\pi(1 + \epsilon)) < 0$, 當 y_j 為 -1 的情況永遠可以預測正確。

Stage 2:

任一 sample x_j 的 $y_j = +1$, 則:

$$\begin{aligned}\alpha \cdot x_j &= \pi(10^{-j} + \sum_{i:y_i=-1} 10^{i-j}) \\ &= \pi(10^{-j} + \sum_{\substack{i:y_i=-1\\ i>j}} 10^{i-j} + \sum_{\substack{i:y_i=-1\\ i<j}} 10^{i-j}), \sum_{\substack{i:y_i=-1\\ i>j}} 10^{i-j} \text{ 為 2 倍數的正整數, 令為 } 2k, \\ &= \pi(10^{-j} + \sum_{\substack{i:y_i=-1\\ i<j}} 10^{i-j}) + 2k\pi,\end{aligned} \tag{3}$$

代入 $\epsilon = 10^{-j} + \sum_{\substack{i:y_i=-1\\ i<j}} 10^{i-j}$, 則 $\alpha \cdot x_j = \pi \cdot \epsilon + 2k\pi$,

由於 $\sin(\alpha \cdot x_j) = \sin(\pi \cdot \epsilon + 2k\pi) = \sin(\pi \cdot \epsilon)$, 僅需討論 $\pi \cdot \epsilon$ 的範圍, 由 Stage 1 可得 $0 < \epsilon < 1$, 則 $0 < \pi \cdot \epsilon < \pi$, 在該區間內 $\sin(\pi \cdot) > 0$, 當 y_j 為 $+1$ 的情況永遠可以預測正確。

綜合 Stage 1 和 Stage 2, 在 label 不同的條件下, 在 $\alpha = \frac{1}{2}(1 + \sum_{i=1}^{N} \frac{1-y_i}{2} 10^i)$ 時必能將所有情況預測正確, 將其 shatter。

Noise and Error

11. Answer: [d]

在 τ 機率干擾下, 看到 $[\![h(\mathbf{x}) \neq y)]\!]$ 發生有可能來自兩種情況, 分別是 y 沒被 flip 且 $h(\mathbf{x})$ 答錯以及 y 被 flip 且 $h(\mathbf{x})$ 答對的情況, 因此, $E_{out}(h, \tau)$ 可以拆成兩情況的疊加, 如下式:

$$\begin{aligned}&E_{out}(h,\tau) = (1-\tau) \cdot \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{P}_0}[\![h(\mathbf{x}) \neq y)]\!] + \tau \cdot \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{P}_0}[\![h(\mathbf{x}) = y)]\!], \\ &\text{代入 } \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{P}_0}[\![h(\mathbf{x}) = y)]\!] = 1 - \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{P}_0}[\![h(\mathbf{x}) \neq y)]\!] \text{ 以及 } E_{out}(h,0) = \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{P}_0}[\![h(\mathbf{x}) \neq y)]\!], \\ &E_{out}(h,\tau) = (1-\tau) \cdot E_{out}(h,0) + \tau \cdot (1 - E_{out}(h,0)), \\ &\text{經移項整理可得 } E_{out}(h,0) = \frac{E_{out}(h,\tau) - \tau}{1-2\tau}\end{aligned} \tag{4}$$

12. Answer: [b]

所有 x, y 產生的 $err(f(\mathbf{x}), y)$ 需要分成 $f(\mathbf{x}) = 1$, $f(\mathbf{x}) = 2$ 和 $f(\mathbf{x}) = 3$ 三種情況討論, 如下式:

$$\begin{aligned}&E_{out}(f) = \mathcal{E}_{(\mathbf{x},y)\sim\mathcal{P}}[err(f(\mathbf{x}), y)] = \\ &\mathbb{P}[f(\mathbf{x}) = 1] \cdot \mathcal{E}_{(\mathbf{x},y)\sim\mathcal{P}|f(\mathbf{x})=1}[err(f(\mathbf{x}), y)](= 0.1 \cdot (1-2)^2 + 0.2 \cdot (1-3)^2 = 0.9) \\ &+\mathbb{P}[f(\mathbf{x}) = 2] \cdot \mathcal{E}_{(\mathbf{x},y)\sim\mathcal{P}|f(\mathbf{x})=2}[err(f(\mathbf{x}), y)](= 0.1 \cdot (2-3)^2 + 0.2 \cdot (2-1)^2 = 0.3) \\ &+\mathbb{P}[f(\mathbf{x}) = 3] \cdot \mathcal{E}_{(\mathbf{x},y)\sim\mathcal{P}|f(\mathbf{x})=3}[err(f(\mathbf{x}), y)](= 0.1 \cdot (3-1)^2 + 0.2 \cdot (3-2)^2 = 0.6) \\ &x \text{ 為平均分布, 因此 } f(\mathbf{x}) \text{ 亦為平均分布, 代入 } \mathbb{P}[f(\mathbf{x}) = 1] = \mathbb{P}[f(\mathbf{x}) = 2] = \mathbb{P}[f(\mathbf{x}) = 3] = \frac{1}{3}, \\ &\text{可得 } E_{out}(f) = \frac{1}{3} \cdot 0.9 + \frac{1}{3} \cdot 0.3 + \frac{1}{3} \cdot 0.6 = 0.6\end{aligned} \tag{5}$$

13. Answer: [b]

同上題所述, 需分段討論在 $f(\mathbf{x})$ 不同的情況下對應的 $f_*(\mathbf{x})$ 和機率:

當 $f(\mathbf{x})=1$ 時: $P(y|\mathbf{x})=\begin{cases}0.7, y=1\\0.1, y=2\\0.2, y=3\end{cases}$ 得 $f_*(\mathbf{x})=1\cdot 0.7+2\cdot 0.1+3\cdot 0.2=1.5$

當 $f(\mathbf{x})=2$ 時: $P(y|\mathbf{x})=\begin{cases}0.2, y=1\\0.7, y=2\\0.1, y=3\end{cases}$ 得 $f_*(\mathbf{x})=1\cdot 0.2+2\cdot 0.7+3\cdot 0.1=1.9$

當 $f(\mathbf{x})=3$ 時: $P(y|\mathbf{x})=\begin{cases}0.1, y=1\\0.2, y=2\\0.7, y=3\end{cases}$ 得 $f_*(\mathbf{x})=1\cdot 0.1+2\cdot 0.2+3\cdot 0.7=2.6$

$$\begin{aligned}\Delta(f,f_*)&=\mathbb{E}_{\mathbf{x}\sim P(\mathbf{x})}(f(\mathbf{x})-f_*(\mathbf{x}))^2\\&=\mathbb{P}[f(\mathbf{x})=1]\cdot\mathbb{E}_{\mathbf{x}\sim P(\mathbf{x}|f(\mathbf{x})=1)}(f(\mathbf{x})-f_*(\mathbf{x}))^2\\&+\mathbb{P}[f(\mathbf{x})=2]\cdot\mathbb{E}_{\mathbf{x}\sim P(\mathbf{x}|f(\mathbf{x})=2)}(f(\mathbf{x})-f_*(\mathbf{x}))^2\\&+\mathbb{P}[f(\mathbf{x})=3]\cdot\mathbb{E}_{\mathbf{x}\sim P(\mathbf{x}|f(\mathbf{x})=3)}(f(\mathbf{x})-f_*(\mathbf{x}))^2\\&=\frac{1}{3}\cdot(1-1.5)^2+\frac{1}{3}\cdot(2-1.9)^2+\frac{1}{3}\cdot(3-2.6)^2=0.14\end{aligned}\tag{6}$$

Decision Stump

14. Answer: [d]

由 Lecture 7 slides 的第 4 頁可得 $\delta=4m_{\mathcal{H}}(2N)\exp\left(-\frac{1}{8}\epsilon^2N\right)$, 代入 $m_{\mathcal{H}}(2N)=2(2N)=4N$, $\epsilon=0.1$ 以及 $\delta=0.1$, 得 $0.1=16\cdot N\cdot exp(-0.00125\cdot N)$。

$16\cdot N\cdot exp(-0.00125\cdot N)\approx\begin{cases}53.096(>\delta), N=6000\\5.8112(>\delta), N=8000\\0.5963(>\delta), N=10000\\0.0587(<\delta), N=12000\\0.0056(<\delta), N=14000\end{cases}$, 得 $N=12000$ 時為滿足條件最小值。

15. Answer: [b]

須分成 $\theta>0$ 和 $\theta\le 0$ 的兩種情形討論, 如下式:

$E_{out}(h_{+1,\theta},0)=\mathbb{P}[\theta>0]\cdot\mathbb{E}_{\mathbf{x}\sim\mathcal{P}}[\![x\in(0,\theta]|\theta>0]\!]+\mathbb{P}[\theta\le 0]\cdot\mathbb{E}_{\mathbf{x}\sim\mathcal{P}}[\![x\in[\theta,0)|\theta\le 0]\!]$

因為 $\mathbf{x}$ 為平均分布, 則 θ 亦為平均分布, 得 $\mathbb{P}[\theta>0]=\mathbb{P}[\theta\le 0]=\frac{1}{2}$,

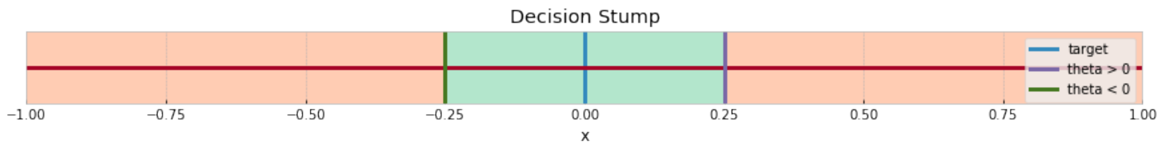
由下圖所示, 右半和左半邊綠色區塊和整個定義域 $[-1,+1]$ 的比例即為不同情況下發生錯誤的機率,

當 $\theta>0$, 右半邊綠色區塊的機率為 $\mathbb{E}_{\mathbf{x}\sim\mathcal{P}}[\![x\in(0,\theta]|\theta>0]\!]=\frac{1}{2}\cdot\theta=\frac{1}{2}|\theta|$,

當 $\theta\le 0$, 左半邊綠色區塊的機率為 $\mathbb{E}_{\mathbf{x}\sim\mathcal{P}}[\![x\in[\theta,0)|\theta\le 0]\!]=\frac{1}{2}\cdot(-\theta)=\frac{1}{2}|\theta|$,

得 $E_{out}(h_{+1,\theta},0)=\frac{1}{2}\cdot\frac{1}{2}|\theta|+\frac{1}{2}\cdot\frac{1}{2}|\theta|=\frac{1}{2}|\theta|$

(7)



Experiment

程式碼實作細節如下，可以透過 parser 的 −−mode 參數決定用 closed form 或 simulation 求取 $E_{out}(h_{s,\theta}, \tau)$:

```
python code.py --mode closedform
python code.py --mode simulate
```

```
import numpy as np
import random
from scipy.stats import bernoulli
import argparse

'''Define Function'''


def generate_data(size, tau):
    x = np.sort(np.random.uniform(-1, 1, size))
    y = np.zeros(size).astype(int)
    y[x > 0] = 1
    y[x <= 0] = -1

    noisy_idx = bernoulli.rvs(tau, size=size) > 0
    y[noisy_idx] = -y[noisy_idx]

    return x, y


def get_g_Ein(x, y):
    theta_set = ([-1] + list(((x[1:] + x[:-1]) / 2))) * 2
    s_set = [-1] * len(x) + [1] * len(x)
    hypothesis_set = np.array(
        sorted(tuple(zip(s_set, theta_set)), key=lambda x: x[0] + x[1]))
    g, Ein = (-1, -1), 1

    for hypothesis in hypothesis_set:
        err = get_err(hypothesis, x, y)
        if Ein > err:
            g, Ein = hypothesis, err
        if Ein == 0:
            break
    return g, Ein
```

```python
def get_err(hypothesis, x, y):
    s, theta = hypothesis
    pred = np.ones(len(y)).astype(int)

    if s > 0:
        pred[x <= theta] = -1
    else:
        pred[x > theta] = -1

    return (y != pred).sum() / len(y)


def get_Eout(hypothesis, tau, IsSimulate=False):
    if IsSimulate:
        x_tst, y_tst = generate_data(100000, tau)
        Eout = get_err(hypothesis, x_tst, y_tst)
        return Eout
    else:
        s, theta = hypothesis
        Eout = 0.5 * np.abs(theta) if s > 0 else 1 - 0.5 * np.abs(theta)
        return (1 - 2 * tau) * Eout + tau


def get_answer(exp_num, size, tau, IsSimulate=False):
    ans = []

    for _ in range(exp_num):
        random.seed(random.randint(1, 10000))
        x_tra, y_tra = generate_data(size, tau)

        g, Ein = get_g_Ein(x_tra, y_tra)
        Eout = get_Eout(g, tau, IsSimulate)
        ans.append(Eout - Ein)

    return np.mean(ans)


def main():
    '''Parsing'''
    parser = argparse.ArgumentParser(
        description='Argument Parser for MLF HW1.')
    parser.add_argument('--mode', default='closedform',
                        choices=['closedform', 'simulate'])
    args = parser.parse_args()
    if args.mode == 'simulate':
        is_simulate = True
        print("Tesing by simulation!")
    elif args.mode == 'closedform':
        is_simulate = False
        print("Tesing by closed form!")
```

```python
    '''Answer questions'''
    print('RUNNING Q16...')
    print('Answer of Q16 : {:.4f}\n'.format(get_answer(
        exp_num=10000, size=2, tau=0, IsSimulate=is_simulate)))

    print('RUNNING Q17...')
    print('Answer of Q17 : {:.4f}\n'.format(get_answer(
        exp_num=10000, size=20, tau=0, IsSimulate=is_simulate)))

    print('RUNNING Q18...')
    print('Answer of Q18 : {:.4f}\n'.format(get_answer(
        exp_num=10000, size=2, tau=0.1, IsSimulate=is_simulate)))

    print('RUNNING Q19...')
    print('Answer of Q19 : {:.4f}\n'.format(get_answer(
        exp_num=10000, size=20, tau=0.1, IsSimulate=is_simulate)))

    print('RUNNING Q20...')
    print('Answer of Q20 : {:.4f}\n'.format(get_answer(
        exp_num=10000, size=200, tau=0.1, IsSimulate=is_simulate)))


if __name__ == "__main__":
    main()
```

16. Answer: [d]	17. Answer: [b]	18. Answer: [e]	19. Answer: [c]	20. Answer: [a]
0.2930	0.0243	0.3671	0.0519	0.0051