3.1) $w = A^{-1}b$

$A^{-1} = \frac{\text{adj } A}{|A|}$ & $\text{adj } A = C^T$

$A = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}$ Cofactor matrix of $A = \begin{bmatrix} 3 & -1 \\ -1 & +4 \end{bmatrix}$

$\text{adj } A = C^T = \begin{bmatrix} 3 & -1 \\ -1 & 4 \end{bmatrix}$

$|A| = 12 - 1 = 11$

$A^{-1} = \frac{1}{11}\begin{bmatrix} 3 & -1 \\ -1 & 4 \end{bmatrix}$

$w = A^{-1}b = \frac{1}{11}\begin{bmatrix} 3-2 \\ -1+8 \end{bmatrix} = \frac{1}{11}\begin{bmatrix} 1 \\ 7 \end{bmatrix}$

verifying : $Aw = b$

$\begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 7 \end{bmatrix}\frac{1}{11} = \frac{1}{11}\begin{bmatrix} 4+7 \\ 1+21 \end{bmatrix} = \frac{1}{11}\begin{bmatrix} 11 \\ 22 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = b$

3.2) 1) $f(x) = x^2 + 3x + 1$

$\frac{df}{dx} = 2x + 3$

2) $g(x_1, x_2) = x_1^2 + 2x_1x_2 + 3x_2^2$

$\frac{\partial g}{\partial x_1} = 2x_1 + 2x_2$

$\frac{\partial g}{\partial x_2} = 2x_1 + 6x_2$

3) $h(x) = \begin{bmatrix} x^2 \\ 3x+1 \end{bmatrix}$

$\frac{dh}{dx} = \begin{bmatrix} 2x \\ 3 \end{bmatrix}$

4) $\nabla_x g = \dfrac{\partial g}{\partial x}$ $\quad$ x is a vector $= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$\therefore \nabla_x g = \begin{bmatrix} \dfrac{\partial g}{\partial x_1} \\ \dfrac{\partial g}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 + 2x_2 \\ 2x_1 + 6x_2 \end{bmatrix}$$

5) $g(x) = x^T A x$ $\quad$ $g(x_1, x_2) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + 2x_1x_2 + 3x_2^2$

$$= \underset{1\times 2}{[x_1 \;\; x_2]} \, A \, \underset{2\times 1}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} = 1\times 1$$

$\therefore$ A ~~is~~ a 2×2 matrix

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

A is symmetric matrix i.e. $A^T = A$ $\quad \therefore b = c$

$$x^T A x = [x_1 \; x_2] \begin{bmatrix} a & b \\ b & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [ax_1 + bx_2 \quad bx_1 + dx_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= ax_1^2 + bx_1x_2 + bx_1x_2 + dx_2^2$$

$\therefore a = 1 \quad d = 3 \quad 2b = 2$
$\quad b = 1$

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}$$

6) $$\frac{\partial}{\partial x}(x^T A x) = \begin{bmatrix} \dfrac{\partial (x^T A x)}{\partial x_1} \\ \dfrac{\partial (x^T A x)}{\partial x_2} \end{bmatrix} =$$

7) $\frac{\partial (A^T x)}{\partial x} = A$

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \cdots a_{1n} \\ a_{21} & \cdots & a_{2n} \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \qquad A^T x = \begin{bmatrix} \sum_{i=1}^{i=n} a_{i1} x_i \\ \sum_{i=1}^{n} a_{2i} x_i \\ \vdots \end{bmatrix}$$

$$\frac{\partial (A^T x)}{\partial x} = \begin{bmatrix} \frac{\partial A^T x}{\partial x_1} \\ \vdots \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & a_{31} & \cdots & a_{n1} \\ a_{12} & a_{22} & a_{32} & \cdots & a_{n2} \\ a_{1n} & & & & a_{nn} \end{bmatrix} = A$$

b) $x^T x = [x_1\ x_2\ \cdots\ x_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1^2 + x_2^2 + \cdots x_n^2$

$$\frac{\partial (x^T x)}{\partial x} = \begin{bmatrix} \frac{\partial (x^T x)}{\partial x_1} \\ \frac{\partial (x^T x)}{\partial x_2} \\ \vdots \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{bmatrix} = 2x$$

c) $x^T A x = [x_1\ x_2\ \cdots\ x_n] \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & & & \\ \vdots & & & \\ a_{n1} & & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$$= [x_1 a_{11} + x_2 a_{21} + x_3 a_{31} \cdots,\ x_1 a_{12} + x_2 a_{22} \cdots,\ \cdots] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= x_1 a_{11} x_1 + x_2 a_{21} x_1 + x_3 a_{31} x_1 \cdots + x_1 a_{12} x_2 + x_2 a_{22} x_2 \cdots$$
$$+ x_1 a_{13} x_3 + x_2 a_{23} x_3 \cdots$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} x_i a_{ij} x_j$$

$$\frac{\partial (x^T A x)}{\partial x} = \begin{bmatrix} \frac{\partial x^T A x}{\partial x_1} \\ \frac{\partial x^T A x}{\partial x_2} \\ \vdots \end{bmatrix}$$

$\frac{\partial x^T A x}{\partial x_1} \to 2x_1 a_{11} + x_2 a_{21} + x_3 a_{31} \cdots + a_{12} x_2 + a_{13} x_3 \cdots$

$$= \sum_{i=1}^{n} a_{i1} x_i + \sum_{j=1}^{n} a_{1j} x_j$$

A is symmetric $\therefore a_{i1} = a_{1i}$

$$= 2 \sum_{i=1}^{n} a_{i1} x_i$$

Sim, $\frac{\partial x^T A x}{\partial x_2} = a_{21} x_1 + x_1 a_{12} + 2 x_2 a_{22} + x_3 a_{32} \dots + a_{23} x_3 + a_{24} x_4 \dots$

$$= \sum_{j=1}^{n} a_{2j} x_j + \sum_{i=1}^{n} x_i a_{i2} = 2 \sum_{i=1}^{n} a_{i1} x_i$$

$$\therefore \frac{\partial (x^T A x)}{\partial x} = \begin{bmatrix} 2 \sum_{i=1}^{n} a_{i1} x_i \\ 2 \sum_{i=1}^{n} x_i a_{i2} \\ \vdots \end{bmatrix} = 2Ax \quad \text{bc } A \text{ is sym } A = A^T$$

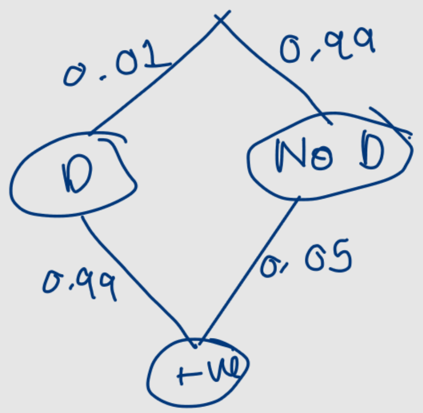
$\to (A + A^T) x$

6) from above result

$$\frac{\partial}{\partial x}(x^T A x) = (A + A^T) x = 2Ax = 2 \begin{bmatrix} \sum_{i=1}^{n} a_{i1} x_i \\ \sum_{i=1}^{n} a_{i2} x_i \\ \vdots \\ \sum_{i=1}^{n} a_{in} x_i \end{bmatrix}$$

---

4) 1)



$P(D \mid +ve)$

$$= \frac{P(D)\, P(+ve \mid D)}{P(D)\, P(+ve \mid D) + P(ND)\, P(+ve \mid ND)}$$

$$= \frac{(1)(99)}{1(99) + (5)(99)}$$

$$\approx \frac{1}{6}$$

4.2) 1) $L(\mu, \sigma^2) = p(x_1, x_2, \ldots\ldots x_n \mid \mu, \sigma^2)$

samples are iid

$\therefore L(\mu, \sigma^2) = p(x_1 \mid \mu, \sigma^2) \times p(x_2 \mid \mu, \sigma^2) \times p(x_3 \mid \mu, \sigma^2) \ldots\ldots$

$$= \prod_{i=1}^{n} p(x_i \mid \mu, \sigma^2)$$

pdf of gaussian random variable

$$p(x_i \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}$$

$$L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2}$$

$$= (2\pi\sigma^2)^{-n/2} e^{-\frac{\sum (x_i-\mu)^2}{2\sigma^2}}$$

2) $\log L(\mu, \sigma^2) = l(\mu, \sigma^2)$

$$l(\mu, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{\sum_{i=1}^{n} (x_i-\mu)^2}{2\sigma^2}$$

→ log likelihood f^n

3) diff wrt μ & $= 0$

$$\frac{\partial l}{\partial \mu} = 0 - \frac{1}{\cancel{2}\sigma^2} \sum_{i=1}^{n} \cancel{2}(x_i-\mu)(-1) = 0$$

$$\Sigma x_i - \mu n = 0$$

$$\mu_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}$$

MLE of mean is the mean of the sample

4) diff wrt σ^2 & $= 0$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\cancel{2\pi}\sigma^2} (\cancel{2\pi}) - \frac{\sum_{i=1}^{n} (x_i-\mu)^2}{2} \left(\frac{-1}{\sigma^4}\right)$$

$$= \frac{-n}{\cancel{2\sigma^2}} + \frac{\Sigma (x_i-\mu)^2}{\cancel{2\sigma^4}} = 0 \Rightarrow \sigma^2_{MLE} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_{MLE})^2$$

5) $\hat{\mu}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad \hat{\sigma}^2_{MLE} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu}_{MLE})^2$

5) 1) (5, 10)

2) bc given data set (2,4) ; (4,8) , (9,18)

I noticed that $4 = 2\times2$ $8 = 4\times2$ $18 = 9\times2$

$\therefore$ I predicted for input = 5 output = 5×2

3) $f(x) = ax^3 + bx^2 + cx + d$

$4 = a(8) + b(4) + 2c + d$

$8 = a(64) + b(16) + 4c + d$

(−)

$\cancel{4} = \cancel{56}a + \cancel{12}b + \cancel{2}c$

$2 = 28 \quad 6$

$18 = a(729) + b\,81 + 9c + d$

$4 = a8 + 4b + 2c + d$

$14 = 721a + 77b + 7c$

$14 = 196a + 42b + 7c$

$0 = \cancel{525}a + \cancel{35}b$

$\cancel{105}$ 7

15

$\overset{5}{\cancel{28}}$
$\times 7$
196

721 − 196
−200+4

521+4

$-15a = b$

a cubic function has 4 parameters (a, b, c, d) but we have only 3 constraints so we can't find a, b, c, d by value but we can find the relation b/ them.
(bc ∞ values can satisfy the relation)

∴ No, the cubic f^n need not necessarily satisfy $f(5) = 10$

4) I assumed that the function is linear. There could be other functions which satisfy the given data pairs. But i took the linear function which is most intuitive

5) The hypothesis of linear regression for a single input x. is that the output is a linear function of the input.

$y = mx + c$ here the used letters are $y = wx + b$

w = weights

6) If each datapoint has d features

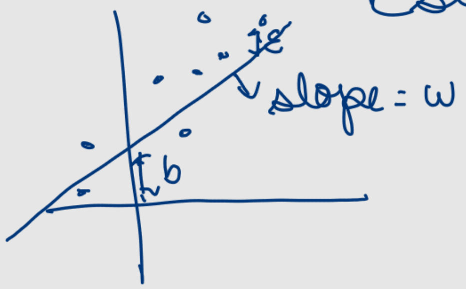
$x = (x_1, x_2 \dots x_d)$

$$y = w_1 x_1 + w_2 x_2 + w_3 x_3 \dots w_d x_d + b$$

or in vector form $y = wx + b$

9) Let the data points inputs be $x_1, x_2 \dots x_n$ (single-feature)

$$y_i = w x_i + b + \varepsilon_i$$



$$\varepsilon_i = y_i - w x_i - b$$

$$\|\varepsilon_i\|_2 = \sqrt{\sum_{i=1}^{n} (y_i - w x_i - b)^2}$$

$$\|\varepsilon_i\|_2^2 = \sum_{i=1}^{n} (y_i - w x_i - b)^2 = \text{Total loss} = E(w, b)$$

10) $$\frac{\partial E}{\partial w} = 2 \sum (y_i - w x_i - b)(-x_i) = 0$$

$$= \sum x_i (y_i - w x_i - b) = 0$$

$$\frac{\partial E}{\partial b} = 2 \sum (y_i - w x_i - b)(-1) = 0$$

$$\sum (y - w x_i - b) = 0$$

$$\rightarrow \sum y_i x_i - w \sum x_i^2 - b n \bar{x} = 0$$

$$\rightarrow \cancel{n} \bar{y} - w \cancel{n} \bar{x} - b \cancel{n} = 0$$

$$\bar{y} = w \bar{x} + b$$

$$\rightarrow \sum y_i x_i - w \sum x_i^2 - (\bar{y} - w \bar{x}) n \bar{x} = 0$$

$$\sum x_i y_i - n \bar{x} \bar{y} = w \left(\sum x_i^2 - n \bar{x}^2 \right)$$

$$w = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$b = \bar{y} - w\bar{x}$$

$$= \bar{y} - \left[\frac{(\sum x_i y_i)\bar{x} - n\bar{x}^2\bar{y}}{\sum x_i^2 - n\bar{x}^2}\right]$$

$$= \frac{(\sum x_i^2)\bar{y} - \cancel{n\bar{x}^2}\bar{y} - \sum x_i y_i \bar{x} + \cancel{n\bar{x}^2}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$= \frac{\sum x_i^2 \bar{y} - \sum x_i y_i \bar{x}}{\sum x_i^2 - n\bar{x}^2}$$

---

11) $\vec{y}_p = \vec{x} \times \vec{w}$ (prediction vector)

$$= \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1d} \\ 1 & x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix} \qquad w = \begin{bmatrix} b \\ w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}$$

$$\therefore \vec{y}_p = \begin{bmatrix} b + w_1 x_{11} + w_2 x_{12} + \cdots w_d x_{1d} \\ b + w_1 x_{21} + w_2 x_{22} \cdots w_d x_{2d} \\ \vdots \\ b + w_1 x_{n1} + \cdots w_d x_{nd} \end{bmatrix}$$

12) $\|\vec{\varepsilon}\|_2 = \|\vec{y} - \vec{x}\vec{w}\|_2$

$\|\vec{\varepsilon}\|_2^2 = \|\vec{y} - \vec{x}\vec{w}\|_2^2$ $\qquad \|x\|^2 = x^T x$ where x is column vector

since here $y - xw$ is a column vect

$$= (y - Xw)^T (y - Xw)$$

13) $$= (y^T - w^T x^T)(y - xw)$$

$$L(w) = y^T y \underbrace{- y^T X w - w^T x^T y}_{} + w^T x^T X w$$

↳ error fn

equal bc $y^T x w$ = scalar (bc $(1 \times n) \times (n \times d) \times (d \times 1)$) & transpose of scalar = scalar

$= 1 \times 1$

diffe wrt ω & equate to 0 to minimise

$$L(\omega) = y^T y - 2(y^T X)\omega + \omega^T (X^T X)\omega$$

$$\frac{\partial L}{\partial \omega} = 0 - 2(y^T X)^T + 2(X^T X)\omega = 0$$

↳ normal equations

* $X^T X$ is symmetric bc $(X^T X)^T = X^T X$

$$\cancel{2} X^T y = \cancel{2}(X^T X)\omega$$

$$\omega = (X^T X)^{-1} X^T y$$

$$\frac{d}{dx} a^T x = a$$

$$x^T x = 2x$$

$$x^T A x = 2Ax$$

↳ where A is symmetric *