Problem 2.1

Equation (2.1) $\Rightarrow$ $\epsilon(M, N, \delta) = \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$

Given $\delta = 0.03$.

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

① $M = 1$

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

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

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

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

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

$$\geq \frac{41997}{50}$$

$$\geq 839.94.$$

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

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,

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

$$\Rightarrow N \geq \frac{8.8048 \times 10^{4}}{2 \times 25} \Rightarrow N \geq \frac{88048}{50} \Rightarrow \boxed{N \geq 1760.96}$$

$\therefore$ We need 1760.96 examples

③ $\underline{M = 10{,}000}$

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

$25 \times 10^{-4} \geq \frac{1}{2N} \ln\left(\frac{2 \times 10^{4}}{0.03}\right)$

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

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

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

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 ~~negative~~ positive rays.

for negative & positive common case: All +1 & All -1.

$\therefore$ subtract 2 dichotomies got from negative rays

as we alredy have it counted in positive rays.

$\therefore$ maximum Dichotomies for positive/negtive rays:-

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

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

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

$$= 2N.$$

$$\therefore \boxed{m_{\mathcal{H}}(N) = 2N}$$

Now largest value of N for which $m_{\mathcal{H}}(N) = 2^N$,

$\therefore$ $N = 2$, $m_{\mathcal{H}}(2) = 4$ & Growth $f^n = 2N = 4$

for $N = 3$, $m_{\mathcal{H}}(3) = 8$ but we get $m_{\mathcal{H}}(3) = 6$.

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

(b) <u>Positive or Negative Intervals</u> :-

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

for negative intervals, we add only (+1, -1, +1) for $N=3$ dichotomy, for $\underline{N=4}$ 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.

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

VC dimension = largest value of N for which $m_{\mathcal{H}}(N) = 2^N$

$\therefore N = 2$, $m_{\mathcal{H}}(N) = \frac{\cancel{4}^{2}}{\cancel{2}} + \frac{\cancel{12}^{6}}{\cancel{2}} - 1 = 7$ $\quad\therefore$ $N = 3$, $m_{\mathcal{H}}(3) = \frac{9}{2} + \frac{9}{2} - 1$

$2^{(2)} = 4 \,!= 7$ $\quad 9 - 1 = 8$

$2^N = 2^3 = 8$.

$\therefore \boxed{\text{VC dimension} = 3}$

(c)

Problem 2.3 (C)

(C) we have function, +1 for $a \le \sqrt{x_1^2 + \ldots + x_d^2} \le b$.

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

$\therefore$ for this circles equation $\Rightarrow$ $x_1^2 + x_2^2 = r^2$

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

$\Rightarrow$ that means, $\sqrt{x_1^2 + \ldots + x_d^2} \longrightarrow$ 'r' between $[a, b]$

$\Rightarrow \underbrace{a \le r \le b}$.

$\Rightarrow$ positive interval where 'r' $\longrightarrow [0, +\infty]$.

$\therefore$ maximum dichotomies of 'C' $\Rightarrow$ maximum dichotomies for positive intervals.

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

$\therefore$ V_c dimension = largest N for which $m_H(N) = 2^N$.

our $m_H(N) = \frac{N^2}{2} + \frac{N}{2} + 1$.

$\therefore N = 2$, $\Rightarrow \cancel{\frac{2^2}{2} + \frac{2}{2} + 1 = 2}$ $\frac{4}{2} + \frac{2}{2} + 1 = 4$ $\quad 2^N = 4$ equal. ~~Not equal~~.

$\therefore N = 3$, $m_H(N) = \frac{9}{2} + \frac{3}{2} + 1 =$ $(N = 3)$ $2^N = 8$ & $m_H(N) = 8$.

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

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

for $N = 4$, $2^N = 16$
$m_H(N) = \frac{16}{2} + \frac{4}{2} + 1 = 11$ Not equal.

problem 2·8

for growth-function $m_{\mathcal{H}}(N)$ we know two conditions,

① we get V_c dimension (d_{vc}) when we have largest 'N' for which $m_{\mathcal{H}}(N) = 2^N$.

② Also, $m_{\mathcal{H}}(N)$ is bounded by $N^{d_{vc}}+1$ which implies.

$\Rightarrow$ $m_{\mathcal{H}}(N) \leq N^{d_{vc}}+1$ for 'N'.

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

(a) For $N+1$,

putting $N=1$, $2^N = 2$ — ①

$m_{\mathcal{H}}(N) = N+1 = 2$ — ②

$\therefore$ ① == ② which satisfies first condition

putting $N=2$, $2^N = 4$ — ①

$m_{\mathcal{H}}(2) = 2+1 = 3$ — ②.

$\therefore$ ① != ② $\therefore$ largest $N = 1 \Rightarrow d_{vc} = 1$

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

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

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

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

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

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

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

1^{st} condition → Get finite d_{vc} where $m_\mathcal{H}(N) = 2^N$.

$\therefore d_{vc} = 2$ ... positive intervals.

2^{nd} condition → It satifies $m_\mathcal{H}(N) = \frac{N^2}{2} + \frac{N}{2} + 1 \leq N^{d_{vc}} + 1$

for all 'N'

Hence 'b' is also a possible growth function

(C) 2^N → we know that for every 'N' we get $m_\mathcal{H}(N) = 2^N$

as our growth function is 2^N.

$\therefore d_{vc} = \infty$ .... This is from convex circles.

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

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

1^{st} condition ⇒ Get d_{vc}, in this case $d_{vc} = 1$,

2nd condition ⇒ for all N, should be bounded by $N^{d_{vc}} + 1$ i.e $N^{(1)} + 1$

$\therefore$ Take $N = 1$, $m_\mathcal{H}(N) = 2^{\sqrt{1}} = 2$
& $N^{(1)} + 1 = 1 + 1 = 2$] true

$N = 4$, $m_\mathcal{H}(N) = 2^{\sqrt{4}} = 4$
$N^{(1)} + 1 = 4 + 1 = 5$] true

For $N = 25$, $m_\mathcal{H}(25) = 2^{\sqrt{25}} = 32$
& $N^{(1)} + 1 = 26$] → Here $m_\mathcal{H}(25) < N^{d_{vc}} + 1$

$\therefore$ (d) is not a possible growth function

(e) $2^{\lfloor N/2 \rfloor}$ $\Rightarrow$ $d_{vc} = 0$, must be bounded by $N^0 + 1 = 2$,

$\therefore$ for $N = 4$, $m_{\mathcal{H}}(N) = 2^{N/2} = 4$
& $N^{d_{vc}} + 1 = 2$] Not true.

$\boxed{\text{Hence this is not a possible growth function}}$

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

$d_{vc} = 1$ $\left(m_{\mathcal{H}}(N) = m_{\mathcal{H}}(2) = 3 \neq 2^2 \right)$.
$m_{\mathcal{H}}(1) = 2 = 2^{(1)}$ $\therefore$ $d_{vc} = 1$.

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

$\therefore$ bound $= N^1 + 1 = N + 1$

for $N = 1$, $N+1 = 2$ & $m_{\mathcal{H}}(N) = 2$ $\rightarrow$ true.
$N = 2$, $N+1 = 3$ & $m_{\mathcal{H}}(2) = 3$ $\rightarrow$ true.
$N = 3$ $N+1 = 4$ & $m_{\mathcal{H}}(3) = 8$ $\Rightarrow$ Not true.

$\boxed{\text{Hence this is also not a possible growth function}}$

Problem 2.12

As from book we know,

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

where ϵ = generalization error

& $1 - \frac{\text{confidence}(\%)}{100} = \delta$.

Given,

∴ $d_{vc} = 10$

$\epsilon = 0.05$

$\delta = 0.05$ . . . . . . $(1 - 0.95)$

we can compute this iteratively, by replacing RHS, 'N'.

Taking $N = 1000$, in RHS,

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

∴ $\boxed{N = 257251.3.}$

Taking $N = 257251.3$, in RHS,

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

$\boxed{N = 434853.08}$

And so on....

Question 6

Prove the 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

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

$\Rightarrow$ We know standard error measure used in logistic regression is based on likelihood.

$\therefore$ our $f(x) = P[y=+1 \mid x]$.

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

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

We have $h(x) = \theta(w^T x)$, Also, $1-\theta(s) = \theta(-s)$

where $\theta(s) = \dfrac{e^s}{(1+e^s)}$

Since $(x_1,y_1)\ldots(x_n,y_n)$ 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 ~~will~~ are monotonically decreasing function.

$\therefore$ Now we can minimize,

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

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

& $\prod \Rightarrow \Sigma$

$\Rightarrow$ No we minimize,

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

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

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

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

$$\Rightarrow \frac{1}{N} \sum_{n=1}^{N} \ln\left(e^{(-y_n w^T x_n)} + 1 \right)$$ .... Divided by $e^{y_n w^T x_n}$ to numerator & denominator

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

Hence proved.

Problem 7

7 $\nabla E_{in}(w(t)) = ?$ in gradient descent algorithm.

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

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

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

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

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

① ⇒ is chain rule of derivation.

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

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

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

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

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~~ — We can divide this for $y_n = +1$ & $y_n = -1$.

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

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

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

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

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

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

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

(b) for $h(x) = \theta(w^T x)$.

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

$\therefore$ putting $h(x_n) = \theta(w^T x_n)$

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

we know, $1 - \theta(w^T x_n) = \theta(-w^T x_n)$.

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

We can rewrite this, by taking y_n inside 'θ'.

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

& $y_n = -1$ will give $\frac{1}{\theta(-w^T x_n)} = \frac{1}{1 - \theta(w^T x_n)}$

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

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

Equation 3.9.

Hence proved.

Exercise 3.13

$Z = \Phi_2(X)$ As per (3.13). $\Phi_2(X) = \{1, x_1, x_2, x_1^2, x_1x_2, x_2^2\}$.

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

$\therefore \quad x_1^2 + 9 - 6x_1 + x_2 = 1$

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

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

$\therefore$ for $\Phi_2(x) = 1, x_1, x_2, x_1^2, x_1x_2, x_2^2$

$\downarrow \quad \downarrow$

$w_0 \quad w_1 \quad w_2 \quad w_3 \quad w_4 \quad w_5$.

$\therefore \boxed{w_0 = 8, \; w_1 = -6, \; w_2 = 1, \; w_3 = 1, \; w_4 = 0, \; w_5 = 0}$

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

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

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

$\Rightarrow \quad x_1^2 + 24 - 6x_1 - 8x_2 + x_2^2 = 0$.

$\boxed{\therefore w_0 = 24, \; w_1 = -6, \; w_2 = -8, \; w_3 = 1, \; w_4 = 0, \; w_5 = 1}$

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

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

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

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

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

$$\boxed{\Rightarrow w_0=33,\ w_1=-12,\ w_2=-8,\ w_3=2,\ w_4=0,\ w_5=1}$$

Problem 3.16

		True classification	
		+1 (correct person)	-1 (intruder)
$g(x)$	+1	0	C_a
	-1	C_r	0

(a)

Expected

$\therefore$ cost(accept) $= (0)\, P[y=+1|x] + C_a(P(y=-1|x))$

$= 0 + (C_a)\, P(y=-1|x)$

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

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

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

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

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

$\therefore$ on threshold $g(x) = K$,

$\therefore$ cost(accept) = cost(reject)

$\therefore C_a(1-g(x)) = C_r\, g(x) \quad \ldots \quad K = g(x)$

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

$\therefore C_a - KC_a = C_r K$

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

$\therefore \boxed{\dfrac{C_a}{C_a+C_r} = K}$ hence proved.

$$K = \frac{C_a}{C_a + C_r}$$

(C) cost matrix for supermarket,

		f: +1	-1
h	+1	0	1
	-1	10	0

$\therefore K = \dfrac{C_a}{C_a + C_r}$, we have $C_a = 1$ & $C_r = 10$.

$\therefore K = \dfrac{1}{1+10} = \cancel{}\, 1/11$ $\therefore$ for supermarket, we want to keep the rejection very low & give discounts which our threshold shows, for $K = 1/11$; i.e 'K' is very small. $\therefore g(x) \geq K$ will be high & $g(x) < K$ will be low.

cost matrix for CIA,

		f: +1	-1
h	+1	0	1000
	-1	1	0

$\therefore K = \dfrac{C_a}{C_a + C_r} = \dfrac{1000}{1000+1} = \dfrac{1000}{1001}$

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

finally to conclude

(1) For supermarket, we are avoiding false rejects so with $K = 1/11 \approx 0$ we will reject very few.

(2) For CIA, we want to avoid false accepts, $\therefore K \approx 1$ (very large probability) we ensure we will accept with very less probability.

Problem 2.22

To prove, $E_D[E_{out}(g^{(D)})] = \sigma^2 + \text{bias} + \text{var}$

Given, without noise in the data we have,

1) $E_{out}(g^{(D)}) = E_{x,y}[(g^{(D)}(x) - f(x))^2]$ — ①.

2) we know $\text{bias}(x) = (\bar{g}(x) - f(x))^2$

3) we also know $\text{var}(x) = E_D[(g^{(D)}(x) - \bar{g}(x))^2]$

where $\bar{g}(x) = E_D[g^{(D)}(x)]$.

Now proceeding with the solution:-

$$E_D[E_{out}(g^{(D)})] = E_D[E_{x,y}[(g^{(D)}(x) - y(x))^2]]$$

$$= E_D[E_{x,y}[g^{(D)}(x)^2] - 2E_{x,y}[g^{(D)}(x)]E_{x,y}[y(x)] + E_{x,y}[y(x)^2]]$$

$$= E_{x,y}[E_D[g^{(D)}(x)^2]] - 2E_{x,y}[E_D[g^{(D)}(x)]E_D[y(x)]] + E_{x,y}[E_D[y(x)^2]]$$

$$= E_{x,y}[\underbrace{(E_D[g^{(D)}(x)^2] - \bar{g}(x)^2)}_{\text{part 1}} + \underbrace{(\bar{g}(x)^2 - 2\bar{g}(x)E_D[y(x)] + E_D[y(x)^2])}_{\text{part 2}}]$$

= equation ~~2~~ — ②

$\therefore$ Part 1 $\Rightarrow$ $E_D\left[g^{(D)}(x)^2\right] - \bar{g}(x)^2$

$= E_D\left[g^{(D)}(x)^2\right] - 2\bar{g}(x)^2 + \bar{g}(x)^2$

$= E_D\left[g^{(D)}(x)^2 - 2g^{(D)}(x)\,\bar{g}(x) + \bar{g}(x)^2\right]$

$= E_D\left[\left(g^{(D)}(x) - \bar{g}(x)\right)^2\right]$

Part 2 $\Rightarrow$ $\bar{g}(x)^2 - 2\bar{g}(x)\,E_D[y(x)] + E_D[y(x)^2]$

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

$= \bar{g}(x)^2 - 2\bar{g}(x)\,E_D[f(x)] - 2\bar{g}(x)\,E_D[\epsilon] + E_D[f(x)^2]$
$+ 2E_D[f(x)\epsilon] +$
$E_D[\epsilon^2]$

$=$ ~~$\bar{g}(x) - f(x)$~~

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

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

$\therefore$ putting back in equation 2,

$$E_D\left[E_{out}(g^{(D)})\right]$$

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

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

$\therefore$ putting 'var' & 'bias' now,

$= \cancel{\text{var} + \text{bias} + \sigma^2}$ ~~hence proved~~.

$\Rightarrow \cancel{E_x[\text{var}(x)] + E_{x,y}[\text{bias}(x)]} + E_{x,y}[\epsilon^2] \cancel{- 2E_{x,y}[(\bar{g}(x) - f(x))\epsilon]}$

$\therefore \Rightarrow$ var + bias + $\text{Var}[\epsilon]$ as $E_\epsilon[\epsilon] = 0$.

$\Rightarrow \boxed{\text{var} + \text{bias} + \sigma^2 \text{ .... hence proved.}}$