1.

$$N(x|\mu, \sigma^2) = f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} , \quad -\infty < x < \infty$$

$$f(x_n, x_m) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_m-\mu)^2}{2\sigma^2}}$$

$$-\infty < x_m < \infty \quad , \quad -\infty < x_m < \infty$$

If $n \neq m$:

$$E[x_n x_m] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_n x_m f(x_n, x_m)\, dx_n\, dx_m$$

$$\Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_n x_m \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_m-\mu)^2}{2\sigma^2}}\, dx_n\, dx_m$$

$$= \int_{-\infty}^{\infty} x_m \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_m-\mu)^2}{2\sigma^2}} \underline{\int_{-\infty}^{\infty} x_n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}\, dx_n}\, dx_m$$

變數變換：令 $x_n = x$

$$\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

符合 Eq (1) $= \mu$

$$= \mu \left(\int_{-\infty}^{\infty} x_m \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_m-\mu)^2}{2\sigma^2}}\, dx_m \right) \Rightarrow \text{令 } x_m = x$$

$$= \mu \cdot \mu = \mu^2$$

If $n = m$

Let $x_n = x$

$$E[x_n x_m] = E[x_n x_n] = E[x_n^2]$$

$$= E(x^2) \Rightarrow \text{符合 Eq (2)}$$

$$= \mu^2 + \sigma^2$$

$\because$

if $m \neq n$, $E[x_n x_m] = \mu^2$

if $m = n$, $E[x_n x_m] = \mu^2 + \sigma^2$

$\therefore E[x_n x_m] = \mu^2 + I_{nm}\sigma^2$ ，其中 $I_{nm} = \begin{cases} 1 & , n = m \\ 0 & , \text{O.W} \end{cases}$

得證 Eq (3) #

ML = maximum likehood

$\mu_{ML} = \overline{x} = \dfrac{x_1 + x_2 + \cdots + x_N}{N}$ (樣本平均數)

$$E[\mu_{ML}] = E\left[\frac{x_1 + x_2 + \cdots + x_N}{N}\right]$$

$$= \frac{1}{N} E[x_1 + x_2 + \cdots + x_N]$$

$$= \frac{1}{N}\left(E[x_1] + E[x_2] + \cdots + E[x_N]\right)$$

每一個 x_i 皆可透過變數變化換成 $x \Rightarrow$ 符合 Eq (1)

$$= \frac{1}{N}(N\mu)$$

$= \mu$ ，得證 Eq (4) #

樣本變異數定義：

$$S^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}$$

σ^2_{ML} = 最大概似的樣本變異數

$$= \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}$$

$$E[\sigma^2_{ML}] = E\left[\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}\right]$$

$$= \frac{1}{N} E\left[\sum_{i=1}^{N} (x_i - \bar{x})^2\right] \quad \cdots\cdots \text{①}$$

$E\left[\sum_{i=1}^{N} (x_i - \bar{x})^2\right]$ 折解成 $\left[(x_i - \mu) + (\mu - \bar{x})\right]^2$

$$= E\left[\sum_{i=1}^{N} \left((x_i - \mu)^2 + (\mu - \bar{x})^2 + 2(\mu - \bar{x})(x_i - \mu)\right)\right]$$

$$= E\left[\sum_{i=1}^{N} (x_i - \mu)^2 + \sum_{i=1}^{N} (\mu - \bar{x})^2 + 2\sum_{i=1}^{N} (\mu - \bar{x})(x_i - \mu)\right]$$

μ 利用 Eq (4) 替換成 $E[\bar{x}]$

$$= \sum_{i=1}^{N} E\left[(x_i - \mu)^2\right] + \sum_{i=1}^{N} E\left[(\bar{x} - E[\bar{x}])^2\right] + 2E\left[(\mu - \bar{x}) \sum_{i=1}^{N} (x_i - \mu)\right]$$

$$\sum_{i=1}^{N} (x_i - \mu)$$

$$= \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} \mu$$

$$= N\bar{x} - N\mu$$

$$= N(\bar{x} - \mu)$$

$$= \sum_{i=1}^{N} E[(x_i-\mu)^2] + \underbrace{\sum_{i=1}^{N} E[(\bar{x}-E(\bar{x}))^2]}_{= N E[(\bar{x}-E(\bar{x}))^2]} - 2N\, E[(\bar{x}-\mu)^2]$$

μ 利用 Eq.(4) 替換成 $E[\bar{x}]$

$$= \sum_{i=1}^{N} E[(x_i-\mu)^2] - N\, E[(\bar{x}-E(\bar{x}))^2]$$

以 $N=3$ 來看

$$\bar{x} = \frac{x_1+x_2+x_3}{3}$$

則 $E[(\bar{x}-E(\bar{x}))^2]$

$$= E\left[\left(\frac{x_1+x_2+x_3}{3} - \frac{3\mu}{3}\right)^2\right]$$

$$= E\left[\left(\frac{(x_1-\mu)(x_2-\mu)(x_3-\mu)}{3}\right)^2\right]$$

$$= \frac{1}{3^2} E\left[(x_1-\mu)^2+(x_2-\mu)^2+(x_3-\mu)^2 + 2(x_1-\mu)(x_2-\mu) + 2(x_1-\mu)(x_3-\mu) + 2(x_2-\mu)(x_3-\mu)\right]$$

$$= \frac{1}{3^2}\left(3\sigma^2 + 0\right)$$

$\because E[(x_1-\mu)(x_2-\mu)]$

$$= E[x_1x_2] - E[\mu x_1] - E[\mu x_2] + E[\mu^2]$$

$$= \mu^2 - \mu^2 - \mu^2 + \mu^2$$

$$= 0$$

$$= \frac{1}{3}\sigma^2 \Rightarrow \text{推得 } \frac{\sigma^2}{3}$$

$$E\left[(y_i - t_i)^2\right] = (y_i - t_i)^2 \quad \because 常數$$

$$E\left[E_D(w)\right] = \frac{1}{2}\sum_{i=1}^{N}\left[(y_i - t_i)^2 + \sum_{j=1}^{D} w_j \sigma^2\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}(y_i - t_i)^2 + \underbrace{\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{D} w_j \sigma^2}_{常數}$$

$$= \underbrace{\frac{1}{2}\sum_{i=1}^{N}(y_i - t_i)^2}_{原本未加雜訊的模型} + A \quad \#$$

∵由結果可得知，在加入雜訊後，僅需針對原本的模型優化即可。

2.

隨機變數	機率密度函數	期望期
x	$f_x(x)$	$E(x)$
y	$f_y(y)$	$E(y)$
$x+y$	$f(x,y)$	$E(x+y)$

證明 $E(x+y) = E(x) + E(y)$

$$E(x+y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y)\,dx\,dy$$

$\because$ x 和 y 獨立

$\therefore$ $f(x,y) = f_x(x)\,f_y(y)$

$$= \int_{-\infty}^{\infty} x f_x(x) \underbrace{\int_{-\infty}^{\infty} f_y(y)\,dy}_{=1}\,dx +$$

$$\int_{-\infty}^{\infty} y f_y(y) \underbrace{\int_{-\infty}^{\infty} f_x(x)\,dx}_{=1}\,dy$$

$$= \int_{-\infty}^{\infty} x f_x(x)\,dx + \int_{-\infty}^{\infty} y f_y(y)\,dy$$

$$= E(x) + E(y) \quad \cdots\cdots \text{①式}$$

令 $\vec{a} = (X_1, X_2, \cdots, X_N)$

$$\vec{b} = (Y_1, Y_2, \cdots, Y_N)$$

$$\vec{Y} = \vec{a} + \vec{b} = (X_1+Y_1, X_2+Y_2, \cdots, X_N+Y_N)$$

$E(\vec{Y}) \rightarrow$ 拆出第 i 個

$$z_i = X_i + Y_i$$

$$E(z_i) = E(X_i + Y_i) \quad \text{by ① 式}$$

$$= E(X_i) + E(Y_i), \quad i = 1, 2, 3, \cdots, N$$

$$E(z_i) = E(X_i) + E(Y_i)$$

$$\begin{bmatrix} E(z_1) \\ E(z_2) \\ \vdots \\ E(z_N) \end{bmatrix} = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_N) \end{bmatrix} + \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_N) \end{bmatrix}$$

$\therefore E[\vec{Y}] = E[\vec{a}] + E[\vec{b}] \Rightarrow$ 得證第二題第1小題

#

定義：

$$\mathrm{COV}(A, A)$$

$$= E\left[(A - E(A))(A - E(A))^T\right]$$

假設 $A = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$= E\left[\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 & x_2 - \mu_2 \end{bmatrix}\right]$$

$$= \begin{bmatrix} E[(x_1 - \mu_1)^2] & a_{12} \\ a_{21} & E[(x_2 - \mu_2)^2] \end{bmatrix}$$

$$a_{12} = a_{21} = E[(x_1 - \mu_1)(x_2 - \mu_2)]$$

可以發現對角線上的都是"變異數"
非對角線上的會以對角線對稱

令 $\vec{a} = (x_1, x_2, \cdots, x_N)$

$\vec{b} = (y_1, y_2, \cdots, y_N)$

$\vec{y} = \vec{a} + \vec{b} = (x_1 + y_1, x_2 + y_2, \cdots, x_N + y_N)$

$$A = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \quad B = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \quad Y = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_N + y_N \end{bmatrix}$$

假設 $cov(A,A) = M_A$

$cov(B,B) = M_B$

$cov(Y,Y) = M_Y$

證明

$M_Y = M_A + M_B$

Step 1：

$M_A = cov(A,A) = E[(A-E(A))(A-E(A))^T]$

根據定義對角線上皆是"變異數"

故 $i=j \Rightarrow \sigma_i^2$

$i \neq j \Rightarrow E[(x_i-u_i)(x_j-u_j)]$

$= E[x_ix_j - x_iu_j - x_ju_i + u_iu_j]$

$= E[x_ix_j] - u_iu_j - \cancel{u_ju_i} + \cancel{u_iu_j}$

$\because x_i$、x_j 彼此獨立

$\iint x_ix_j f(x_i,x_j)\, dx_i\, dx_j$

$= (\int x_i f_{x_i}(x_i)dx_i)(\int x_j f_{x_j}(x_j)\, dx_j)$

$= u_i \cdot u_j$

$= u_iu_j - u_iu_j$

$= 0 \Rightarrow$ 矩陣之非對角線上皆為0

故 $M_A = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_N^2 \end{bmatrix}$

同理 $M_B = \begin{bmatrix} \hat{\sigma}_1^2 & 0 & \cdots & 0 \\ 0 & \hat{\sigma}_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \hat{\sigma}_N^2 \end{bmatrix}$

故 $M_A + M_B = \begin{bmatrix} \sigma_1^2 + \hat{\sigma}_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 + \hat{\sigma}_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_N^2 + \hat{\sigma}_N^2 \end{bmatrix}$

Step 2:

$$M_Y = \mathrm{cov}(Y, Y)$$
$$= \mathrm{cov}(A+B, A+B)$$
$$= E\left[[(x+y) - E(x+y)][(x+y) - E(x+y)]^T\right]$$

$i = j$ (在對角線上):

$$(M_Y)_{ii} = E\left[[(x_i + y_i) - (\mu_i + \hat{\mu}_i)]^2\right]$$
$$= E\left[[(x_i - \mu_i) + (y_i - \hat{\mu}_i)]^2\right]$$
$$= E\left[(x_i - \mu_i)^2 + (y_i - \hat{\mu}_i)^2 + 2(x_i - \mu_i)(y_i - \hat{\mu}_i)\right]$$

$$= E[(x_i-\mu_i)^2] + E[(y_i-\hat{\mu}_i)^2] + 2E[(x_i-\mu_i)(y_i-\hat{\mu}_i)]$$

$$= \sigma_i^2 + \hat{\sigma}_i^2 + 2\underbrace{E[x_i y_i - x_i\hat{\mu}_i - y_i\mu_i + \mu_i\hat{\mu}_i]}_{\text{在 step 1 有證明過 } = 0}$$

$$= \sigma_i^2 + \hat{\sigma}_i^2$$

$i \neq j$（非對角線）：

$$(M_y)_{ij} = E[[(x_i+y_i)-(\mu_i+\hat{\mu}_i)][(x_j+y_j)(\mu_j+\hat{\mu}_j)]]$$

$$= E[[(x_i-\mu_i)(y_i-\hat{\mu}_i)][(x_j-\mu_j)(y_j-\hat{\mu}_j)]]$$

$$= \underbrace{E[(x_i-\mu_i)(x_j-\mu_j)]}_{\text{step 1 有證明過 } = 0} + E[(x_i-\mu_i)(y_j-\hat{\mu}_j)] +$$

$$E[(y_i-\hat{\mu}_i)(x_j-\mu_j)] + E[(y_i-\hat{\mu}_i)(y_j-\hat{\mu}_j)]$$

$$= E[x_i y_j - x_i\hat{\mu}_j - \mu_i y_j + \mu_i\hat{\mu}_j] +$$

$$E[y_i x_j - y_i\mu_j - \hat{\mu}_i x_j + \hat{\mu}_i\mu_j]$$

皆在前面有證明過 $= 0$

$$= 0$$

故 $\begin{cases} i = j \Rightarrow \sigma_i^2 + \hat{\sigma}_i^2 \\ i \neq j \Rightarrow 0 \end{cases}$

$$M_Y = \begin{bmatrix} \sigma_1^2 + \hat{\sigma}_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 + \hat{\sigma}_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_N^2 + \hat{\sigma}_N^2 \end{bmatrix}$$

$$= M_A + M_B$$

$\because$ step 1 與 step 2 的結果相同

$\therefore$ 得證 #

3、

根據課本 3.8 所得到的結論

$$S_{N+1}^{-1} = S_N^{-1} + \beta \phi_{N+1} \phi_{N+1}^T$$

故 3.59 可被改寫成

$$\sigma_{N+1}^2(x) = \frac{1}{\beta} + \phi(x)^T S_{N+1} \phi(x)$$

$$\Rightarrow (S_{N+1}^{-1})^{-1} = (S_N^{-1} + \beta \phi_{N+1} \phi_{N+1}^T)^{-1}$$

令 $M = S_N^{-1}$

$$\beta \phi_{N+1} \phi_{N+1}^T = V V^T$$

$$\downarrow$$

$$(\sqrt{\beta}\, \phi_{N+1})(\sqrt{\beta}\, \phi_{N+1})^T$$

$$\Rightarrow V = \sqrt{\beta}\, \phi_{N+1}$$

$$\Rightarrow S_{N+1} = (M + V V^T)^{-1}$$

$$= M^{-1} - \frac{(M^{-1} V)(V^T M^{-1})}{1 + V^T M^{-1} V}$$

$$= (S_N^{-1})^{-1} - \frac{((S_N^{-1})^{-1} \sqrt{\beta}\, \phi_{N+1})((\sqrt{\beta}\, \phi_{N+1})^T (S_N^{-1})^{-1})}{1 + (\sqrt{\beta}\, \phi_{N+1})^T (S_N^{-1})^T (\sqrt{\beta}\, \phi_{N+1})}$$

$$= S_N - \frac{\beta S_N \phi_{N+1} \phi_{N+1}^T S_N}{1 + \beta \phi_{N+1}^T S_N \phi_{N+1}}$$

令 $\phi_{N+1}^T S_N \phi_{N+1}$ 為二次型

∵ S_N 為正定矩陣

∴ $\phi_{N+1}^T S_N \phi_{N+1}$ = 正定型二次式

⇒ 令 $\phi_{N+1}^T S_N \phi_{N+1} = k^+ > 0$

⇒ 令 $S_N(\phi_{N+1}\phi_{N+1}^T)S_N = Q^+$

$$\Rightarrow S_{N+1} = S_N - \beta \frac{Q^+}{1+\beta k^+}$$

$$\begin{aligned}
\Rightarrow \sigma_{N+1}^2 &= \frac{1}{\beta} + \phi^T \left(S_N - \beta \frac{Q^+}{1+\beta k^+}\right)\phi \\
&= \frac{1}{\beta} + \phi^T S_N \phi - \beta \phi^T \frac{Q^+}{1+\beta k^+}\phi \\
&= \left(\frac{1}{\beta} + \phi^T S_N \phi\right) - \beta\left(\frac{\phi^T Q^+ \phi}{1+\beta k^+}\right) \\
&= \sigma_N^2 - \beta \frac{\phi^T Q^+ \phi}{1+\beta k^+}
\end{aligned}$$

$$\beta \frac{\overbrace{\phi^T Q^+ \phi}^{\text{二次型}}}{1+\beta k^+} = \sigma_N^2 - \sigma_{N+1}^2$$

∵ 分母、分子皆 > 0，且 $\beta > 0$

∴ $\sigma_N^2 - \sigma_{N+1}^2 > 0 \Rightarrow \sigma_N^2 > \sigma_{N+1}^2$，得證 #

4.

線性模型：$y(\overset{\text{向量}}{x}, \overset{\text{向量}}{w}) = w_0 + \sum_{j=1}^{D} w_j x_j$

對輸入加高斯誤差

$$\underbrace{\hat{y}_i}_{= y(x_i, w)} = w_0 + \sum_{j=1}^{D} w_j (x_{ij} + \varepsilon_{ij})$$

誤差函數：

$$E_D(w) = \frac{1}{2}\sum_{i=1}^{N} (\hat{y}_i - t_i)^2$$

$$= \frac{1}{2}\sum_{i=1}^{N} \left(w_0 + \sum_{j=1}^{D} w_j (x_{ij} + \varepsilon_{ij}) - t_i \right)^2$$

$$= \frac{1}{2}\sum_{i=1}^{N} \left(\underbrace{w_0 + \sum_{j=1}^{D} w_j x_{ij}}_{= y_i} + \sum_{j=1}^{D} w_j \varepsilon_{ij} - t_i \right)^2$$

$$= \frac{1}{2}\sum_{i=1}^{N} \left(y_i + \sum_{j=1}^{D} w_j \varepsilon_{ij} - t_i \right)^2$$

$$E\left[E_D(w) \right] = \frac{1}{2} E\left[\sum_{i=1}^{N} \left(y_i + \sum_{j=1}^{D} w_j \varepsilon_{ij} - t_i \right)^2 \right]$$

$$= \frac{1}{2}\sum_{i=1}^{N} E\left[y_i^2 + \left(\sum_{j=1}^{D} w_j \varepsilon_{ij}\right)^2 + t_i^2 - 2 y_i t_i + 2 y_i \sum_{j=1}^{D} w_j \varepsilon_{ij} - 2 t_i \sum_{j=1}^{D} w_j \varepsilon_{ij} \right]$$

$$= \frac{1}{2}\sum_{i=1}^{N} E\Big[(y_i - t_i)^2 + \Big(\sum_{j=1}^{D} w_j \varepsilon_{ij}\Big)^2 - 2t_i \sum_{j=1}^{D} w_j \varepsilon_{ij}$$

$$+ 2y_i \sum_{j=1}^{D} w_j \varepsilon_{ij}\Big]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\Big[E\big((y_i - t_i)^2\big) + E\Big(\Big(\sum_{j=1}^{D} w_j \varepsilon_{ij}\Big)^2\Big) -$$

$$E\Big(2t_i \sum_{j=1}^{D} w_j \varepsilon_{ij}\Big) + E\Big(2y_i \sum_{j=1}^{D} w_j \varepsilon_{ij}\Big)\Big]$$

$$E\Big[2t_i \sum_{j=1}^{D} w_j \varepsilon_{ij}\Big] = \underbrace{2t_i \sum_{j=1}^{D} w_j}_{\text{常數可移出}} \; \underbrace{E[\varepsilon_{ij}]}_{\text{題目已知為 } 0}$$

$$= 0$$

同理， $E\Big[2y_i \sum_{j=1}^{D} w_j \varepsilon_{ij}\Big] = 0$

$E\Big[\Big(\sum_{j=1}^{D} w_j \varepsilon_{ij}\Big)^2\Big]$ ⇒ 以 2 項來看 ⇒ $(w_1\varepsilon_{i1} + w_2\varepsilon_{i2})^2$

⇒ 個別平方項 + 2(相乘項)

$= E[$個別平方項之和$] + 2E[$任 2 項乘積之和$]$

$$= \sum_{j=1}^{D} w_j^2 \underbrace{E[\varepsilon_{ij}^2]}_{\text{題目已知為 } \sigma^2} + 2\sum w_m w_n \underbrace{E[\varepsilon_{im}]}_{0} \underbrace{E[\varepsilon_{in}]}_{0}$$

$$= \sum_{j=1}^{D} w_j^2 \sigma^2$$

$$= \sum_{i=1}^{N} E[(x_i - \mu)^2] - N \cdot \left(\frac{\sigma^2}{N}\right)$$

$$= N\sigma^2 - \sigma^2$$

$$= (N-1)\sigma^2 \quad \cdots \quad ②$$

將 ② 代回 ① 式

$$E[\sigma^2_{ML}] = \frac{1}{N}(N-1)\sigma^2$$

$$= \frac{N-1}{N}\sigma^2$$ ，得證 Eq (5) #