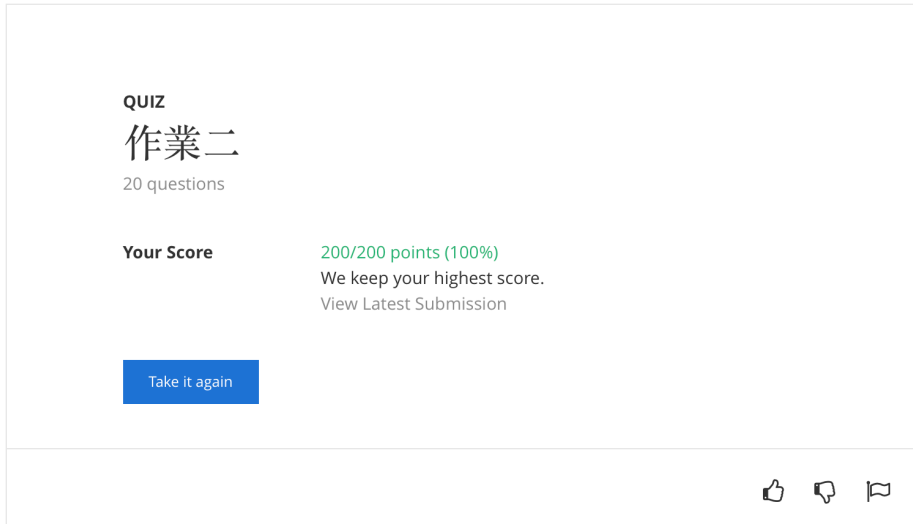


Machine Learning Foundation HW2

B04902004 王佑安

1.



2. positive intervals 的 $m_{\mathcal{H}}(N) = C_2^{N+1}$, negative 跟 positive 一樣, 扣掉 interval = $[-\infty, r]$ 跟 $[l, \infty]$ 會重複, 這兩種情況分別會重複 $N - 1$ 種, 因此 positive-and-negative intervals 的 growth function $m_{\mathcal{H}}(N) = 2 \times C_2^{N+1} - 2 \times (N - 1) = N^2 - N + 2$ 。
3. 將 $x^1 \dots x^D$ 視為一個 D 維的 vector 就可以用 PLA 找出 $d + 1$ 個 input 時的所有 dichotomies 因此 $d_{vc} \geq d + 1$, 而在 $d = 1$ 時可以找到 $\times \circ \times$ 的 input 不能被 shatter 因此 $d_{vc} \leq d + 1 \Rightarrow d_{vc} = d + 1$
4. 假設 $x_i = 2^i, 0 \leq i \leq N$, $(1 + \frac{k}{2^N}) \leq \alpha \leq (1 + \frac{k+1}{2^N}), 0 \leq k \leq 2^N$ 可以表示全部 2^N 種 dichotomies, 因此 $d_{vc} = \infty$
5. $\mathcal{H}_1 \subseteq \mathcal{H}_2$ 表示 \mathcal{H}_1 的 dichotomies $\leq \mathcal{H}_2$ 的 dichotomies $\Rightarrow m_{\mathcal{H}_1}(N) \leq m_{\mathcal{H}_2}(N)$, 根據 VC dimension 的定義 $d_{vc}(\mathcal{H}) = \text{largest } N \text{ for which } m_{\mathcal{H}}(N) = 2^N \Rightarrow d_{vc}(\mathcal{H}_1) \leq d_{vc}(\mathcal{H}_2)$
6. positive-ray 的 $m_{\mathcal{H}}(N) = N + 1$, 扣掉全為 \times 跟全為 \circ 的情況重複, $m_{\mathcal{H}_1 \cup \mathcal{H}_2}(N) = 2 \times (N + 1) - 2 = 2N$, $N = 2$, $2N = 2^2 = 4 \Rightarrow d_{vc}(\mathcal{H}_1 \cup \mathcal{H}_2) = 2$

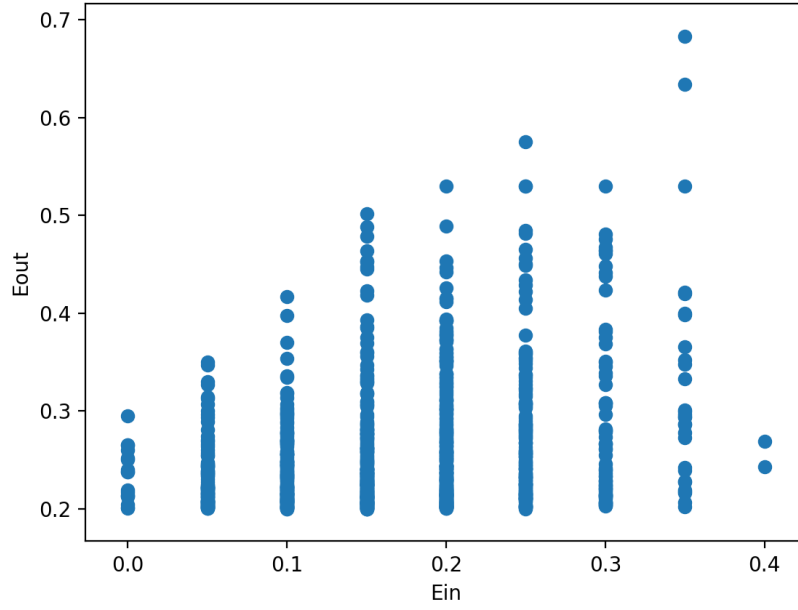
7. $s = +1 \Rightarrow \mu = 0.5|\theta|$

$s = -1 \Rightarrow \mu = 1 - 0.5|\theta|$

$\mu = \frac{s+1}{2} \times 0.5|\theta| + \frac{1-s}{2} \times (1 - 0.5|\theta|) = 0.5 \times (s|\theta| - s + 1)$

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

8.



E_{out} 是錯誤發生的機率，因此可以看出基本上錯誤率越高，在 data 上的 E_{in} 就愈高。