

### Problem 2.1

$$\text{Equation (2.1)} \Rightarrow \epsilon(M, N, \delta) = \sqrt{\frac{1}{2N} \ln \left( \frac{2M}{\delta} \right)} < 0.05.$$

Given  $\delta = 0.03$ .

$$\text{for } \epsilon \geq 0.05, \text{ we get } \sqrt{\frac{1}{2N} \ln \left( \frac{2M}{\delta} \right)} \leq 0.05.$$

$$\textcircled{1} M = 1$$

$$0.05 \geq \sqrt{\frac{1}{2N} \ln \left( \frac{2M}{\delta} \right)}$$

$$0.05 \geq \sqrt{\frac{1}{2N} \ln \left( \frac{2}{0.03} \right)}$$

$$0.05 \geq \sqrt{\frac{1}{2N} \times 4.1997}$$

$$\therefore \text{Squaring both sides} \Rightarrow 25 \times 10^{-4} \geq \frac{1}{2N} \times 4.1997$$

$$\Rightarrow N \geq \frac{1 \times 4.1997 \times 10^4}{2 \times 25}$$

$$\Rightarrow 41997$$

so

$$\geq 41997$$

$$\geq 839.94$$

$\therefore$  We need 839.94 examples to make  $\epsilon \leq 0.05$  for  $M=1$

$$\textcircled{2} M = 100 \quad 0.05 \geq \sqrt{\frac{1}{2N} \ln \left( \frac{200}{0.03} \right)} \Rightarrow 0.05 \geq \sqrt{\frac{1}{2N} (8.8048)}$$

Squaring both sides,

$$25 \times 10^{-4} \geq \frac{1}{2N} (8.8048)$$

$$\Rightarrow N \geq \frac{8.8048 \times 10^4}{2 \times 25}$$

$\therefore$  We need 1760.96 examples

so

③  $M = 10,000$

$$0.05 \geq \sqrt{\frac{1}{2N} \ln \left( \frac{2M}{\delta} \right)}$$

$$25 \times 10^{-4} \geq \frac{1}{2N} \ln \left( \frac{2 \times 10^4}{0.05} \right)$$

$$25 \times 10^{-4} \geq \frac{1}{2N} \left( 13.4100 \right) \Rightarrow N \geq \frac{13.4100 \times 10^4}{50}$$

$$\therefore N \geq \frac{134100}{50} \Rightarrow N \geq 2682$$

$\boxed{\therefore \text{No. of examples} = 2682}$

### Problem 2-3

(a) Positive or Negative Ray:-

Growth function for positive rays =  $N+1$

for negative rays, all the dichotomies will be doubled.  
as we have ' $+1$ ' for ' $-1$ ' in ~~positive~~ rays.

For negative & positive common case: All  $+1$  & All  $-1$ .

$\therefore$  subtract 2 dichotomies got from negative rays

maximum as we already have it counted in positive rays.

i. Dichotomies for positive/negative rays:-

$$= \text{dichotomies for +ve rays} + \text{dichotomies for -ve rays} - 2.$$

$$= (N+1) + (N+1) - 2$$

$$= N+1 + N-1$$

$$= 2N.$$

$$\therefore m_x(N) = 2N$$

Now largest value of  $N$  for which  $m_x(N) = 2^N$ ,

$$\therefore N=2, m_x(2)=4 \text{ & Growth f}^h = 2N=4$$

$$\text{For } N=3, m_x(3)=8 \text{ but we get } m_x(2)>6.$$

$$\boxed{\text{VC dimension} = 2}$$

(b) Positive or Negative Intervals:-

for positive intervals,  $m_x(N) = \frac{N^2}{2} + \frac{N}{2} + 1$

for negative intervals, we add only (+1, -1, H)

dichotomy, for  $N=4$  we add only (+1, +1, -1, +1)  
& (+1, -1, +1, +1).

Thus we say, we add  $N-2$  dichotomies for negative intervals.

for  $N=1$ , we already get 2 dichotomies covered in +ve intervals, hence no negative dichotomies are different to add.

$$\therefore \boxed{\text{for } N=1, m_x(N)=2 \text{ & } N>1, \frac{N^2}{2} + \frac{N}{2} + 1 + N-2} \\ = \frac{N^2}{2} + \frac{3N}{2} - 1$$

Vcdimension = largest value of  $N$  for which  $m_x(N) = 2^N$

$$\therefore N=2, m_x(2) = \frac{4}{2} + \frac{2}{2} - 1 = 7 \quad \therefore N=8, m_x(8) = \frac{64}{2} + \frac{8}{2} - 1$$

$$2^4 = 4! = 7 \quad 2^8 = 2^3 = 8$$

$$\boxed{\text{Vcdimension} = 3}$$

(Q)

Problem 23 (C)

(C)

we have function, +1 for  $a \leq \sqrt{x_1^2 + \dots + x_d^2} \leq b$ .

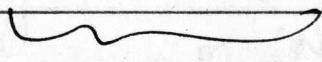
∴ we can say, we have to select  $[a, b]$  where  $\{\sqrt{x_1^2 + \dots + x_d^2}\}$  lies. This can be mapped again to our positive intervals dichotomies.

∴ for this circle equation  $x_1^2 + x_2^2 = r^2$

$$\therefore \sqrt{x_1^2 + x_2^2} = r.$$

⇒ that means,  $\sqrt{x_1^2 + \dots + x_d^2} \rightarrow 'r'$  between  $[a, b]$

$$\Rightarrow a \leq r \leq b.$$

⇒ 

positive interval where ' $r$ ' →  $[0, \infty]$ .

∴ maximum dichotomies of '(C)' ⇒ maximum dichotomies for positive intervals.

$$= \frac{N^2}{2} + \frac{N}{2} + 1$$

∴  $V_c$  dimension = largest  $N$  for which  $m_N(N) = 2^N$

$$\text{our } m_N(N) = \frac{N^2}{2} + \frac{N}{2} + 1$$

$$\therefore N^2, \quad \approx 2^2 \times 2^2 + \frac{2^2}{2} + 1 = 4 \quad 2^N = 4 \text{ equal.}$$

$$\therefore \text{Ans. } m_{N=3}(N) = \frac{3^2}{2} + \frac{3}{2} + 1 = 8 \quad (N=3) \quad 2^3 = 8 \quad \text{and } m_{N=3}(N) = 8.$$

$$\therefore V_c \text{ dimension} = 3$$

$$\text{max. equal. } \frac{9}{2} + \frac{3}{2} + 1 = 8$$

$$\text{for } N=4, 2^4 = 16 \quad ) \text{Not equal.}$$

$$m_{N=4}(N) = \frac{16}{2} + \frac{4}{2} + 1 = 11 \quad ) \text{Not equal.}$$

problem 2.8

for growth-function  $m_x(N)$  we know two conditions,

① we get  $V_c$  dimension ( $d_{rc}$ ) when we have largest  $N'$  for which  $m_{x'}(N) = 2^{N'}$ .

② Also,  $m_x(N)$  is bounded by  $N^{d_{rc}} + 1$  which implies.  
 $\Rightarrow m_x(N) \leq N^{d_{rc}} + 1$  for (N).

∴ for each we will calculate  $d_{rc}$  & see if it is  $\leq N^{d_{rc}} + 1$

(a) For (N=1),

putting  $N=1$ ,  $2^N = 2 - ①$

obviously  $m_x(N) = N+1 = 2 - ②$ .

∴ ① = ② which satisfies first condition

putting  $N=2$ ,  $2^N = 4 - ①$

$m_x(2) = 2+1 = 3 - ②$ .

∴ ① ≠ ② ∴ largest  $N=1 \Rightarrow d_{rc} = 1$

Now check the bound  $N^{d_{rc}} + 1$ ,

for  $N=1$ ,  $N^{d_{rc}} + 1 = 1^1 + 1 = 2 \therefore$  true as  $N+1=2$

for  $N=2$ ,  $N^{d_{rc}} + 1 = 2^1 + 1 = 3 \therefore$  true as  $N+1=3$ .

Hence for all 'N' we get the bound also true.

$\Rightarrow$   $N+1$  is a possible growth function

$$b) 1 + N + \frac{N(N-1)}{2} \Rightarrow 1 + N + \frac{N^2 - N}{2}$$

$$\Rightarrow \frac{N^2}{2} + \frac{N}{2} + 1$$

1<sup>st</sup> condition  $\rightarrow$  get finite  $dvc$  where  $m_x(N) = 2^N$ .

$\therefore dvc = 2 \dots$  positive intervals.

2<sup>nd</sup> condition  $\rightarrow$  It satisfies  $m_x(N) = \frac{N^2}{2} + \frac{N}{2} + 1 \leq N^{dvc} + 1$   
for all 'N'

Hence 'b' is also a possible growth function.

(c)  $2^N$   $\rightarrow$  we know that for every 'N' we get  $m_x(N) = 2^N$   
as our growth function is  $2^N$ .  
 $\therefore dvc = \infty \dots$  This is from convex circles.

$\therefore$  This is also a possible growth function.

(d)  $2^{\sqrt{N}}$

1<sup>st</sup> condition  $\Rightarrow$  Get  $dvc$ , in this case  $dvc = 1$ ,  
for all N,

2<sup>nd</sup> condition  $\Rightarrow$  should be bounded by  $N^{dvc+1}$ ; i.e.  $N^{(1)}+1$

$$\therefore \text{Take } N=1, m_x(N) = 2^{\sqrt{1}} = 2 \quad \boxed{1+1=2} \quad \text{true}$$

$$\cdot \quad N=4, m_x(N) = 2^{\sqrt{4}} = 4 \quad \boxed{4+1=5} \quad \text{true}$$

$$\text{For } N=25, m_x(25) = 2^{\sqrt{25}} = 32 \quad \boxed{\text{Here } m_x(25) < N^{(1)+1}} \\ \& N^{(1)}+1 = 26$$

possible

∴ (d) is not a growth function

(e)  $2^{(N^2)}$  →  $d_{rc} = 0$ , must be bounded by  $N^0 + 1 = 2$ ,

for  $N=4$ ,  $m_N(N) = 2^{N/2} = 4$  ] not true.  
 $\& N^{d_{rc}} + 1 = 2$

Hence this is not a possible growth function

$$(f) 1 + N + \frac{(N)(N-1)(N-2)}{6}$$

$$d_{rc} = 1 \quad (m_{rc}(N) = m_{rc}(2) = 3 \neq 2^2)$$

$$m_{rc}(1) = 2 = 2^{(1)} \Rightarrow d_{rc} = 1.$$

Now it should be bounded by  $N^{d_{rc}} + 1$  for all  $Ns$ .

$$\therefore \text{bound} = N^1 + 1 = N + 1$$

$$\text{for } N=1, N+1 = 2 \& m_{rc}(N) = 2 \rightarrow \text{true.}$$

$$N=2, N+1 = 3 \& m_{rc}(2) = 3 \rightarrow \text{true.}$$

$$N=3 \quad N+1=4 \quad \& m_{rc}(3) = 8 \Rightarrow \text{Not true.}$$

Hence this is also not a possible growth function

problem 2.12

As from book we know,

$$N = \frac{8}{\epsilon^2} \ln \left[ \frac{4[(2N)^{\epsilon} + 1]}{8} \right]$$

where  $\epsilon$  = generalization error

$$\epsilon \text{ loss} = 1 - \text{confidence}(.) = \frac{\delta}{100}$$

Given,

$$\therefore dvc = 10$$

$$\epsilon = 0.05$$

$$\delta = 0.05 \quad \dots \quad (1 - 0.95)$$

We can compute this iteratively, by replacing RHS,  $N$ .

Taking  $N = 1000$ , in RHS,

$$N = \frac{8}{(0.05)^2} \ln \left[ \frac{4((2000)^{10} + 1)}{0.05} \right]$$

$$\therefore N = 257251.3$$

257251.3

Taking  $N = 257251.3$ , in RHS,

$$N = \frac{8}{(0.05)^2} \ln \left[ \frac{4((2 \times 257251.3)^{10} + 1)}{0.05} \right]$$

$$N = 434853.08$$

And so on ...

Question 6

Prove that selecting the hypothesis  $h$  that maximizes the likelihood  $\prod_{n=1}^N p(y_n | x_n)$  is equivalent to minimizing the cross entropy error

$$E(\omega) = -\sum_{n=1}^N \ln(1 + e^{-y_n w^T x_n})$$

c) Introduce standard error measure used in logistic regression is based on likelihood.

$$\text{Ans } f(x) = p[y = +1 | x].$$

$$\Rightarrow p(y|x) = \begin{cases} f(x) & \text{for } y = +1 \\ 1 - f(x) & y = -1 \end{cases} \quad \text{--- (1)}$$

$$\therefore \text{Based on (1), we say, } p(y|x) = \begin{cases} h(x) & y = +1 \\ 1 - h(x) & y = -1 \end{cases}$$

$$\text{We have } h(x) = \sigma(w^T x), \text{ also, } 1 - \sigma(s) = \sigma(-s)$$

$$\text{where } \sigma(s) = \frac{e^s}{1 + e^s}$$

Since  $(x_1, y_1), \dots, (x_N, y_N)$  are independently generated, the probability of getting all 'y's' given 'x' will be the product.

$$\Rightarrow \prod_{n=1}^N p(y_n | x_n). \quad \text{--- (2)}$$

The method of maximum likelihood selects hypothesis 'h' which maximizes this probability (2).

∴ multiplying,  $-\frac{1}{N}$  & taking  $\ln$  with <sup>as</sup> monotonically decreasing function.

∴ Now we can minimize,

$$-\frac{1}{N} \ln \left( \prod_{n=1}^N P(y_n | x_n) \right)$$

Getting the '-' sign inside, inverts the probability.

$$\Delta \Pi \Rightarrow \Sigma$$

∴ Now we minimize,

$$\frac{1}{N} \sum_{n=1}^N \ln \left( \frac{1}{P(y_n | x_n)} \right).$$

$$\Rightarrow \frac{1}{N} \sum_{n=1}^N \ln \left( \frac{1}{\sigma(y_n w^T x_n)} \right)$$

$$\therefore \sigma(y_n w^T x_n) = \frac{e^{y_n w^T x_n}}{1 + e^{y_n w^T x_n}}$$

$$\frac{1}{N} \sum_{n=1}^N \ln \left( \frac{1 + e^{y_n w^T x_n}}{e^{y_n w^T x_n}} \right)$$

$$\Rightarrow \frac{1}{N} \sum_{n=1}^N \ln \left( e^{-y_n w^T x_n} + 1 \right) \quad \dots \text{Divided by } e$$

$$\Rightarrow \boxed{\frac{1}{N} \sum_{n=1}^N \ln \left( 1 + e^{-y_n w^T x_n} \right)}$$

$y_n w^T x_n$   
to numerator &  
denominator

Hence proved.

## Problem 7

(7)

$\nabla E_{in}(w(t)) = ?$  in gradient descent algorithm.

$$E_{in}(w(t)) = \frac{1}{N} \sum_{n=1}^N \ln(1 + e^{-y_n w^T x_n})$$

$$\frac{\nabla E_{in}(w(H))}{\partial w} = \frac{1}{N} \sum_{n=1}^N \frac{\partial [\ln(1 + e^{-y_n w^T x_n})]}{\partial w}$$

$$\begin{aligned} \frac{\partial [\ln(1 + e^{-y_n w^T x_n})]}{\partial w} &= \left( \frac{1}{1 + e^{-y_n w^T x_n}} \right) \times \left( \frac{\partial e^{-y_n w^T x_n}}{\partial w} \right) + \frac{\partial (1 - e^{-y_n w^T x_n})}{\partial w} \\ &= \frac{e^{-y_n w^T x_n} \cdot (-y_n x_n)}{1 + e^{-y_n w^T x_n}} \end{aligned}$$

= multiplying by  $e^{y_n w^T x_n}$  to numerator  
& denominator.

①  $\Rightarrow$  is chain rule of derivation.

$$= \frac{(1) (-y_n x_n)}{e^{y_n w^T x_n} + 1} = \left( \frac{-y_n x_n}{1 + e^{y_n w^T x_n}} \right).$$

putting this in  $\nabla E_{in}(w(H))$

$$\nabla E_{in}(w(H)) = \frac{1}{N} \sum_{n=1}^N \left( \frac{-y_n x_n}{1 + e^{y_n w^T x_n}} \right).$$

$$\Rightarrow \boxed{-\frac{1}{N} \sum_{n=1}^N \frac{y_n x_n}{1 + e^{y_n w^T x_n}}}$$

8) Exercise 3.6 [cross entropy error measure]

(a) We know that,  $\prod_{n=1}^N p(y_n | x_n)$  maximizing this is equal to minimizing  $\frac{1}{N} \sum_{n=1}^N \ln \left( \frac{1}{p(y_n | x_n)} \right)$ .

& for  $y_n = +1$  -- we can divide this for  $y_n = +1$   
&  $y_n = -1$ .

$\therefore$  for  $y_n = +1$  maximizing  $\prod_{n=1}^N p(y_n = +1 | x_n)$   
 $\therefore y_n = -1$

$$= \frac{1}{N} \left[ p(y_n = +1 | x_n) + \right.$$

$$\left. p(y_n = -1 | x_n) \right]$$

$$\Rightarrow \frac{1}{N} \sum_{n=1}^N \ln \left[ \left( \frac{1}{p(y_n = +1 | x_n)} \right) + \left( \frac{1}{p(y_n = -1 | x_n)} \right) \right]$$

$$\Rightarrow \frac{1}{N} \sum_{n=1}^N \left[ \ln \left( \frac{1}{p(y_n = +1 | x_n)} \right) + \ln \left( \frac{1}{p(y_n = -1 | x_n)} \right) \right]$$

We know,  
for  $\Rightarrow P(y_n | x_n) = \begin{cases} h(x) & y_n = +1 \\ 1 - h(x) & y_n = -1 \end{cases}$

$$\Rightarrow \frac{1}{N} \sum_{n=1}^N \left[ \ln \left( \frac{1}{h(x_n)} \right) + \ln \left( \frac{1}{1-h(x_n)} \right) \right]$$

$$\Rightarrow \frac{1}{N} \sum_{n=1}^N \left[ [y_n = +1] \ln \left( \frac{1}{h(x_n)} \right) + [y_n = -1] \ln \left( \frac{1}{1-h(x_n)} \right) \right]$$

(b) for  $h(n) = O(w^T x_n)$

$$\therefore E_{\text{loss}} = \frac{1}{N} \sum_{n=1}^N [y_n = +1] \ln \left( \frac{1}{h(x_n)} \right) + [y_n = -1] \ln \left( \frac{1}{1-h(x_n)} \right)$$

putting  $h(x_n) = O(w^T x_n)$

$$= \frac{1}{N} \sum_{n=1}^N [y_n = +1] \ln \left( \frac{1}{O(w^T x_n)} \right) + [y_n = -1] \ln \left( \frac{1}{1-O(w^T x_n)} \right)$$

$$\text{we know, } 1 - O(w^T x_n) = O(-w^T x_n).$$

$$\therefore \frac{1}{N} \sum_{n=1}^N [y_n = +1] \ln \left( \frac{1}{O(y_n w^T x_n)} \right) + [y_n = -1] \ln \left( \frac{1}{O(-y_n w^T x_n)} \right)$$

We can rewrite this, by taking  $y_n$  inside ' $O$ '.

$$\therefore \frac{1}{N} \sum_{n=1}^N \ln \left( \frac{1}{O(y_n w^T x_n)} \right) \quad \text{where } y_n = +1 \text{ will give} \\ \frac{1}{O(w^T x_n)}$$

&  $y_n = -1$  will give

$$\frac{1}{O(-w^T x_n)} = \frac{1}{1 - O(w^T x_n)}$$

$$\Rightarrow \frac{1}{N} \sum_{n=1}^N \ln \left( \frac{1}{1 + e^{y_n w^T x_n}} \right)$$

$$\Rightarrow \underbrace{\frac{1}{N} \sum_{n=1}^N \ln \left( 1 + e^{-y_n w^T x_n} \right)}$$

$$\Rightarrow \frac{1}{N} \sum_{n=1}^N \ln \left( \frac{1 + e^{-y_n w^T x_n}}{e^{y_n w^T x_n}} \right)$$

Equation 3.9.

Hence proved.

Exercise 3.13

$Z = \Phi_2(x)$  As per (3.13).  $\Phi_2(x) = \{1, x_1, x_2, x_1^2, x_1 x_2, x_2^2\}$

(a) Parabola  $(x_1 - 3)^2 + x_2^2 = 1$

$$\begin{aligned} & x_1^2 + 9 - 6x_1 + x_2^2 = 1 \\ & \Rightarrow x_1^2 + 8 - 6x_1 + x_2^2 = 0. \end{aligned}$$

: Comparing coefficients to map in hyperplane 'Z',  
the  $\tilde{w} = ?$

$$\text{for } \Phi_2(x) = \underbrace{1}_{w_0}, \underbrace{x_1}_{w_1}, \underbrace{x_2}_{w_2}, \underbrace{x_1^2}_{w_3}, \underbrace{x_1 x_2}_{w_4}, \underbrace{x_2^2}_{w_5}.$$

$$\therefore [w_0 = 8, w_1 = -6, w_2 = 1, w_3 = 1, w_4 = 0, w_5 = 0]$$

(b) The circle,  $(x_1 - 3)^2 + (x_2 - 4)^2 = 1$

$$\Rightarrow x_1^2 + 9 - 6x_1 + x_2^2 + 16 - 8x_2 = 1$$

$$\Rightarrow x_1^2 + 9 + 16 - 1 - 6x_1 - 8x_2 + x_2^2 = 0.$$

$$\Rightarrow x_1^2 + 25 - 6x_1 - 8x_2 + x_2^2 = 0.$$

$$\therefore [w_0 = 25, w_1 = -6, w_2 = -8, w_3 = 1, w_4 = 0, w_5 = 1]$$

(c) The ellipse,  $2(x_1 - 3)^2 + (x_2 - 4)^2 = 1$

$$\Rightarrow 2(x_1^2 + 9 - 6x_1) + x_2^2 + 16 - 8x_2 = 1$$

$$\Rightarrow 2x_1^2 + 18 - 12x_1 + x_2^2 + 16 - 8x_2 = 1 = 0.$$

$$\Rightarrow 2x_1^2 + (18 + 16 - 1) - 12x_1 + x_2^2 - 8x_2 = 0. \quad (\text{Ans})$$

$$2x_1^2 + x_2^2 - 12x_1 - 8x_2 + 33 = 0$$

$$\boxed{\omega_0 = 33, \omega_1 = -12, \omega_2 = -8, \omega_3 = 2, \omega_4 = 0, \omega_5 = 1}$$

Problem 3.16

True classification  
 $+1$  (correct person)  $\rightarrow$  (introduction)

(a)

$g(x)$	+1	0	$C_a$
	-1	$C_r$	0

Expected

$$\begin{aligned} \text{cost(accept)} &= (0) P[y = +1|x] + C_a (P[y = -1|x]) \\ &= 0 + (C_a) P[y = -1|x] \end{aligned}$$

Given,  $g(x) = P[y = +1|x] \therefore -1 - g(x) = P[y = -1|x]$ .

$$\therefore \text{cost(accept)} = C_a(1 - g(x)) \quad \boxed{\text{hence proved.}}$$

Similarly,  $\text{cost(reject)} = (C_r) P(y = +1|x) + 0 (P(y = -1|x))$

$$\therefore \text{cost(reject)} = C_r g(x), \text{ hence proved.}$$

b) On  $g(x)$ , that means on threshold we have equal probability of getting accepted or rejected.

$$\therefore \text{on threshold } g(x) = k,$$

$$\text{cost(accept)} = \text{cost(reject)}$$

$$\therefore C_a(1 - g(x)) = C_r g(x) \quad \therefore k = g(x)$$

$$\therefore C_a(1 - k) = C_r k$$

$$\therefore C_a - k C_a = C_r k$$

$$\therefore C_a = (C_a + C_r) k$$

$$\boxed{\frac{C_a}{C_a + C_r} = k} \quad \text{hence proved.}$$

$$K = \frac{C_A}{C_A + C_R}$$

(c) Cost matrix for supermarket,

	+1	-1
+1	0	1
-1	10	0

$$\therefore K = \frac{C_A}{C_A + C_R} \quad \text{we have } C_A = 1 \quad \& \quad C_R = 10.$$

$\therefore K = \frac{1}{1+10} = \frac{1}{11}$  : for supermarket, we want to keep the rejection very low

threshold shows, for  $K=1$  since ' $K$ ' is very small.  $\therefore g(m) \geq K$  will be high &  $g(m) < K$  will be low.

	+1	-1
+1	0	1000
-1	1	0

$$= \frac{1000}{1000 + 1} = \frac{1000}{1001}$$

$\therefore$  Here we get threshold ' $K$ ' very high.  $\therefore g(m) \geq K$  will accept very few instance which is expected for CTA & a lot of rejections as  $g(m) < K$  will be more probable.

finally to conclude

For supermarket, we are avoiding false rejects so with  $K=1$ , & we will reject very few.

For CTA, we want to avoid false accepts,  $\therefore K$  will be very large & probability to ensure we will accept with very less probability.

Problem 2.22

References for LFD book page 63, 64.

Bias = To prove =

$$E_D(E_{out}(g^{(D)})) = \text{Eny} \cdot \sigma^2 + \text{bias} + \text{var.}$$

$\therefore$  We know, bias =  $(\bar{g}(n)^2 - f(n)^2)$  where  $\bar{g}(n)$

$$= E_D[g^{(D)}(n)]$$

Given,

$$\text{Eny}[E_{out}(g^{(D)})] = \text{Eny}[(g^{(D)}(n) - g(n))^2]$$

LHS where  $y(n) = f(n) + \epsilon$

$$\therefore \text{Eny}[(g^{(D)}(n) - y(n))^2]$$

$$= \therefore E_D[E_{out}(g^{(D)})] = E_D[\text{Eny}[(g^{(D)}(n) - y(n))^2]]$$

LHS RHS.

LHS =

$$\text{Eny}[E_D[(g^{(D)}(n) - g(n))^2]]$$

..... from page 63.  
of LFD.

$$= \text{Eny}[(E_D[g^{(D)}(n)]) - y(n)^2].$$

$$= \text{Eny}[E_D[g^{(D)}(n)^2] + (y(n))^2 - 2E_D[g^{(D)}(n)]y(n)]$$

$$= \text{Eny}[E_D[g^{(D)}(n)^2] + (f(n) + \epsilon)^2 - 2E_D[g^{(D)}(n)](f(n) + \epsilon)]$$

$\curvearrowright$  put  $y(n) = f(n) + \epsilon$

$$= E_{n,y} \left[ E_D \left[ (g^{(D)}(n))^2 \right] - 2E_D \left[ g^{(D)}(n) \right] (f(n) + \epsilon) + (f(n) + \epsilon)^2 \right]$$

~~we know,~~  $\bar{g}(n) = E_D \left[ g^{(D)}(n) \right]$

$\therefore$  substituting with  $\bar{g}(n)$ ,

$$= E_{n,y} \left[ E_D \left[ g^{(D)}(n)^2 \right] - 2\bar{g}(n) (f(n) + \epsilon) + f(n)^2 + (\epsilon)^2 + 2f(n)(\epsilon) \right]$$

$$= E_{n,y} \left[ E_D \left[ g^{(D)}(n)^2 \right] - 2\bar{g}(n) f(n) - 2\bar{g}(n)f(\epsilon) + f(n)^2 + (\epsilon)^2 + 2f(n)(\epsilon) \right]$$

$\therefore$  add & subtract  $\bar{g}(n)^2$ .

$$\therefore E_{n,y} \left[ E_D \left[ g^{(D)}(n)^2 \right] - \bar{g}(n)^2 + \bar{g}(n)^2 - 2\bar{g}(n)f(n) - 2\bar{g}(n)(\epsilon) + f(n)^2 + (\epsilon)^2 + 2f(n)(\epsilon) \right]$$

$$= E_{n,y} \left[ \underbrace{\left[ E_D \left[ g^{(D)}(n)^2 \right] - \bar{g}(n)^2 \right]}_{-2\bar{g}(n)(\epsilon) + (\epsilon)^2} + \underbrace{\left[ \bar{g}(n) - 2\bar{g}(n)f(n) + f(n)^2 \right]}_{+2f(n)(\epsilon)} \right]$$

$$= E_{n,y} \left[ \underbrace{\left( E_D \left[ g^{(D)}(n)^2 \right] - \bar{g}(n)^2 \right)}_{\text{Var}(n)} + \underbrace{\left( \bar{g}(n) - f(n) \right)^2}_{-2\bar{g}(n)\epsilon + 2f(n)\epsilon + (\epsilon)^2} \right]$$

$$= E_{n,y} \left[ \text{Var}(n) + \text{bias}(n) - 2\bar{g}(n)(\epsilon) + 2f(n)(\epsilon) + (\epsilon)^2 \right]$$

$$= E_x[\text{Var}(n)] + E_y[\text{bias}(x)] - 2E_x(\bar{g}(n) - f(n)) + (\epsilon)^2$$

$$= \text{Var } f \text{ bias} + \sigma^2$$

$$\text{where } \sigma^2 = -2(\epsilon)E_x \left[ \bar{g}(m) - f(m) \right] + (\epsilon)^2$$