Homework 2 of Machine Learning

B01902032 江東峻

1. Answer: c
   Expect Ein = 0.01*(1-9/N) >0.008
   N>45 =>選c
2. Answer: a d e (*)
   [a]: $H^* = H^T$ (H 是在實數裡)，
   $H^T = (X(X^TX)^{-1}X^T)^T = (X^T)^T (X(X^TX)^{-1})^T = X((X^TX)^{-1})^T X^T$
   $= X(X^TX)^{-1}X^T = H$
   又HH = H (看[e])
   => $H^*H = H^TH = HH = H$
   又H*H 必為positive semi-definite => H為positive semi-definite
   [b]: H是projection matrix —> HH = H, 若H^{-1}存在=> H=Identity matrix=> not always invertible

   [c]: $Hb = \lambda b$ b: eigenvector
   $Hb = HHb = H\lambda b = \lambda Hb = \lambda\lambda b = \lambda b$
   $\lambda(\lambda-1)b = 0$ 但b不為0 => $\lambda= 0$ or 1
   [d]: 因為$\lambda= 0$ or 1，eigenvalue 1 的數量= non-zero eigenvalue的數量
   the number of non-zero eigenvalue = rank(H) = rank(X) = d+1
   [e]: $HH = X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = H$，連乘1126次也還是H
3. Answer: a b e
   [a]: 當$\text{sign}(w^Tx) = y$, $[\text{sign}(w^Tx)\neq y] = 0$,
   若$yw^Tx \geq 1$則err=0, 若$0 < yw^Tx < 1$，err = $1-yw^Tx >0$ —>bound
   當$\text{sign}(w^Tx) \neq y$, $[\text{sign}(w^Tx)\neq y] = 1$,
   $yw^Tx < 0$, $1-yw^Tx >1$, err = $1-yw^Tx >1$ => bound
   [a]為upper bound
   [b]: 當$\text{sign}(w^Tx) = y$, $[\text{sign}(w^Tx)\neq y] = 0$,
   若$yw^Tx \geq 1$則err=0, 若$0 < yw^Tx < 1$，err = $(1-yw^Tx)^2 >0$ —>bound
   當$\text{sign}(w^Tx) \neq y$, $[\text{sign}(w^Tx)\neq y] = 1$,
   $yw^Tx < 0$, $1-yw^Tx >1$, err = $(1-yw^Tx)^2 >1$ => bound
   [b]為upper bound
   [c]: 當$\text{sign}(w^Tx) = y$, $[\text{sign}(w^Tx)\neq y] = 0$,
   而$yw^Tx \geq 0$=> err=0 —>bound
   當$\text{sign}(w^Tx) \neq y$, $[\text{sign}(w^Tx)\neq y] = 1$,
   $yw^Tx < 0$, $-yw^Tx >0$, err = $-yw^Tx >0$ => 不一定bound
   [c]不是upper bound
   [d]: $0 < \theta(-yw^Tx) < 1$，當$\text{sign}(w^Tx) \neq y$, $[\text{sign}(w^Tx)\neq y] = 1$ —>not bound
   [d]不是upper bound
   [e]: 當$\text{sign}(w^Tx) = y$, $[\text{sign}(w^Tx)\neq y] = 0$,
   而$yw^Tx \geq 0$=> err ≥ 0 —>bound
   當$\text{sign}(w^Tx) \neq y$, $[\text{sign}(w^Tx)\neq y] = 1$,
   $yw^Tx < 0$, $-yw^Tx >0$, err = $\exp(-yw^Tx)^2 >1$ => bound
   [e]為upper bound

4. Answer: b d e

   [a]: 若$yw^Tx \geq 1$則err=0 —>differentiable
   若$yw^Tx < 1$, err = $1-yw^Tx$ —> differentiable
   當$yw^Tx = 1$, err= 0(右邊) = $1-yw^Tx$ (左邊) —> continuous
   左側在$yw^Tx = 1$微分 ≠ 0 = 右側在$yw^Tx = 1$微分 —> 這點不可微
   [a]不為differentiable functions of w everywhere

   [b]: 若$yw^Tx \geq 1$則err=0 —>differentiable
   若$yw^Tx < 1$, err = $(1-yw^Tx)^2$ —> differentiable
   當$yw^Tx = 1$, err= 0(右邊) = $(1-yw^Tx)$^2 (左邊) —> continuous
   左側在$yw^Tx = 1$微分 = 0 = 右側在$yw^Tx = 1$微分 —> 這點可微
   [b]為differentiable functions of w everywhere

   [c]: 同[a]，
   左側在$yw^Tx = 1$微分 ≠ 0 = 右側在$yw^Tx = 1$微分 —> 這點不可微
   [c]不為differentiable functions of w everywhere

   [d]: $\theta(-yw^Tx) = 1/(1+e^{(yw^Tx)})$，在任何地方可微
   [b]為differentiable functions of w everywhere

   [e]: err = $\exp(-yw^Tx)^2$ 在任何地方可微
   [e]為differentiable functions of w everywhere

5. Answer: c

   PLA: err = $[\text{sign}(w^Tx) \neq y]$
   [c]: err = $\max(0, -yw^Tx)$
   當 w^Tx與y不同號 ，PLA-err = 1
   [c]-err = $-yw^Tx > 0$
   當 w^Tx與y同號 ，PLA-err = 0
   [c]-err = 0 ($-yw^Tx < 0$)

6. Answer: d

   $\nabla E(u, v) = (\partial E/\partial u, \partial E/\partial v)$
   $=(e^u+ve^{uv}+2u-2v-3, 2e^v+ue^{uv}-2u+4v-2)$
   帶入(u, v) = (0, 0): $\nabla E(0, 0) = (-2, 0)$ = [d]

7. Answer: c

	u	v	∂E/∂u	∂E/∂v
$(u_0$	0	0	-2	0
$(u_1$	0.02	0	-1.93979866	-0.02
$(u_2$	0.03939799	0.0002	-1.881219644	-0.0377975
$(u_3$	0.05821018	0.00057798	-1.824219836	-0.0535831
$(u_4$	0.07645238	0.00111381	-1.768758161	-0.0675305
$(u_5$	0.09413996	0.00178911		
E(u	2.82500036			

   $E(u_5, v_5) = 2.82500036$ 較接近c

8. Answer: b

$E_2(0+\Delta u, 0+\Delta v) = E(0, 0) + (\Delta u\partial/\partial u+\Delta v\partial/\partial v)E(0,0)+1/2! *(\Delta u\partial/\partial u+\Delta v\partial/\partial v)^2 * E(0,0)$

b_{uu} = 0.5 * ∂^2E(0,0)/∂u^2 = 0.5*(e^u + v^2 * e^(uv)+2) where (u,v) = (0,0) = 1.5

b_{vv} = 0.5 * ∂^2E(0,0)/∂v^2 = 0.5*(4e^(2v) + u^2 * e^(uv) + 4) where (u,v) = (0,0) = 4

b_{uv} = ∂^2E(0,0)/∂u∂v = uve^(uv) + e^(uv) -2 where (u,v) = (0,0) = -1

b_u = ∂E(0,0)/∂u = e^u + ve^(uv) +2u-2v-3 where (u,v) = (0,0) = -2

b_v = ∂E(0,0)/∂v = 2e^(2v) + u*e^(uv) -2u +4v -2 where (u,v) = (0,0) = 0

b = E(0,0) = 3

($b_{uu}, b_{vv}, b_{uv}, b_u, b_v$, b) = (1.5, 4, -1, -2, 0, 3) = [b]

9. Answer: a

$E2(\Delta u, \Delta v) = E(u,v) + \nabla E(u,v)*(\Delta u, \Delta v) + 0.5*[(\Delta u, \Delta v)]^T \nabla^2 E(u,v)*(\Delta u, \Delta v)$

取微分=0 => $\nabla E(u,v) +\nabla^2 E(u,v)*(\Delta u, \Delta v) = 0$

則$(\Delta u, \Delta v) = -(\nabla^2 E(u,v))^{-1}\nabla E(u,v)$ = [a]

10. Answer: c

	u	v	Δu	Δv	u’	v’
(u	0	0	0.69565216	0.08695652	0.69565216	0.08695652
(u	0.69565216	0.08695652	-0.081889959	-0.009349643	0.613762201	0.077606877
(u	0.613762201	0.077606877	-0.00193331	-0.001676787	0.611828891	0.07593009
(u	0.611828891	0.07593009	-8.3229E-06	-0.00117218	0.611820568	0.07475791
(u	0.611820568	0.07475791	0.69565216	0.08695652	0.69565216	0.08695652
(u	0.69565216	0.08695652				
E(u	2.36087349					

11. Answer: e

讓x=(z1, z2)，設6維的hypothesis = $(1, z_1, z_2, z_1^2, z_1z_2, z_2^2)$。

X = 每一行放一個x的矩陣， 若對於任意的長條的y, 若存在一個w使得y = Xw

$w = X^{-1}y$ —> 這個w存在 <=> X有反矩陣。

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} \quad \det(X) = -16,\ X^{-1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0.25 & 0.25 & -0.25 & -0.25 & 0 & 0 \\ 0.25 & -0.25 & -0.25 & 0.25 & 0 & 0 \\ -0.25 & -0.25 & 0.25 & 0.25 & -1 & 1 \\ 0.25 & -0.25 & 0.25 & -0.25 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 & -1 \end{bmatrix}$$

=> 對任意的 y， $\{x_1, x_2, x_3, x_4, x_5, x_6\}$都可以被union of quadratic, linear, or constant hypotheses of x shatter => biggest subset = $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ = [e]

12. Answer: e

    z0 = (1, 0, 0, …, 0)

    Φ(x_1) = $\mathbf{z_1}$(向量z)= (1, 1, 0, 0, …, 0)　(最前面的1是為了閾值的第0項)

    Φ(x_2) = $\mathbf{z_2}$= (1, 0, 1, 0, …, 0)

    Z = ($\mathbf{z_0}$, $\mathbf{z_1}$, $\mathbf{z_2}$, …, $\mathbf{z_n}$) , $Z = \begin{bmatrix} z_0 \\ \ldots \\ z_N \end{bmatrix}$ 顯然對於這個 **Z**對於任何的**y**，

    都可找到**w**使得**y=Zw**($\mathbf{Z}^{-1}$ exist) —>**w** = $\mathbf{Z}^{-1}\mathbf{y}$ —> shatter

    加入一個x_{N+1}，每個**z**也多一維， $Z = \begin{bmatrix} z_0 \\ \ldots \\ z_{N+1} \end{bmatrix}$ 且Z顯然有反矩陣存在—>shatter

    => 不斷加入x，都還是shatter —> d^{vc}(HΦ)—> ∞

13. Answer: c
14. Answer: a
15. Answer: a
16. Answer: d

    對於一個x，找最有可能的h ＝ max h_y(x) ，等同於找 max 的 ln(h_y(x))

    <=> $$\ln\left(\frac{\exp\left(w_{y_n}^T x\right)}{\sum_{i=1}^{K}\exp\left(w_i^T x\right)}\right) = \ln\left(\exp\left(w_{y_n}^T x\right)\right) - \ln\left(\sum_{i=1}^{K}\exp\left(w_i^T x\right)\right) = w_{y_n}^T x - \ln\left(\sum_{i=1}^{K}\exp\left(w_i^T x\right)\right)$$

    等同於找min 的 $$\ln\left(\sum_{i=1}^{K}\exp\left(w_i^T x_n\right)\right) - w_{y_n}^T x_n$$

    有n個Ｘ => $$\frac{1}{N}\sum_{n=1}^{N}\left(\ln\left(\sum_{i=1}^{K}\exp\left(w_i^T x_n\right)\right) - w_{y_n}^T x_n\right)$$

    =[d]

17. Answer: c

$$\frac{\partial}{\partial w_i}\left(\frac{1}{N}\sum_{n=1}^{N}\left(\ln\left(\sum_{i=1}^{K}\exp\left(w_i^T x_n\right)\right) - w_{y_n}^T x_n\right)\right) = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{\partial}{\partial w_i}\left(\ln\left(\sum_{i=1}^{K}\exp\left(w_i^T x_n\right)\right) - w_{y_n}^T x_n\right)\right)$$

$$= \frac{1}{N}\sum_{n=1}^{N}\left(\frac{x_n\exp\left(w_i^T x_n\right)}{\sum_{j=1}^{K}\exp\left(w_j^T x_n\right)} - [y_n=i]x_n\right) = \frac{1}{N}\sum_{n=1}^{N}\left(h_i(x_n) - [y_n=i]\right)x_n$$ = [c]

18. Answer: a
19. Answer: d
20. Answer: a

21. Answer: N+1

設h: h(x) = 0, RMSE(h)= $\sum_{n=1}^{N} y_n^{\ 2}$ 再設h'(x)=[x=x_1] ，RMSE(h') = $\sum_{n=2}^{N} y_n^{\ 2} + (y_1\text{-}1)^2$

兩式相減可算出y_1，照這個方法可算出y_1~y_N，再算上h(x) = 0 ->總共需要N+1次RMSE

22. Answer: 2

N(RMSE(h))^2

$$= \sum_{n=1}^{N}\left(y_n^{\ 2}\text{-}2y_nh(x_n)+h^2(x_n)\right) = \sum_{n=1}^{N} y_n^{\ 2}\ \text{-}2\sum_{n=1}^{N} y_nh(x_n)\ +\sum_{n=1}^{N} h^2(x_n) = N(RMSE(h(x)=0))^2\text{-}2\sum_{n=1}^{N} y_nh(x_n)\ +\sum_{n=1}^{N} h^2(x_n)$$

其中 $\sum_{n=1}^{N} y_nh(x_n)$ 是所求， $\sum_{n=1}^{N} h^2(x_n)$ 不需要用RMSE，所以 $\sum_{n=1}^{N} y_nh(x_n)$ 可以用

2次RMSE , (RMSE(h)和RMSE(h(x)=0))算出來。

23. Answer: K+1

當所求的min發生時， $\left(\frac{\partial RMSE(H)}{\partial w_1}, \frac{\partial RMSE(H)}{\partial w_2}, \dots \frac{\partial RMSE(H)}{\partial w_K}\right) = (0, 0, \dots, 0)$

=> $\frac{\partial RMSE(H)}{\partial w_i}$ = 0 => $\frac{1}{N}\sum_{n=1}^{N}\left(h_i(x_n)\sum_{k=1}^{K} w_kh_k(x_n)\ \text{-}y_nh_i(x_n)\right)=0$ 其中 $\frac{1}{N}\sum_{n=1}^{N} y_nh_i(x_n)$ 可用22題

的方法算出來，其中RMSE(h(x)=0)可以共用，剩下K次的RMSE(h_i)，總共K+1次。

$\frac{1}{N}\sum_{n=1}^{N}\left(h_i(x_n)\sum_{k=1}^{K} w_kh_k(x_n)\right)$ 是變數為w_1,w_2,...w_K的K維方程式，總共有i=1~K, K個等式

解聯立方程式可算出w1,w_2,...,w_k

總共需要K+1次的RMSE。