

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

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

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

If  $n \neq m$  :

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

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

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

變數變換 : 令  $x_n = x$

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

符合 Eq (1) =  $\mu$

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

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

If  $n = m$

Let  $x_n = x$

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

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

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

"

$$\text{if } m \neq n, \quad E[X_n X_m] = \mu^2$$

$$\text{if } m = n, \quad E[X_n X_m] = \mu^2 + \sigma^2$$

$$\therefore E[X_n X_m] = \mu^2 + I_{nm} \sigma^2 \quad \text{其中 } I_{nm} = \begin{cases} 1, & n = m \\ 0, & \text{o.w.} \end{cases}$$

得證 Eq (3) #

ML = maximum likelihood

$$\mu_{ML} = \bar{X} = \frac{X_1 + X_2 + \dots + X_N}{N} \quad (\text{樣本平均數})$$

$$E[\mu_{ML}] = E\left[\frac{X_1 + X_2 + \dots + X_N}{N}\right]$$

$$= \frac{1}{N} E[X_1 + X_2 + \dots + X_N]$$

$$= \frac{1}{N} (E[X_1] + E[X_2] + \dots + E[X_N])$$

每個  $X_i$  皆可透過變數變化  
換成  $x \Rightarrow$  符合 Eq (1)

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

$$= \mu, \quad \text{得證 Eq (4) #}$$

樣本變異數定義：

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

$G_{ML}^2$  = 最大概似的樣本變異數

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

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

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

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

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

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

$\mu$  利用 Eq (4) 替換成  $E[\bar{x}]$

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

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

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

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

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

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

μ 利用 Eq (4) 替换成 E[x]

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

以  $N=3$  來看

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

$$\text{則 } E[(\bar{x} - E(\bar{x}))^2]$$

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

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

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

$$= \frac{1}{3^2} (36^2 + 0)$$

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

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

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

$$= 0$$

$$= \frac{1}{3} 6^2 \Rightarrow \text{推得 } \frac{6^2}{3}$$

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

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

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

將 (2) 代入 (1) 式

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

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

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

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

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

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

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

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

$$\text{令 } \vec{a} = (x_1, x_2, \dots, x_N)$$

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

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

$$E(\vec{y}) \rightarrow \text{抓出第 } i \text{ 個}$$

$$z_i = x_i + y_i$$

$$E(z_i) = E(x_i + y_i) \quad \text{by ① 式}$$

$$= E(x_i) + E(y_i), \quad i = 1, 2, 3, \dots, N$$

$$E(z_i) = E(x_i) + E(y_i)$$

$$\begin{bmatrix} E(z_1) \\ E(z_2) \\ \vdots \\ E(z_N) \end{bmatrix} = \begin{bmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_N) \end{bmatrix} + \begin{bmatrix} E(y_1) \\ E(y_2) \\ \vdots \\ E(y_N) \end{bmatrix}$$

$$\therefore E[\vec{y}] = E[\vec{a}] + E[\vec{b}] \Rightarrow \text{得證第} = \text{題第 1 小題} \quad \#$$

定義：

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

$$= E[(A - E(A))(A - E(A))^T]$$

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

$$\text{令 } \vec{a} = (x_1, x_2, \dots, x_N)$$

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

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

$$A = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

$$B = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

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



假設  $\text{cov}(A, A) = M_A$

$\text{cov}(B, B) = M_B$

$\text{cov}(Y, Y) = M_Y$

證明

$M_Y = M_A + M_B$

step 1:

$M_A = \text{cov}(A, A) = E[(A - E(A))(A - E(A))^T]$

根據定義對角線上皆是“變異數”

故  $i=j \Rightarrow \sigma_i^2$

$i \neq j \Rightarrow E[(x_i - \mu_i)(x_j - \mu_j)]$

$= E[x_i x_j - x_i \mu_j - x_j \mu_i + \mu_i \mu_j]$

$= E[x_i x_j] - \mu_i \mu_j - \mu_j \mu_i + \mu_i \mu_j$

$\because x_i, x_j$  彼此獨立

$\int \int x_i x_j f(x_i, x_j) dx_i dx_j$

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

$= \mu_i \cdot \mu_j$

$= \mu_i \mu_j - \mu_i \mu_j$

$= 0 \Rightarrow$  矩陣之非對角線上皆為 0

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

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

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

Step 2:

$$M_Y = \text{COV}(Y, Y)$$

$$= \text{COV}(A+B, A+B)$$

$$= E[(X+Y) - E(X+Y)][(X+Y) - E(X+Y)]^T]$$

$i = j$  (在對角線上):

$$(M_Y)_{ii} = E[(X_i + Y_i) - (\mu_i + \hat{\mu}_i)]^2]$$

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

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

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

$$= \hat{\sigma}_i^2 + \hat{\sigma}_i^2 + 2E[x_i y_i - x_i \hat{u}_i - y_i \mu_i + \mu_i \hat{u}_i]$$

在 step 1 有證明過 = 0

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

$i \neq j$  (非對角線) =

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

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

$$= E[(x_i - \mu_i)(x_j - \mu_j)] + E[(x_i - \mu_i)(y_j - \hat{u}_j)] +$$

↓ step 1 有證明過 = 0

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

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

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

皆在前面有證明過 = 0

$$= 0$$

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

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

$$= M_A + M_B$$

∴ step 1 與 step 2 的結果相同

∴ 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

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

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

$$\text{令 } M = S_N^{-1}$$

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

↓

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

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

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

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

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

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

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 = 次式

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

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

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

$$\Rightarrow G_{N+1}^2 = \frac{1}{\beta} + \phi^T \left( S_N - \beta \frac{Q^+}{1 + \beta k^+} \right) \phi$$

$$= \frac{1}{\beta} + \phi^T S_N \phi - \beta \phi^T \frac{Q^+}{1 + \beta k^+} \phi$$

$$= \left( \frac{1}{\beta} + \phi^T S_N \phi \right) - \beta \left( \frac{\phi^T Q^+ \phi}{1 + \beta k^+} \right)$$

$$= G_N^2 - \beta \frac{\phi^T Q^+ \phi}{1 + \beta k^+}$$

$$\beta \frac{\phi^T Q^+ \phi}{1 + \beta k^+} \overset{\text{二次型}}{=} G_N^2 - G_{N+1}^2$$

∵ 分母、分子皆  $> 0$ ，且  $\beta > 0$

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

4.

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

$\uparrow$  向量       $\uparrow$  向量  
 $x$        $w$

↓ 對輸入加高斯誤差

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

$\underbrace{\hspace{1.5cm}}_{y_i(x_i, w)}$

誤差函數 =

$$\begin{aligned} E_D(w) &= \frac{1}{2} \sum_{i=1}^N (\hat{y}_i - t_i)^2 \\ &= \frac{1}{2} \sum_{i=1}^N \left( w_0 + \sum_{j=1}^D w_j (x_{ij} + \varepsilon_{ij}) - t_i \right)^2 \\ &= \frac{1}{2} \sum_{i=1}^N \left( w_0 + \underbrace{\sum_{j=1}^D w_j x_{ij}}_{y_i} + \sum_{j=1}^D w_j \varepsilon_{ij} - t_i \right)^2 \\ &= \frac{1}{2} \sum_{i=1}^N \left( y_i + \sum_{j=1}^D w_j \varepsilon_{ij} - t_i \right)^2 \end{aligned}$$

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

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

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

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

$$E\left[2t_i \sum_{j=1}^D w_j \varepsilon_{ij}\right] = \underbrace{2t_i \sum_{j=1}^D w_j}_{\text{常數可移出}} E[\varepsilon_{ij}]$$

↓ 題目已知為 0

$$= 0$$

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

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

$\Rightarrow$  個別平方項 + 2(相乘項)

$$= E[\text{個別平方項之和}] + 2 E[\text{任 2 項乘積之和}]$$

$$= \sum_{j=1}^D w_j^2 E[\varepsilon_{ij}^2] + 2 \sum w_m w_n E[\varepsilon_{im}] E[\varepsilon_{in}]$$

↓ 題目已知為  $\sigma^2$

↓ 0

↓ 0

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



$$E[(y_i - \bar{t}_i)^2] = (y_i - \bar{t}_i)^2 \quad \because \text{常數}$$

$$\begin{aligned} E[E_D(w)] &= \frac{1}{2} \sum_{i=1}^N \left[ (y_i - \bar{t}_i)^2 + \sum_{j=1}^D w_j \sigma^2 \right] \\ &= \frac{1}{2} \sum_{i=1}^N (y_i - \bar{t}_i)^2 + \underbrace{\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^D w_j \sigma^2}_{\text{常數}} \end{aligned}$$

$$= \frac{1}{2} \sum_{i=1}^N (y_i - \bar{t}_i)^2 + A \quad \#$$

↓  
原本未加雜訊的模型

∴ 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。