

Machine Learning Homework4

B01902032 江東峻

1. Answer: b

因為 $H' \subset H$ ，所以 H' 有更多的資訊被覆蓋 \Rightarrow deterministic noise 增加。

2. Answer: c

$$H(10, 0, 3) = w_0 L_0(x) + w_1 L_1(x) + w_2 L_2(x)$$

$$H(10, 0, 4) = w_0 L_0(x) + w_1 L_1(x) + w_2 L_2(x) + w_3 L_3(x)$$

$$H(10, 1, 3) = w_0 L_0(x) + w_1 L_1(x) + w_2 L_2(x) + L_3 + L_4 + \dots + L_{10}$$

$$H(10, 1, 4) = w_0 L_0(x) + w_1 L_1(x) + w_2 L_2(x) + w_3 L_3(x) + L_4 + \dots + L_{10}$$

$$[c]: H(10, 0, 3) \cap H(10, 0, 4) = \{h | h(x) = w^T z \in H_{10}; w_q = 0 \text{ for } q \geq 3\} = w_0 L_0(x) + w_1 L_1(x) + w_2 L_2(x) = H_2$$

3. Answer: a d

要做 minimize $E_{\text{aug}}(w_t) \Rightarrow$ 往 $E_{\text{aug}}(w_t)$ 梯度的反方向走 $\Rightarrow w_{t+1} = w_t - \eta \nabla E_{\text{aug}}(w_t) = [a]$

$$\text{又 } E_{\text{aug}}(w_t) = E_{\text{in}}(w_t) + \lambda/N * w^T w$$

$$\rightarrow \nabla E_{\text{aug}}(w_t) = \nabla E_{\text{in}}(w_t) + 2\lambda/N * w \text{ 代入 [a] 得 [d]}$$

4. Answer: b c

[b]: $w_{\text{reg}}(\lambda)$ 是被限制在 $w^T w \leq C$ (一個圓) 的條件下。若 w_{lin} 在 C 外，則 $\|w_{\text{reg}}(\lambda)\| < \|w_{\text{lin}}\|$

若 w_{lin} 在 C 內，則 $\|w_{\text{reg}}(\lambda)\| = \|w_{\text{lin}}\|$ ，所以 $\|w_{\text{reg}}(\lambda)\| \leq \|w_{\text{lin}}\|$ 。

[c]: 當 $\lambda=0$ 時， $\|w_{\text{reg}}(\lambda)\| = \|w_{\text{lin}}\|$ ，而根據 [b]，在 $\lambda>0$ 時， $\|w_{\text{reg}}(\lambda)\| \leq \|w_{\text{lin}}\|$ ，而 $\|w_{\text{lin}}\|$ 是定值，所以 $\|w_{\text{reg}}(\lambda)\|$ 是 non-increasing function。

5. Answer: c

for constant model:

拿出 $(-1, 0)$ ，剩下是 $y=1/2$ ， $e_1=1/4$

拿出 $(p, 1)$ ，剩下是 $y=0$ ， $e_2=1$

拿出 $(1, 0)$ ，剩下是 $y=1/2$ ， $e_3=1/4$

for linear model:

拿出 $(-1, 0)$ ，剩下是 $y=(x-1)/(p-1)$ ， $e_1=4/(p-1)^2$

拿出 $(p, 1)$ ，剩下是 $y=0$ ， $e_2=1$

拿出 $(1, 0)$ ，剩下是 $y=(x+1)/(p+1)$ ， $e_3=4/(p+1)^2$

$$\text{tied} \Rightarrow 1/4 + 1 + 1/4 = 4/(p-1)^2 + 1 + 4/(p+1)^2$$

$$\Rightarrow p^4 - 18p^2 - 15 = 0, \quad p^2 = 9 + 4\sqrt{6}, \quad p = \sqrt{9 + 4\sqrt{6}} = [c]$$

6. Answer: a c

[a]: 每一局都有輸贏，共有五局 $\Rightarrow 2^5 = 32$ 種預測。

[c]: 當第一局的結果出來後，只要對於第一局結果預測正確的人送信，剩下4局的預測有16種可能，每一種可能送給一個人，至少要有16個人。

[d]: 一開始先選32個人平均送出預測結果，至少會有16人預測正確，在對這16人送出結果，... 預測5次總共送信次數 = $32 + 16 + 8 + 4 + 2 = 62$

7. Answer: b

如果要有一個人是全部預測正確，至少要送出62封，再加上一封要錢的，總共要37封，支出 = 630，收入 = 1000，賺370元。

8. Answer: a

因為 data 已經固定，且只有一個 function，剛好是 perfect prediction。

\Rightarrow 只有一個 hypothesis，size of hypothesis set = 1

9. Answer: c

$$\text{Hoeffding's Inequality: } P[|v-u|>0.1] \leq 2 * \exp(-2 * 0.01^2 * 10000) = 0.2706 \dots = [c]$$

10. Answer: a c

[a]: 若 $a(x) = 1$ ，則全部的資料都會被拿來做 g 的學習，沒有 sample bias

[c]: $g(x)$ 是只有在 $a(x)=1$ 的條件下在做篩選， $a(x)$ 沒過=0，即使過了 $a(x)$ ， $g(x)$ 沒過也是0

過了 $a(x)$ 等同進入 $a(x)=1$ 的條件，因此 $g(x)$ 就可以做篩選，篩選條件= $a(x)$ 過且在 $a(x)=1$ 條件下 $g(x)$ 過 $\Rightarrow a(x) \text{ AND } g(x) = [c]$

11. Answer: d

最小值發生在微分=0: N+K不看

原式 $\Rightarrow (w^T X - y)^T (w^T X - y) + (w^T \tilde{X} - \tilde{y})^T (w^T \tilde{X} - \tilde{y})$

微分: $-2X^T y + 2X^T X w - 2\tilde{X}^T \tilde{y} + 2\tilde{X}^T \tilde{X} w = 0$

移項後: $w = (X^T X + \tilde{X}^T \tilde{X})^{-1} (X^T y + \tilde{X}^T \tilde{y}) = [d]$

12. Answer: b

最小值: 將右式微分後=0

得 $\lambda w + X^T (Xw - y) = 0$

$w = (X^T X + \lambda I)^{-1} X^T y$

若要11題的 $w = 12$ 題的 w

代入各個選項至11題的 w 表示式:

[a]: $(X^T X + I)^{-1} X^T y$

[b]: $(X^T X + \lambda I)^{-1} X^T y$

[c]: $(X^T X + \lambda^2 I)^{-1} (X^T y + \lambda)$

[d]:

選項[b]的組合使11題的 $w = 12$ 題的 w

13. Answer: d

14. Answer: d

15. Answer: d

16. Answer: e

17. Answer: a

18. Answer: c

19. Answer: e

20. Answer: b

21.

替換掉原本問題12的regularizer:

$w_{\text{reg}} = \arg \min_w \lambda/N * w^T \Gamma^T \Gamma w + 1/N * (Xw - y)^T (Xw - y)$

微分=0求最小值:

$2\lambda/N * \Gamma^T \Gamma w + 2/N * (X^T X w - X^T y) = 0$

$\Rightarrow w = (X^T X + \lambda \Gamma^T \Gamma)^{-1} X^T y$

以11題的方法表示: $\tilde{X} = \sqrt{\lambda} * \Gamma$, $\tilde{y} = 0$ 。

22.

替換掉原本問題12的regularizer:

$w_{\text{reg}} = \arg \min_w \lambda/N * (w - w_{\text{hint}})^T (w - w_{\text{hint}}) + 1/N * (Xw - y)^T (Xw - y)$

微分=0求最小值:

$2\lambda/N * (w - w_{\text{hint}}) + 2/N * (X^T X w - X^T y) = 0$

$\Rightarrow w = (X^T X + \lambda I)^{-1} (X^T y + \lambda w_{\text{hint}})$

以11題的方法表示: $\tilde{X} = \sqrt{\lambda} * I$, $\tilde{y} = \sqrt{\lambda} * w_{\text{hint}}$ 。