Machine Learning Homework 6

B01902032 江東峻

1. Answer: a

$$F(A,B) = \frac{1}{N}\sum_{n=1}^{N} \ln(1+\exp(-y_n(Az_n+B)))$$

$$\nabla F(A,B) = \left(\frac{1}{N}\sum_{n=1}^{N} \frac{\partial \ln(G)}{\partial G} \frac{\partial(1+\exp(F))}{\partial F} \frac{\partial\left(-y_n\left(Az_n+B\right)\right)}{\partial A},\ \frac{1}{N}\sum_{n=1}^{N} \frac{\partial \ln(G)}{\partial G} \frac{\partial(1+\exp(F))}{\partial F} \frac{\partial\left(-y_n\left(Az_n+B\right)\right)}{\partial B} \right)$$

$$= \left(\frac{1}{N}\sum_{n=1}^{N} \frac{1}{G}\ \exp(F)(-y_n z_n),\ \frac{1}{N}\sum_{n=1}^{N} \frac{1}{G}\ \exp(F)(-y_n) \right) = \left(\frac{1}{N}\sum_{n=1}^{N} p_n(-y_n z_n),\ \frac{1}{N}\sum_{n=1}^{N} p_n(-y_n) \right)$$

=[a]

2. Answer: a

$$\frac{\partial F}{\partial A} = \frac{1}{N}\sum_{n=1}^{N} -y_n p_n z_n \qquad \frac{\partial F}{\partial B} = \frac{1}{N}\sum_{n=1}^{N} -y_n p_n$$

$$\frac{\partial^2 F}{\partial A^2} = \frac{1}{N}\sum_{n=1}^{N} -y_n z_n \frac{\partial\theta(H)}{\partial H}\frac{\partial H}{\partial A} = \frac{1}{N}\sum_{n=1}^{N} -y_n z_n \theta(H)(1-\theta(H))(-y_n z_n) = \frac{1}{N}\sum_{n=1}^{N} z_n^2\ p_n(1-p_n)$$

$$\frac{\partial^2 F}{\partial A\partial B} = \frac{1}{N}\sum_{n=1}^{N} -y_n z_n \frac{\partial\theta(H)}{\partial H}\frac{\partial H}{\partial B} = \frac{1}{N}\sum_{n=1}^{N} -y_n z_n \theta(H)(1-\theta(H))(-y_n) = \frac{1}{N}\sum_{n=1}^{N} z_n\ p_n(1-p_n)$$

$$\frac{\partial^2 F}{\partial B^2} = \frac{1}{N}\sum_{n=1}^{N} -y_n \frac{\partial\theta(H)}{\partial H}\frac{\partial H}{\partial B} = \frac{1}{N}\sum_{n=1}^{N} -y_n \theta(H)(1-\theta(H))(-y_n) = \frac{1}{N}\sum_{n=1}^{N} p_n(1-p_n)$$

H(F) = [a]

3. Answer: b

Kernel matrix裏，$K(x_n, x_m)$是表示第n個x與第m個x做轉換後的內積，所以K的大小是取決于data 的多寡=> n,m都從1~N => K = N * N的矩陣

4. Answer: d

當$|y_n - w^T\phi(x_n) - b| \le \varepsilon$時，$\xi_n$都為0，而當$|y_n - w^T\phi(x_n) - b| > \varepsilon$時，不是 >ε 就是 < -ε ，只會滿足其中一種可能：其中一邊的ξ_n^2 = $(|y_n - w^T\phi(x_n) - b| - \varepsilon)$^2，另一邊的$\xi_n$便為0。

則綜合來說，$C\sum(\xi_n^{\vee 2} + \xi_n^{\wedge 2}) = C\sum(\max(0,|y_n - w^T\phi(x_n) - b| - \varepsilon))$^2 => [d]

5. Answer: a

$$F(b,w) = \frac{1}{2}w^T w + C\sum_{n=1}^{N}\left(\max\left(0,|y_n - w^T\phi(x_n) - b| - \varepsilon\right)\right)^2$$

$$F(b,\beta) = \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\beta_n\beta_m K(x_n, x_m) + C\sum_{n=1}^{N}\left(\max\left(0,\left|y_n - \sum_{m=1}^{N}\beta_m K(x_m, x_n) - b\right| - \varepsilon\right)\right)^2$$

$$= \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\beta_n\beta_m K(x_n, x_m) + C\sum_{n=1}^{N}[|y_n - s_n| \ge \varepsilon](\ \mathrm{sign}(y_n - s_n)(y_n - s_n) - \varepsilon)^2$$

$$\frac{\partial F}{\partial\beta_m} = \sum_{n=1}^{N}\beta_n K(x_n, x_m) + 2C\sum_{n=1}^{N}[|y_n - s_n| \ge \varepsilon](\ \mathrm{sign}(y_n - s_n)(y_n - s_n) - \varepsilon)\ \frac{\partial\left(\ \mathrm{sign}\left(y_n - s_n\right)\left(y_n - s_n\right) - \varepsilon\right)}{\partial\beta_m}$$

$$= \sum_{=1}^{N}\beta_n K(x_n, x_m) + 2C\sum_{n=1}^{N}[|y_n - s_n| \ge \varepsilon](\ |y_n - s_n| - \varepsilon)\ \mathrm{sign}(y_n - s_n)\left(-\frac{\partial s_n}{\partial\beta_m}\right)$$

$$= \sum_{=1}^{N}\beta_n K(x_n, x_m) + 2C\sum_{n=1}^{N}[|y_n - s_n| \ge \varepsilon](\ |y_n - s_n| - \varepsilon)\ \mathrm{sign}(y_n - s_n)(-K(x_n, x_m))$$ =[a]

6. Answer: a

$$\sum_{m=1}^{M} g_t(\tilde{x_m})\tilde{y_m} = -\frac{1}{2}\left\{\sum_{m=1}^{M}\left[g_t(\tilde{x_m}) - \tilde{y_m}\right]^2 - \sum_{m=1}^{M}\left[\tilde{y_m}\right]^2 - \sum_{m=1}^{M}\left[g_t(\tilde{x_m})\right]^2\right\} = -\frac{1}{2}(Me_t - Me_0 - Ms_t) = \frac{M}{2}(e_0 - e_t + s_t) = [a]$$

7. Answer: b

對任意一個資料在pN次取樣裡都沒被取到的機率是(1-1/N)^(pN)，
當N非常大：(1-1/N)^N = e^(-1), 則沒被取到的機率是e^(-p)，有N筆資料：
沒被取到的資料數的期望值為N*e^(-p) = [b]

8. Answer: b c e

[a]: X會被L,R bound ，且x都落在整數點上=>有限個。
[b]: $g_{+1,1,L-1}$ =全為+1 ， $g_{-1,3,R+1}$也是全為+1 ，則 $g_{+1,1,L-1} = g_{-1,3,R+1}$
[c]: 若$g_{s,i,\theta} = g_{s,i,\text{ceiling}(\theta)}$ 則 $s \cdot \text{sign}(x_i - \theta) = s \cdot \text{sign}(x_i - \text{ceiling}(\theta))$
=> $\text{sign}(x_i - \theta) = \text{sign}(x_i - \text{ceiling}(\theta))$
若 $x_i \ge \theta$ 且 $x_i \ge \text{ceiling}(\theta)$ => sign都為1
若 $x_i \ge \theta$ 且 $x_i < \text{ceiling}(\theta)$ => 不會發生，因為x_i為格子點
若 $x_i < \theta$ 且 $x_i \ge \text{ceiling}(\theta)$ => 不會發生，因為$\text{ceiling}(\theta) \ge \theta$
若 $x_i < \theta$ 且 $x_i < \text{ceiling}(\theta)$ => sign都為-1
=>得 $g_{s,i,\theta} = g_{s,i,\text{ceiling}(\theta)}$
[d]: X有(R-L+1)*d種，與decision stump的數量= 2d*(R-L)+2不同。
[e]: 對某個d來說：有(R-L)*2+2種，有兩個d，但對於全+1或全-1是一樣的([b])，在扣掉重複的2種，得12*2-2 = 22種decision stump。

9. Answer: c

$$K_{ds}(x, x') = \Phi_{ds}(x)^T \Phi_{ds}(x') = \sum_{i=1}^{|G|} g_i(x) g_i(x')$$

又可分為g(x)g(x')相乘為1或-1： $= \sum_{g_i(x)g_i(x')>0}^{\|} g_i(x)g_i(x') - \sum_{g_i(x)g_i(x')<0}^{\|} |g_i(x)g_i(x')|$

$$= \sum_{g_i(x)g_i(x')>0}^{\|} g_i(x)g_i(x') + \sum_{g_i(x)g_i(x')<0}^{\|} |g_i(x)g_i(x')| - 2\cdot \sum_{g_i(x)g_i(x')<0}^{\|} |g_i(x)g_i(x')| = |G| - 2\cdot \sum_{g_i(x)g_i(x')<0}^{\|} |g_i(x)g_i(x')|$$

又|G|為所有不同decision stump = 2d*(R-L)+2
g(x)g(x')要異號：theta必須選在兩個x之間，有$\|x-x'\|_1$種可能，每種可能有兩個方向：

$$= 2d(R-L)+2 - 2 \cdot 2\,\|x-x'\|_1 = [c]$$

10. Answer: d

$$\text{argmax}_{1\le n\le N}\, u^{T+1} = \text{argmax}_{1\le n\le N}\, u^1 \cdot \prod_{t=1}^{T}\left(d(t) \textit{ if error}, \tfrac{1}{d(t)} \textit{ if correct}\right)$$
$$= \text{argmax}_{1\le n\le N}\, \ln(u^1) + \sum_{t=1}^{T}\{\ln(d(t)) \textit{ if } y_n g_t(x_n) < 0, -\ln(d(t)) \textit{ if } y_n g_t(x_n) > 0\}$$
$$= \text{argmax}_{1\le n\le N} \sum_{t=1}^{T} -y_n g_t(x_n) \ln(d(t)) = \text{argmax}_{1\le n\le N} \sum_{t=1}^{T} -y_n g_t(x_n)\, \alpha_t = [d]$$

11. Answer: a c d

[a]: $\sum_{n=1}^{N} u_n^{(1)} = \sum_{n=1}^{N} \frac{1}{N} = 1$
[b]: if Ein(g_t) = 1, $U^{(t+1)}$ = sum of ($u^{(t)}$ * 0(因為$\varepsilon_t = 1$)) = 0，矛盾。
[c]: $U^{t+1} = \sum_{n=1}^{N} u_n^{t+1} = \sum_{n=1}^{N} \exp\left(\ln u^1 - y_n \sum_{t'=1}^{t} \alpha_{t'} g_{t'}(x_n)\right) = \frac{1}{N}\sum_{n=1}^{N} \exp\left(-y_n \sum_{t'=1}^{t} \alpha_{t'} g_{t'}(x_n)\right) = \frac{1}{N}\sum_{n=1}^{N} \exp(-y_n G^t(x_n))$

而exp的error必大於0/1 error ，若正確: exp > 1，0/1 = 1，若錯誤: exp > 0，0/1 = 0。
則可知 $U^{t+1} > \text{Ein}(G_t)$ 。
[d]: $U^{t+1} = \sum_{n=1}^{N} u_n^{t+1} = \sum_{n=1}^{N} u_n^t \left(\sqrt{\frac{1-\varepsilon}{\varepsilon}}, \textit{if error}, \sqrt{\frac{\varepsilon}{1-\varepsilon}}, \textit{if correct}\right) = U^t\left(\sqrt{\frac{1-\varepsilon}{\varepsilon}}\cdot\varepsilon, \textit{if error}, \sqrt{\frac{\varepsilon}{1-\varepsilon}}\cdot(1-\varepsilon), \textit{if correct}\right)$ (ε的定義)
乘入根號內使用算幾不等式得： $U^{t+1} = U^t\left(\sqrt{\varepsilon(1-\varepsilon)}_{err} + \sqrt{\varepsilon(1-\varepsilon)}_{correct}\right) \le U^t$
“=“發生在ε = 0.5 => if ε < 0.5 ,then $U^{t+1} < U^t$
[e]: 根據[d]，當ε = 1時， $U^{t+1} = 0$, 而U^t不一定為0 。

12. Answer: a
13. Answer: c
14. Answer: b
15. Answer: d
16. Answer: c
17. Answer: a
18. Answer: c
19. Answer: c
20. Answer: a or Φ
21. No

    若有某個adaBoost-stump的所有可能decision stump個數為G。讓T>G，
    則根據鴿籠原理，g_t只有G種，但是有T個g_t => 必有重複的g_t 。
22. Yes