Machine Learning Homework #4

Qing-Cheng Li

R01922024

1. A

In general, there is a too powerful / complex model, so that overfit the f , so there is a big deterministic noise. If we choice H’ to replace H, it will have not so powerful / complex model, this will help tp decrease the deterministic noise in general.

2. C

In H(10,0,3), $w_3 = 0$, so this is at most equal to H_2 – (a) and (b) are wrong. H(10,0,3) ∩ H(10,0,4) is H(10,0,3) which $w_q = 0, q > 3$, is equal to H_2 – (c) is correct. H_1’s w_3 and w_4 must be 0, but in (d) it is 1, so (d) is wrong.

3. A,D

Similar to the update function of w_{lin} , use (a), not (b) as the update rule. And,
$\nabla E_{arg}(w) = \nabla E_{in}(w) + \frac{2\lambda}{N} w$, so select (d), too.

4. B,C

(a)(b) If $w_{lin}^T w_{lin} \le C$, $w_{reg} = w_{lin}$, but sometimes w_{reg} is bound by a constant C, w_{lin} not, so $|w_{lin}| \ge |w_{reg}|$, select (b).

(c)(d) $w_{reg} = (Z^T Z + \lambda I)^{-1} Z^T y$, when λ increasing, w_{reg} *will* decreasing (in general), so it is a non-increasing function.

5. C

$h_0 : y = 0$, err = 1 ; $y = 0.5$, err = 3*0.5^2, Total Error = $1^2 + 2(\frac{1}{2}^2) = \frac{3}{2}$.

$h_1 : y = 0$, err = 1 ; $y = \frac{1}{1+\rho} x + \frac{1}{1+\rho}$, err = $(\frac{2}{1+\rho})^2$; $y = \frac{-1}{1-\rho} x + \frac{1}{1-\rho}$, err = $(\frac{2}{1-\rho})^2$, Total Error = $1 + (\frac{2}{1-\rho})^2 + (\frac{2}{1+\rho})^2$

$\frac{1}{2} = \frac{4}{1+2\rho+\rho^2} \frac{4}{1-2\rho+\rho^2} = \frac{8+8\rho^2}{1-2\rho^2+\rho^4}$, $\rho^4 - 18\rho^2 - 15 = 0$,
$\rho^2 = \frac{18 \pm \sqrt{18^2 - 4 \cdot 1 \cdot 15}}{2} = 9 \pm 4\sqrt{6}$, $\rho > 0$, $\rho = \sqrt{9 + 4\sqrt{6}}$, (c)

6. A,C

(a) $\{win, lose\}^5 = 32$

(b) at least 32 people, not 5.

(c) After first game, there are half of people (32/2 = 16) should be targeted.

(d) 32 + 16 + 8 + 4 + 2 != 64

7. B

(32 + 16 + 8 + 4 + 2 + 1)*10 = 630, 1000 - 630 = 370, (b)

8. C

$\{approve, not\ approve\}^N = 2^N$, (c)

9. C

$2e^{-2\epsilon^2 N} = 2e^{-2 \star 0.01^2 10000}$ ~= 0.27067056647, (c)

10. A,C

(a) 會發生這種問題是因為訓練資料裡面只有被核發信用卡的用戶，而不包含沒有被核發信用卡的客戶。拿到這種測試資料大概會學出幾乎直接同意發出信用卡的模型，拿到真實有的核發，有的不核發的情境下自然就會出問題。不過如果銀行本來就是來者不拒，那就不存在不被核發的狀況，當然就不會發生此題所述的狀況。

(b)(c)(d) 同上所述，我們學出來的模型只針對了銀行過去會核發信用卡的狀況，所以如果先以銀行的$\alpha(x)$ 分出同意核發的人再用$g(x)$ 判斷整體來說結果會變好。

11. D

$\triangledown \frac{1}{N+K}(|Xw - y|^2 + |\tilde{X}w - \tilde{y}|^2) = \frac{2}{N+K}(X^TXw - X^Ty + \tilde{X}^T\tilde{X}w - \tilde{X}^T\tilde{y})$, let it be 0, $(X^TX + \tilde{X}^T\tilde{X})w = X^Ty + \tilde{X}^T\tilde{y}$, $w = (X^TX + \tilde{X}^T\tilde{X})^{-1}(X^Ty + \tilde{X}^T\tilde{y})$, (d)

12. B

$\triangledown w_{reg} = \frac{2\lambda}{N}w + \frac{2}{N}X^TXw - \frac{2}{N}X^Ty$, let this be 0, $\lambda w + X^TXw - X^Ty = 0$, $(\lambda I + X^TX)w = X^Ty$, $w = (\lambda I + X^TX)^{-1}X^Ty$, so $\tilde{X}^T\tilde{X} = \lambda I$ and $\tilde{X}^T\tilde{y} = 0$, $\tilde{X} = \sqrt{\lambda} I$, $\tilde{y} = 0$,(b)

13. E_{in} = 0.05, E_{out} = 0.045

$ python hw4-13.py

14. $\lambda = 10^{-10}$,E_{in} = 0.015, E_{out} = 0.02

15. $\lambda = 10^{-7}$,E_{in} = 0.03, E_{out} = 0.015

```
$ python hw4-14.py
```

16. $\lambda = 10^{-9}$, $E_{train}(g_{\lambda}^{-}) = 0.0$, $E_{val}(g_{\lambda}^{-}) = 0.1$, , $E_{out}(g_{\lambda}^{-}) = 0.038$

17. $\lambda = 10^{-7}$, $E_{train}(g_{\lambda}^{-}) = 0.0333$, $E_{val}(g_{\lambda}^{-}) = 0.0375$, , $E_{out}(g_{\lambda}^{-}) = 0.0375$

18. $E_{in}(g_{\lambda}) = 0.03$, $E_{out}(g_{\lambda}) = 0.015$

```
$ python hw4-16.py
```

19. $\lambda = 10^{-8}$, $E_{vc} = 0.03$

20. $E_{in}(g_{\lambda}) = 0.015$, $E_{out}(g_{\lambda}) = 0.02$

```
$ python hw4-19.py
```