(1) Because x_n, x_m are from a Gaussian distribution

$$E[x_n] = \int_{-\infty}^{\infty} N(x_n|\mu_n,\sigma_n^2) = \mu_n$$

$$E[x_m] = \int_{-\infty}^{\infty} N(x_m|\mu_m,\sigma_m^2) = \mu_m$$

$E[x_n \cdot x_m]$, If $n \neq m$, 表示 x_n and x_m 不是同個 random variable

所以 $E[x_n x_m] = E[x_n]E[x_m] = \mu_n \cdot \mu_m$, 又 x_n, x_m from a distribution

$\mu_m = \mu_n = \mu \Rightarrow E[x_n x_m] = \mu^2$ — ①

If x_n and x_m 是同一個 r.v

$$E[x_n^2] = E[x_m^2] = E[x^2] = \int_{-\infty}^{\infty} N(x|\mu,\sigma^2) = \mu^2+\sigma^2 \text{ — ②}$$

$\Rightarrow$ ① + ② 結合

$E[x_n x_m] = \mu^2 + I_{nm}\sigma^2$, if $n=m$, $I_{nm}=1$, otherwise $I_{nm}=0$

$$\mu_{ML} = \frac{1}{N}\sum_{j=1}^{N} x_j \Rightarrow E[\mu_{ML}] = \frac{1}{N} E\left[\sum_{j=1}^{N} x_j\right] = \frac{1}{N}\left[E[x_1]+E[x_2]+\cdots+E[x_N]\right]$$

$$= \frac{1}{N} N \cdot \mu = \mu$$

$$\sigma_{ML}^2 = \frac{1}{N}\sum_{j=1}^{N}(x_j-\mu_{ML})^2$$

$$E[\mu_{ML}^2] = E\left[\frac{1}{N^2}\sum_{j=1}^{N}x_j \cdot \sum_{j=1}^{N}x_j\right] = \frac{N}{N^2}(\sigma^2+\mu^2) + \frac{N^2-N}{N^2}\mu^2$$

$$\Rightarrow E[\sigma_{ML}^2] = \frac{1}{N}E\left[\sum_{j=1}^{N}(x_j^2 - 2x_j\mu_{ML} + \mu_{ML}^2)\right], \quad E[2x_j\mu_{ML}] = 2E\left[x_j \frac{1}{N}\sum_{j=1}^{N}x_j\right] = \frac{\sigma^2+\mu^2}{N} + \frac{N-1}{N}\mu^2$$

$E[x_j^2]$

$$= \frac{1}{N}(N\cdot(\sigma^2+\mu^2)) - \frac{1}{N}2N\cdot\left(\frac{\sigma^2+\mu^2}{N} + \frac{N-1}{N}\mu^2\right) + \frac{1}{N}\cdot N\left(\frac{N}{N^2}(\sigma^2+\mu^2) + \frac{N^2-N}{N^2}\mu^2\right)$$

$$= \frac{N-1}{N}\sigma^2 \#$$

(2)

$$p(a, b) = p(a)p(b)$$

$$\vec{y} = \vec{a} + \vec{b} \quad \text{mean } y = \frac{\vec{a}+\vec{b}}{2} = \frac{\vec{a}}{2} + \frac{\vec{b}}{2} \ \#$$

$$\mathrm{cov}(\vec{a}+\vec{b}) = E\left[(\vec{a}+\vec{b} - E[\vec{a}+\vec{b}])(\vec{a}+\vec{b} - E[\vec{a}+\vec{b}])^T\right]$$

$$\mathrm{cov}(\vec{a}) = E\left[(\vec{a} - E[\vec{a}])(\vec{a} - E[\vec{a}])^T\right]$$

$$\mathrm{cov}(\vec{b}) = E\left[(\vec{b} - E[\vec{b}])(\vec{b} - E[\vec{a}])^T\right]$$

又 $E[\vec{a}+\vec{b}] = E[\vec{a}] + E[\vec{b}]$

$$\mathrm{cov}(\vec{a}+\vec{b}) = E\left[(\vec{a} - E[\vec{a}] + b - E[\vec{b}])(\vec{a} - E[\vec{a}] + \vec{b} - E[\vec{b}])^T\right]$$

$X = \vec{a} - E[\vec{a}]$, $Y = \vec{b} - E[\vec{b}] \Rightarrow \ = E[(X+Y)(X+Y)^T]$

$$= E[XX^T + XY^T + YX^T + YY^T] = E[XX^T] + E[YY^T] + E[XY^T] + E[YX^T]$$

又 $P(a,b) = P(a)P(b)$, we can know that $\vec{a}$ and $\vec{b}$ are independent

$$= \mathrm{cov}(X) + \mathrm{cov}(Y) + \mathrm{cov}(X,Y) + \mathrm{cov}(Y,X)$$

又 $\vec{a}$ and $\vec{b}$ are independent random varible, $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X) = 0$

$$= \mathrm{cov}(X) + \mathrm{cov}(Y) \ \#$$

(4)

ε_i 是 Gaussian noise，所以

$$y(x,w) = w_0 + \sum_{i=1}^{D} w_i x_i \Rightarrow \tilde{y}(x,w) = w_0 + \sum_{i=1}^{D} w_i(x_i + \varepsilon_i)$$
$$= y(x,w) + \sum_{i=1}^{D} \varepsilon_i w_i$$

Error function

$$E_r = \frac{1}{2}\sum_{n=1}^{N}\left(\tilde{y}(x_n,w) - t_n\right)^2$$
$$= \frac{1}{2}\sum_{n=1}^{N}\left(\tilde{y}^2(x_n,w) - 2\tilde{y}(x_n,w)t_n + t_n^2\right)$$
$$= \frac{1}{2}\sum_{n=1}^{N}\left(y(x_n,w)^2 + 2y(x,w)\sum_{i=1}^{D}\varepsilon_i w_i - 2y(x_n,w)t_n - 2\sum_{i=1}^{D}\varepsilon_i w_i t_n + t_n^2 + \left(\sum_{i=1}^{D} w_i\varepsilon_i\right)^2\right)$$

期望值 $E[\varepsilon_i] = 0$，所以對 $E[E_r]$，有 $E[\varepsilon_i]$ 都等於 0（因為是題目的條件都給期望值條件）

且 $E\left[\left(\sum_{i=1}^{D} w_i\varepsilon_i\right)^2\right] \Rightarrow \because$ w_i and ε_i are uncorrelated, so

$$E\left[\left(\sum_{i=1}^{D} w_i\varepsilon_i\right)^2\right] = E\left[(w_1\varepsilon_1 + \cdots w_n\varepsilon_n)(w_1\varepsilon_1 + \cdots + w_n\varepsilon_n)\right]$$

$\Rightarrow$ When $E[\varepsilon_i\varepsilon_j] = \delta_{ij}\sigma^2$ when $i \neq j$, $\delta_{ij} = 1$

if $i = j$, $E[\varepsilon_i\varepsilon_i] = 0$

$E\left[\left(\sum_{i=1}^{D} w_i\varepsilon_i\right)^2\right] = \sum_{i=1}^{D} w_i^2\sigma^2$，剩下的 $y(x_n,w)^2 - 2y(x_n,w)t_n + t_n^2 = E_D(w)$

$$\Rightarrow E[E_r] = E[E_D] + \frac{1}{2}\sum_{i=1}^{D} w_i^2\sigma^2$$

左邊為 Error function with noise input，如果找最小值，

等於找右式 sum of squared without noise input 的最小值再加上 weight decay regulation term #

(3) From Neural Networks book

$\sigma^2(x) = \sigma_V^2(x) + \phi^T(x) S^{-1} \phi(x)$, We add a single point located at $x^0 \Rightarrow \sigma^2(x)$

$= \sigma_V^{-2}(x)\,\phi(x^0)\,\phi^T(x^0)$

inverse Hessian $\Rightarrow (M+vv^T)^{-1}(M+vv^T) = M^{-1}(M+vv^T) - \dfrac{(M^{-1}v)(v^TM^{-1})}{1+v^TM^{-1}v}(M+vv^T)$

prior covariance matrix

$\Rightarrow C(x,x') = \phi^T S^{-1} \phi(x') \Rightarrow \sigma^2(x) = \sigma_V^2 + C(x,x) - \dfrac{C(x^0,x^0)^2}{\sigma_V^2 + C(x^0,x^0)}$

Now, we have N data points, and add x^{N+1} to the data

A is Hessian matrix $\Rightarrow A = \dfrac{1}{\sigma_V^2}\sum_{n=1}^{N}\phi(x^n)\phi(x^n)^T + S = \dfrac{1}{\sigma_V^2}\phi^T\phi + S$ (on the Qazaz book)

$\Rightarrow A_{N+1} = A_N + \underline{\sigma_V^{-2}\phi(x^{N+1})\phi^T(x^{N+1})}$

$\Rightarrow (A_{N+1})^{-1} = A_N^{-1} - \dfrac{A_N^{-1}\phi(x^{N+1})\phi^T(x^{N+1})A_N^{-1}}{\sigma_V^2 + \phi^T(x^{N+1})A_N^{-1}\phi(x^{N+1})}$

☆ Total variance of output predictions is (NN book)

內部雜訊 不確定性對於 weights.

$\sigma^2(x) = \sigma_V^2 + \sigma_W^2(x) = \sigma_V^2 + \phi^T(x) A^{-1} \phi(x)$

將$(A_{N+1})^{-1}$放入

$\sigma_{N+1}^2(x) = \sigma_N^2(x) - \dfrac{[\phi^T(x^{N+1})A_N^{-1}\phi(x)]^2}{\sigma_V^2 + \phi^T(x^{N+1})A_N^{-1}\phi(x^{N+1})}$,

從 A 可以知道, A is positive definite, 因為 $\frac{1}{\sigma_V^2} > 0$ $\phi \frac{1}{\sigma_V^2} \phi^T > 0$, 所以是 positive definite

所以 $\phi^T(x^{N+1})A_N^{-1}\phi(x) > 0$, $\sigma_{N+1}(x) = \sigma_N^2(x) - $(positive number)

$\Rightarrow \sigma_{N+1}^2(x) \le \sigma_N^2(x)$