ML HW2
姓名: 呂宗澤
學號: R10922A17

程式題我發現我每次測試，結果的變化幅度還滿大的，Q16 的 linear regression 誤差約略上下起伏 0.005 左右，也許是跟我 seed 取自時間有關。Q13、Q15 都很穩，唯一可能出現錯誤的就是計算誤差的方式，但我測試誤差計算函數後也證實能反映出 w 的正確性，總之誤差越小應該是越好(?)，希望助教不要看到結果就把我打錯。

Output:

```
#13, #14
Squar error of linear regression of Error-in  0.2837211363951444
|Error-in - Error-out| of linear regresstion  0.0019640000000000005
#15
Error-out of linear regression  0.05555000000000001
Error-out of logistic regression  0.058904000000000005
#16
Error-out of linear regression 0.07699
Error-out of logistic regression  0.05882
```

手寫的 Q6 是錯的，而 Q5 則是需要額外補充。
我把 Q5- [c]的補充寫在這
Q5 [c]:

設 X = Matrix(5 x 4+1)，代表 4+1 維的 5 筆資料。
此題我要論無法 shatter 任何 X。
我在手寫 Q5 利用 Cramer’s rule，這會牽涉到除以 det(X)，我當時並沒有討論到 det(X) = 0 情形。
簡而言之，若 det(X) 非 0，則用 Cramer’s rule 可以告訴我們 X 無法被 shatter；而當 det(X) = 0，表示有 data 為線性相依，表示必有三筆 data 可以在 4+1 維是共線情形，故同樣無法被 shatter。
所以 X 無論如何，都無法被在 w_0 > 0 限制下被 shatter。

Q6:

|H| =1126，相當於說 H 能表示的 dichotomy 最多只有 1126 種
而當 X = data , |X| = 11 時，要想 shatter X 就需要 2024 種 dichotomy，H 一定無法生成這麼多種 dichotomy。
但若|X|=10，只有 1024 種 dichotomy，H 是有機會的全部生出來的。

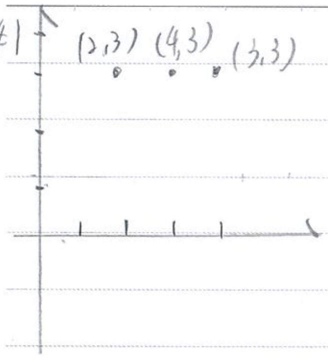
HL - HW2

R10722A17 呂宗澤 HW2

No.: P1

Subject: .... Date: ..../........./..........

#1



<u>Fact 1</u>: 3點共線者無法被 shatter

eg. x o x

以3點為例

Lemma 1: k-dim 中, 任取 k-1 維抓出來看, 若3點在此k-1維可被 shatter

⇒ 在 k-dim 維可被 shatter

eg. $(2,3,4), (4,3,2), (3,3,3)$

取 1 dim & 3 dim 來看: $(2,4), (4,2), (3,3)$ 可被 shatter

⇒ $\exists\, w_1, w_3$ s.t. $\text{sign}(\langle \hat{x}_i, \underbrace{\hat{w}}_{(w_1, w_3)} \rangle) = f(\hat{x}_i)$, $\hat{x}_1 = (2,4)$, $\hat{x}_2 = (4,2)$, $\hat{x}_3 = (3,3)$

設 $w_2 = 0$, 則 (w_1, w_2, w_3) 可分割 $\hat{x}_1, \hat{x}_2, \hat{x}_3$

補充: Lemma I 還須假設 取k-1維 後, 所有點皆不同

Complement of definition:

Data Set $= \{\hat{x}_i\}_{i=1}^{n}$, $\hat{x}_i \in \mathbb{R}^d$; s.t. := such that

Expandition of Lemma I: (Consider $\hat{x} \in \mathbb{R}^3$)

If $\exists\, w = \langle w_1, w_2, w_3 \rangle$ s.t. w can separate $\{\hat{x}_i\}_{i=1}^{n}$

where $\hat{x}_1 = (x_{11}, x_{12}, x_{13})$

⇒ then $\exists\, w' = \langle w_0, w_1, w_2, w_3 \rangle$ s.t. w' can separate $x_0 \oplus \{\hat{x}_i\}$ $x_0 = 1$

i.e. 我說不用管 $x_0 = 1$, 只要能 separate $\{\hat{x}_i\}_{i=1}^{n}$, 則就算 $\hat{x}_i$ 加上 x_0, 也一樣能 separate p2 choose "any $\mathbb{R}^{d-1} \subseteq \mathbb{R}^{d}$"

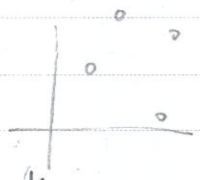
Subject : .... No. : Date : ..../........./..........

#1. Lemma 2: If $\{x_i\}_{i=1}^{n} \subseteq \mathbb{R}^{d-1}$ cannot be shattered, then $\{x_i\}_{i=1}^{n} \subseteq \mathbb{R}^{d}$ cannot be shattered.

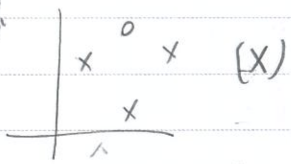
pf: 不会。而且可能根本是錯的。
[a]: 3点太稀 X
[d]: 5点 > 3+1=4点，極有可能是 X
$(d+1)$
[b]: 實際用电脑畫出来檢查，可被 shattered.

Ans: [b] 很難用標準化的定理去證明。

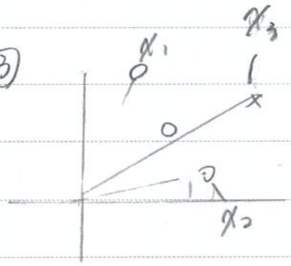
#2. Consider on 2D Euclidean space with case $n=4$
Case1: Four points are on the same 象限
WLOG, let them on first 象限。

① 
ok

I should fix 4 point in principle. However, in order to maximize numbers of possible lines, I will slightly adjust some of their location.

② 

Warning: 這頁解得全部沒用到
直接翻面即可

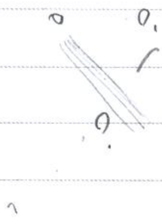
③ 

Obviously their is no any line can separate. We can observe the "range" of possible lines are limited somehow by two point with different sign.

For example, in ③, All the possible line should locate between $(0, x_3)$ and $(0, x_2)$

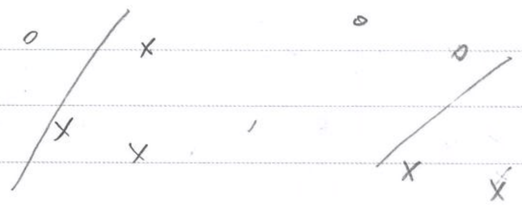
Double A

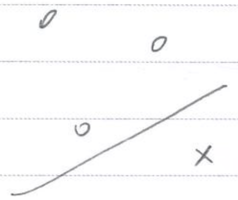
#2 所以換个思維想，先決定2个異号点當起頭，

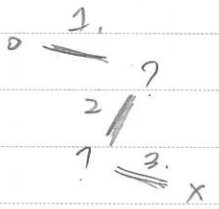


?: possible range of lines.

?: undetermined points.







Observation: 決定好2点字o,x後，剩下可能線段位置為 N-1 个(同时也決定dichotomies)
(当然点的罷放要夠好，讓這N-1种皆能有線穿过。)

∴ 起始2点可為 ox or xo ⇒ 共 2(N-1) 种

若起始2点為 oo or xx?

則 dichotomies 只能是 o,o,...,o or x,x,...,x, (resp.)

⇒ 共 $2(N-1)+2=2N$ 种

#3 $h(x) = \begin{cases} +1 & \text{if } a \le \sum_{i=1}^{d} x_i^2 \le b \\ -1 & \text{o.w.} \end{cases}$

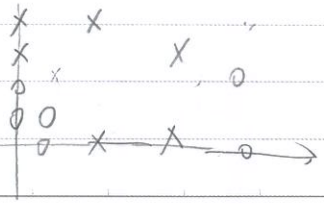
Consider $d=2$ ①

观察後可知：For any input $\{x,y\}$ if $\exists h$ can separate them

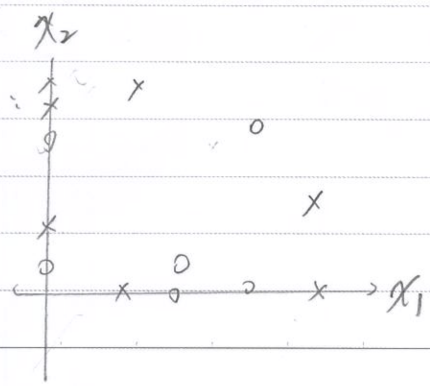
⇒ (x_1,x_2) project on x_1-axis and x_2-axis can be separated by "positive interval hypothesis"

⇒ Grown func. $\le \binom{N+1}{2}+1$

good case:



bad case:



Subject : p4

H_p := positive intervals hypothesis

No. :
Date : ..../........./..........

#3 ② 如果 x_1-projection $\in H_p$，但x_2-projection $\notin H_p$.
$\Rightarrow (x_1, x_2) \notin H_d$. H_d := dount hypothesis.

③ 在(x_1 projection H_p)中的任意(x,y)，皆可調整x_2使 $(x,y) \in H_{dount}$
$\Rightarrow m_{H_d}(N) \geq m_{H_p}(N)$

By ① & ③, $m_{H_d}(N) = m_{H_p}(N) = \binom{N+1}{2} + 1$, $d=2$

Consider $d=3$: Input data (x,y) where $x \in \mathbb{R}^3$
$(x,y) \in H \Rightarrow x \in H$ on x-axis projection
y-axis projection
z-axis projection

$\therefore m(H \text{ on } 3\text{-dim})(N) \leq m(H \text{ on } 2\text{-dim})(N) = \binom{N+1}{2} + 1$
$\Rightarrow m(H \text{ on } 3\text{-dim})(N) = \binom{N+1}{2} + 1$

以此類推, while considering generally case $x \in \mathbb{R}^d$,
$m_H(N) = \binom{N+1}{2} + 1$.

#4 $m_H(N) \in O(N^2)$ $\therefore$ vc = 2

#5
(a) 相当於說給N个点最多只能有2个正区間
For N=4:
O X O X $\Rightarrow$ 最多2个正区間之后
For N=5:
O X O X O $\Rightarrow$ 3个正区間 $\rightarrow$ can't shatter $\therefore$ vc = 4

AM

P5 HW2

Subject : ....

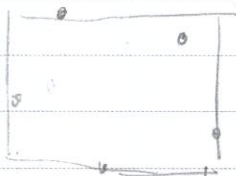
No. :

Date : ..../........./..........

#5 [b] 存在4点可被 shatter $d_{vc}=4$

但5点時，

如果最多圖的



点為正 ⇒ 就会讓 rectangle 含蓋到其它可能為負的点

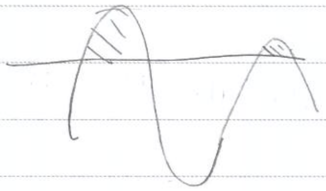
[c] 需在最後一面 $d_{vc}=4$.

Ans: [e]

[d] $h(x)=\text{sign}(w_0+w_1x+w_2x^2+w_3x^3)$

其實就是 [a] 的另种樣貌 ⇒ $d_{vc}=4$

圖示：



▨ : positive interval

Ans: [b]

#6

Fact: ∃ 3点共線的 input data 無法被 shatter

所以以下討論 data 都假設不存在 3点共線

(In other words, $\{x\}$ = input data is a linearly indep. set)

$\forall X \subseteq$ input data, $X := \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_{d+1}^T \end{bmatrix}$, $x_i \in \mathbb{R}^d$

從 data 中任取 d+1 筆

By assumption of Fact, $x_{d+1} \in$ linear combination of $\{x_i\}_{i=1}^{d}$

let $x_{d+1} = a_1x_1 + a_2x_2 + \dots + a_d x_d$

$\Rightarrow w^T x_{d+1} = \underbrace{a_1 w^T x_1}_{>0} + \underbrace{a_2 w^T x_2}_{>0} + \dots + \underbrace{a_d w^T x_d}_{>0}$

(let $a_2, \dots, a_d < 0$ and $w^T x_2, \dots, w^T x_d < 0$)

(let $a_1 > 0$, $w^T x_1 > 0$)

$\Rightarrow w^T x_{d+1} > 0$ $\therefore$ H cannot shatter while $N = d+1$

$\Rightarrow d_{vc} = d$ $(x \in \mathbb{R}^d)$

Double A

Ans: [a] 1/26

Subject : MLHW P6

No. :

Date : ..../........../...........

#7 $P[\exists h \in H \text{ s.t. } |E_{in}(h) - E_{out}(h)| > \varepsilon] \le 2M\exp(-2\varepsilon^2 N)$

$\Updownarrow$

$P[g$ can hornorly reponse E_{out} by $E_{in}(g)] \le 2M\exp(-2\varepsilon^2 N)$

or just says $P[|E_{in}(g) - E_{out}(g)| > \varepsilon] \le 2M\exp(-2\varepsilon^2 N)$

$\Leftrightarrow P[|E_{in}(g) - E_{out}(g)| \le \varepsilon] \ge 1 - 2M \cdot \exp(-2\varepsilon^2 N)$

Let $\delta = 2M \cdot \exp(-2\varepsilon^2 N)$

then $\frac{\delta}{2M} = \exp(-2\varepsilon^2 N) \Rightarrow \ln\left(\frac{\delta}{2M}\right) = -2\varepsilon^2 N$;

$\Rightarrow \ln\frac{2M}{\delta} = 2\varepsilon^2 N \Rightarrow \sqrt{\frac{1}{2N}\ln\frac{2M}{\delta}} = \varepsilon$

Conclude that $P[|E_{in}(g) - E_{out}(g)| \le \varepsilon] \ge 1-\delta$

where $\varepsilon = \sqrt{\frac{1}{2N}\ln\frac{2M}{\delta}}$

$\Rightarrow$ With probability more than $1-\delta$, we have

$$-\varepsilon \le E_{in}(g) - E_{out}(g) \le \varepsilon$$

$$\Rightarrow E_{in}(g) - \varepsilon \le E_{out}(g) \le E_{in}(g) + \varepsilon;$$

Similarly, we can get $P[|E_{in}(g^*) - E_{out}(g^*)| < \varepsilon] \ge 1 - \frac{\delta}{M}$

from Hoffding bound but just fix $h = g^*$.

Note that probability of $1 - \frac{\delta}{M} > 1 - \delta$

Conclusion: With probability more that $1-\delta$, we have

$E_{in}(g) - \varepsilon \le E_{out}(g) \le E_{in}(g) + \varepsilon$

$E_{in}(g^*) - \varepsilon \le E_{out}(g^*) \le E_{in}(g^*) + \varepsilon$

$\Rightarrow E_{out}(g) - E_{out}(g^*) \le \underbrace{E_{in}(g) - E_{in}(g^*)}_{<0} + 2\varepsilon$ Ans: [C] 2ε

Double A

MLHW P7

Subject: ...... No.: 吕宗澤 Date: ......

#8 $d_{vc}=1$, $P[\text{bad data}] \le 4\,(2N)^{d_{vc}}\exp(-\frac{1}{8}0.01N)$

$\Rightarrow 8N\exp(-\frac{1}{800}N)=\delta$

$\Rightarrow$ 用电脑将[a]~[e]由小至大代入N

得 $N=11000$ 時，$\delta=0.09$ Ans: [b]

#9

$E(w)\approx E(u)+b_E(u)^T(w-u)+\frac{1}{2}(w-u)^T A_E(u)(w-u)$

想法：將v視為 $v(t)=\begin{pmatrix}v_1(t)\\ \vdots \\ v_n(t)\end{pmatrix}$，$v(t)$代表在$\mathbb{R}^n$中隨t=time 用向量 $\begin{pmatrix}v_1(t)\\ \vdots \\ v_n(t)\end{pmatrix}$ 移动的点

$w \leftarrow u+v \Leftrightarrow v=w-u$

題目所求為 Taylor's expansion 的誤差之极最小值

$\Rightarrow$ 先將誤差 func. 化為对v变数 再去求 critical point

Let G ...

① $E(u+v)-E(u)=b_E(u)^T(v)+\frac{1}{2}(v)^T A_E(u)(v)$

constants constant

$\frac{\partial}{\partial v}[E(u+v)-E(u)]=b_E(u)+A_E(u)v = 0$

Wanted

$\Rightarrow v=-(A_E(u))^{-1}b_E(u)$ is a critical point.

① 計算有誤，更於背面

② Commit that $v=-(A_E(u))^{-1}b_E(u)$

最便宜的作法一直接代入，或是去計算 $E(u+v)-E(u)$ 的 Hessian 是否為 positive define.

正確來說應是計算 $\text{Norm}_2(E(u+v)-E(u))$ 的 Hessian

$\text{Norm}_2: \mathbb{R}^n \longrightarrow \mathbb{R}$

HLHW2

Subject : P.8   No. :

Date : ..../........./..........

#9-② $\frac{d}{dv}\left(Norm_2\left[E(u+v)-E(u)\right]\right)$

$= Norm_2'\left(E(u+v)-E(u)\right) \cdot E'(u+v)$ ← inner product

$Norm_2(x_1,x_2,\dots,x_n) = (x_1^2+x_2^2+\dots+x_n^2)^{\frac{1}{2}}$

$Norm_2'(x_1,\dots,x_n) = \vec{\nabla} Norm_2(x_1,\dots,x_n) = (x_1,x_2,\dots,x_n)$

$\therefore \frac{d}{dv}\left(Norm_2\left(E(u+v)-E(u)\right)\right)$

$= \left[E(u+v)-E(u)\right] \cdot \underline{E'(u+v)}$ → 0

$\hookrightarrow -(A_E(u)^{-1} b_E(u))$

① 我應該將 $E(u+v)-E(u)$ 套入 $Norm_2$ 再去求 critical point,

不过結果可看出 $-(A_E(u)^{-1} b_E(u))$ 仍為唯一可能。

如果是 CS 学生，其實可以直接用电腦不斷丟数字去驗證。我手很酸了，就寫到這。

#10 $E_{in}: \mathbb{R}^n \to \mathbb{R}$; $E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N} \ln\left(1+\exp(-y_n w^T x_n)\right)$

方便观察，設 $N=2$：

$E_{in}(w) = \frac{1}{2}\left[\ln(1+\exp(-y_1 w^T x_1)) + \ln(1+\exp(-y_2 w^T x_2))\right]$

$\Rightarrow \vec{\nabla} E_{in}(w) = \frac{1}{2}\left[\frac{\exp(-y_1 w^T x_1)}{1+\exp(-y_1 w^T x_1)}(-y_1 x_1) + \frac{\exp(-y_2 w^T x_2)}{1+\exp(-y_2 w^T x_2)}(-y_2 x_2)\right]$

$= \frac{1}{2}\left[\theta(-y_1 w^T x_1)(-y_1(x_{11},x_{12})) + \theta(-y_2 w^T x_2)(-y_2(x_{21},x_{22}))\right]$

$= \frac{1}{N}\left[\sum_{n=1}^{N} \theta(-y_n w^T x_n)(-y_n x_n)\right] = \left(\frac{\partial E_{in}}{\partial w_1}, \frac{\partial E_{in}}{\partial w_2}\right)$

↓ $N=2$

#10 $\left(\frac{\partial E_{in}}{\partial w_1}, \frac{\partial E_{in}}{\partial w_2}\right) = \frac{1}{N}\left[\sum_{n=1}^{N} \theta(-y_n w^T x_n)(-y_n x_n)\right]$, $x_n \in \mathbb{R}^2$, $N=2$

$$\left(\frac{\partial^2 E_{in}}{\partial w_1^2}, \frac{\partial E_{in}}{\partial w_1 \partial w_2}\right) = \frac{\partial}{\partial w_1}\left(\frac{\partial E_{in}}{\partial w_1}, \frac{\partial E_{in}}{\partial w_2}\right)$$

$$= \frac{1}{N}\sum_{n=1}^{N} \frac{\partial \theta(-y_n w^T x_n)}{\partial(-y_n w^T x_n)} \frac{\partial(-y_n w^T x_n)}{\partial w_1} \cdot (-y_n x_n)$$

Compute $\frac{\partial \theta}{\partial s} = \frac{d}{ds}\theta(s) = \frac{d}{ds}\left(\frac{1}{1+e^{-s}}\right) = \frac{e^{-s}}{(1+e^{-s})^2}$

Apply s to $-y_n w^T x_n$, then we have

$$= \frac{1}{N}\sum_{n=1}^{N} \frac{\exp(y_n w^T x_n)}{[1+\exp(y_n w^T x_n)]^2} \cdot (-y_n x_{n1})(-y_n x_n)$$

$$= \frac{1}{N}\sum_{n=1}^{N} \theta(-y_n w^T x_n) \cdot h(y_n w^T x_n) \cdot (\cancel{-y_n} x_{n1})(\cancel{-y_n} x_n)$$

$-y_n \cdot -y_n = 1$

同時推得 $\frac{\partial}{\partial w_2}\left(\frac{\partial E_{in}}{\partial w_1}, \frac{\partial E_{in}}{\partial w_2}\right)$

$$= \frac{1}{N}\sum_{n=1}^{N} \theta(-y_n w^T x_n) h(y_n w^T x_n)(x_{n2} x_n)$$

$x_n \in \mathbb{R}^2$

$$\therefore \text{Hessian}(E_{in}) = \left[\frac{1}{2}\sum_{n=1}^{2} \frac{\exp(y_n w^T x_n)}{[1+\exp(y_n w^T x_n)]^2} \begin{matrix} x_{n1} x_n \\ \updownarrow \\ (x_{n2} x_n) \end{matrix}\right]$$

$$\frac{\exp(y_n w^T x_n)}{[1+\exp(y_n w^T x_n)]^2} = \frac{\exp(y_n w^T x_n)}{1+\exp(y_n w^T x_n)} \frac{1}{1+\exp(y_n w^T x_n)}$$

$$\frac{1}{1+\exp(-y_n w^T x_n)} \frac{1}{h_t(y_n x_n)} = h_t(-y_n x_n)\, h_t(y_n x_n)$$

Ans: $\nabla^2 E_{in} = \frac{1}{N}\sum_{n=1}^{N} [h_t(y_n x_n)][h_t(-y_n x_n)]\, x_n x_n^T$

HL HW2

P10

Subject : .... No. : Date : ..../........../..........

#11 $X = U\Sigma V^T$, $X^\dagger = V\Sigma^\dagger U^T$

[a] 直將用 $X = U\Sigma V^T$ 代入計算

$(X^TX)^{-1} = (V\Sigma U^T U \Sigma V^T) = (V\Sigma^2 V^T)^{-1}$ (如果 $(X^TX)^{-1}$ exists)

$\because (V\Sigma^2 V^T)^{-1}(V\Sigma^{-2} V^T) = I_{d+1}$

$\therefore (X^TX)^{-1} = V\Sigma^{-2}V^T$

$\therefore (X^TX)^{-1}X^T = (V\Sigma^{-2}V^T)(V\Sigma U^T) = V\Sigma^{-1}U^T = X^\dagger$

[b] 同樣代入計算

$(XX^\dagger)^k = (U\Sigma V^T \cdot V\Sigma^\dagger U^T)^k = (U\Sigma\Sigma^\dagger U^T)^k$

$\Sigma\Sigma^\dagger = \begin{pmatrix} I_j & \\ & \ddots 0 \end{pmatrix}_{N\times N}$ j = rank of X

$\Rightarrow U\begin{bmatrix} I_j & \\ & 0 \end{bmatrix}U^T = [col_1(U), col_2(U), \ldots, col_j(U), 0, \ldots, 0]\, U^T$

$= [col_1(U), \ldots, col_j(U), 0, \ldots, 0]\begin{bmatrix} col_1(U) \\ \vdots \\ col_j(U) \\ \vdots \\ col_N(U) \end{bmatrix} = [col_1(U), \ldots, col_j(U)]\begin{bmatrix} col_1(U) \\ \vdots \\ col_j(U) \end{bmatrix}$

$= U_1U_1^T \in j\times j$ matrix, $j = \text{rank}(X)$, $U = [U_1, U_2]$

Note: $\text{trace}(X) = \text{trace}(U_1U_1^T) = \text{trace}(U_1^TU_1) = j = \text{rank}(X)$

Claim $U_1U_1^T$ is a projection matrix

$(U_1U_1^T)^2 = U_1\underbrace{U_1^TU_1}_{=I_j}U_1^T = U_1U_1^T \in j\times j$ matrix

結論：只有當 rank(X)=n 時 [c] 才為 True.

Double A

ML HW2 P11

Subject :

No. :

Date : ..../........../...........

#12 Given x_n, then $p(y_n | x_n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y_n - w^T x_n}{\sigma}\right)^2\right)$

∴ Given data set $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$

$$\underset{(y_1,\dots,y_N)}{\underset{\parallel}{p(\underset{\sim}{y}}} \,|\, \underset{(x_1,\dots,x_N)}{\underset{\parallel}{\underset{\sim}{x}}}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\underset{\sim}{y} - w^T \underset{\sim}{x}}{\sigma}\right)^2\right)$$

where $w^T = [\;—\;]_{1 \times d+1}$, $\underset{\sim}{x} = \overbrace{\begin{bmatrix} & \\ & \end{bmatrix}}^{N}_{d+1 \cdot N}$

只要压制 $\exp\left(-\frac{1}{2}\left(\frac{\underset{\sim}{y} - w^T \underset{\sim}{x}}{\sigma}\right)^2\right) \Leftrightarrow p(\underset{\sim}{y} | \underset{\sim}{x}) \uparrow\uparrow$

$\Updownarrow$

minimize $(y - w^T x)^2$, $\Leftrightarrow$ Use linear regression

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

補充:

#5-[c] Positively-biased perceptrons over $x \in \mathbb{R}^4$, which contains perceptrons with $w_0 > 0$

$\underset{(1, x_1, x_2, x_3, x_4)}{\underset{\parallel}{x \in \mathbb{R}^5}}$ ∴ $\underset{(w_0 > 0, w_1, w_2, w_3, w_4)}{\underset{\parallel}{w \in \mathbb{R}^5}}$

Claim: 5 input cannot be shattered by X

For any given X, $Xw = y$, for some un-determined y and w.

Use Cramer's rule to present w_0:

$w_0 = \frac{\det(M_0)}{\det(X)}$ by Cramer's rule, where $M_0 = \begin{bmatrix} y_0 & & & & \\ y_1 & & & & \\ y_2 & & & & \\ y_3 & & & & \\ y_4 & & & & \end{bmatrix}$

$\det(X)$ is a constant.

$\det(M_0) = C_{11} y_0 + C_{12} y_1 + C_{13} y_2 + C_{14} y_3 + C_{15} y_4$,

$\{C_{1j}\}_{j=1,\dots,5}$ are cofactor of X, they are constants.

2st ~ 5th column of X

Double A

ML Hw2

Subject : p12 No. : Date : ..../......../........

#5-[c] Following above statement...

$$w_0 = \frac{\det(M_0)}{\det(X)} = \frac{C_{11}}{\det(X)}y_0 + \frac{C_{12}}{\det(X)}y_1 + \frac{C_{13}}{\det(X)}y_2 + \frac{C_{14}}{\det(X)}y_3 + \frac{C_{15}}{\det(X)}y_4$$

constants

$\left\{\frac{C_{1j}}{\det(X)}\right\}$ are indepent to y !

For any X, let $y_j = -\text{sign}\left(\frac{C_{1j}}{\det(X)}\right)$, then $\nexists$ w s.t. $\text{sign}(Xw) = \text{sign}(y)$

Note: ① y_0, y_1, y_2, y_3, y_4 下標是指data編号,正確寫法應是 $y_{1\sim5}$

② 正確寫法應該是"5 input cannot be shattered by H."
$H = \{h \mid h = (1, x_1, x_2, x_3, x_4) \in \mathbb{R}^5\}$

Clam $\exists$ input with size 4 can be shattered by X !

Let $X = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$, $Xw = X\begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} = w_0\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}$

$\forall y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$, let $w = \begin{bmatrix} 1 \\ y_1 - 1 - w_1 \\ y_2 - 1 - w_2 \\ y_3 - 1 - w_3 \\ y_4 - 1 - w_4 \end{bmatrix}$, then $Xw = y$

Double A

Subject : ....   matrix ✓   No. :
Date : ..../........./..........

先備觀念: Given any $X \in \text{Matrix}(N \times M)$

① # of distinct non-zero eigenvalues of X = rank(X)

Note: $Xv = 0 = 0v$, 0也是种 eigenvalue

② $\forall$ non-zero eigenvalue λ of XX^* (or X^*X)

We have $\lambda > 0$

$\Downarrow$

$\exists\, \sigma > 0$ s.t. $\sigma^2 = \lambda$

③ If U is a unitary matrix (By def., $UU^* = U^*U = I$)
$U \in \text{Matrix}(N \times N)$

then we can denote U as $[\,u_1 \,|\, u_2 \,|\, \cdots \,|\, u_N\,]$

where $\{u_1, u_2, \ldots, u_N\}$ is othonormal basis

先將U寫成 $[\,u_1 \,|\, u_2, \ldots, u_N\,]$, 会發現 $UU^* =$

$$[\,u_1 | u_2 \cdots | u_N\,] \begin{bmatrix} -u_1^T- \\ -u_2^T- \\ \vdots \\ -u_N^T- \end{bmatrix} \overset{?}{=} \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}_{N \times N}$$

If

$\Uparrow$

$$\langle u_i, u_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

回頭看 SVD 的結果

Subject : .... No. : Date : ..../........../..........

For any matrix $X \in \text{Matrix}(N \times M)$, $\exists\, U \in \text{Matrix}(M \times M)$, $V \in \text{Matrix}(N \times N)$

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & 0 \ddots \end{bmatrix}_{M \times N}$$

with U: unitary, V: unitary,

s.t.

$$X = U \Sigma V^T$$

$V = \{v_1, v_2, \dots, v_M\}$, $U = \{u_1, u_2, \dots, u_N\}$

$\Downarrow$

$$X v_k = U \Sigma V^T v_k = U \Sigma \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v^T \end{bmatrix} \begin{bmatrix} v_k \end{bmatrix} = U \Sigma \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_N \leftarrow k$$

$1 \le k \le M$

$$= U \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots \sigma_r & \\ & & & 0 \ddots \end{bmatrix}_{M \times N} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_N \leftarrow k$$

(忘了說，事實上 $\sigma = \sqrt{\lambda}$，λ 為 $X^* X$ 的 eigenvalue)

$X^* X$ 最多有 r 個 $\lambda > 0$，我們只看這些 λ，使 $\sigma = \sqrt{\lambda}$

r: $\text{rank}(X^* X) = \text{rank}(X)$

$$= U \begin{bmatrix} \Sigma \text{ 第 } k \text{ 行} \end{bmatrix} =$$

如果 $k \le \text{rank}(X) = r$ 呢？ $\Rightarrow$ Σ 第 k 行為 $\begin{bmatrix} 0 \\ \vdots \\ \sigma_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow k$

($\text{rank}(X) = \text{rank}(X^* X)$)

$$\Rightarrow X v_k = U \begin{bmatrix} 0 \\ \vdots \\ \sigma_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma_k u_k$$

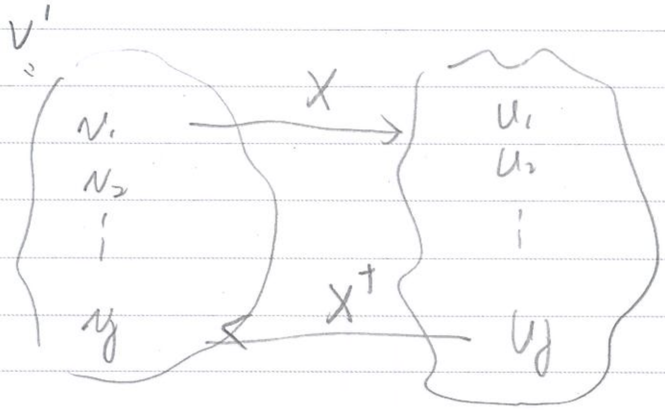
反之，若 $k > \text{rank}(X) = r$

$$\Rightarrow X v_k = 0$$

這是不是在說，当我們討論 $span(v_1, ..., v_j)$ 與 $span(u_1, ..., u_j)$時

$V \supseteq V' = span(v_1, ..., v_j)$　　　$U \supseteq U' = span(u_1, ..., u_j)$

有類似X-inverse的存在呢?



其實我們仍無法完美的用$X^\dagger$解 $Xw=y$ 問題

正確來說

$y \in span(u_1, ..., u_j) \xrightarrow{X^\dagger}$ 可以用$X^\dagger$推回 $w_1, ..., w_j$

$$y = \sum_{1 \le k \le j} u_k + c_{j+1}u_{j+1} + ... + c_N u_N$$

$\sum_{1 \le k \le j} u_k \Big\downarrow X^\dagger$ 可推回$X^\dagger$

$c_{j+1}u_{j+1} + ... + c_N u_N$：存在c非0 $\Big\downarrow X^\dagger$ $X^\dagger(u_k) = 0$, for $k > j$

用$X^\dagger$推回的只有部份的v

$u_k > 0$全部變為0