2)

$y(x,w) = w^T \phi(x)$

where y is a K-dimensional column vector, W is an $M*K$ matrix of parameters, and $\phi(x)$ is a M-dimensional column vector with elements $\phi_j(x)$ with $\phi_0(x) = 1$

i) we assume that the target variable

$$t = y(x,w) + \underbrace{\epsilon}_{\text{Guassian noise}}$$

Here ϵ is zero mean Gaussian random variable With precision β.

β is inverse variance

i.e. $p(t|x,w,\beta) = N(t|y(x,w), \beta^{-1})$

∴ likelihood function with parameters w and β is

$$\boxed{p(t|X,w,\beta) = \prod_{n=1}^{N} N(t_n|w^T\phi(x_n), \beta^{-1})}$$

since we are not seeking to model the input variable we will drop x

Taking logarithm on both sides we get

$$\ln p(t|w,\beta) = \sum_{n=1}^{N} \ln N(t_n|w^T\phi(x_n), \beta^{-1})$$

$$= \sum_{n=1}^{N} \ln\left(\frac{1}{(2\pi\beta^{-1})^{1/2}} \exp\left(\frac{-1}{2\beta^{-1}}(t_n - w^T\phi(x_n))^2\right)\right)$$

$$= \sum_{n=1}^{N}\left(\frac{1}{2}\ln\left(\frac{\beta}{2\pi}\right) - \frac{\beta}{2}(t_n - w^T\phi(x_n))^2\right)$$

$$= \frac{N}{2}\ln\left(\frac{\beta}{2\pi}\right) - \frac{\beta}{2}\sum_{n=1}^{N}(t_n - w^T\phi(x_n))^2$$

Maximizing with respect to w

Here we can ignore the first term as it doesn't depend upon w and maximizing the likeli-hood function is equivalent to minimizing the second term since it is negative.

$$\therefore \nabla \ln p(t|w,\beta) = \sum_{n=1}^{N}(t_n - w^T\phi(x_n))\,\phi(x_n)^T\beta$$

setting this to zero we get

$$\beta\left(\sum_{n=1}^{N}(t_n - w^T\phi(x_n))\,\phi(x_n)^T\right) = 0$$

$$0 = \sum_{n=1}^{N}\left(t_n\phi(x_n)^T - w^T\phi(x_n)\phi(x_n)^T\right)$$

$\therefore$ solving for w

we get $\phi^T\phi\, w_{ML} = \phi^T t$

$$\Rightarrow \boxed{w_{ML} = (\phi^T\phi)^{-1}\phi^T t}$$

For multioutput let $T = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}$

w_{ML} become $\boxed{w_{ML} = (\phi^T\phi)^{-1}\phi^T T}$

MAP:

Now introducing prior distribution which will be in the form of gaussian with mean 0 and variance σ^2

~~p(w)~~

we write posterior as

$$\underbrace{P(w \mid x, t, \sigma^2, \beta)}_{\text{posterior}} \propto \underbrace{P(t \mid x, w, \beta)}_{\text{likelihood}} \underbrace{P(w)}_{\text{prior}} \quad -① $$

~~$p(w) = N(t \mid y(x, w$~~

$$P(w) = N(w \mid 0, \sigma^2)$$

$$= \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{w^T w}{2\sigma^2}\right)$$

applying log on prior we get

$$\ln(P(w)) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{w^T w}{2\sigma^2}$$

Therefore maximizing posterior is maximizing likehood and prior which is minimizing the negative terms in both.

$\therefore$ maximizing posterior is minimizing $\frac{\beta}{2}\sum_{n=1}^{N}(y(x_n, w) - t_n)^2$ ~~and~~ → from likelihood

and minimizing $\frac{w^T w}{2\sigma^2}$ → from prior.

i.e minimizing $\frac{\beta}{2}\sum_{n=1}^{N}(y(x_n, w) - t_n)^2 + \frac{w^T w}{2\sigma^2}$ ($\because$ log will be applied) substituting $y = w^T \phi(x)$

we need to minimize

$$\beta/2 \sum_{n=1}^{N} (w^T \phi(x_n) - t_n))^2 + \frac{w^T w}{2\sigma^{2}}$$

differentiating w.r.t w

we finally get

$$w = \left(\phi(x)^T \phi(x) + \frac{1}{\sigma^2 \beta} NI\right)^{-1} \phi(x)^T \cdot t$$

For multioutput let $T = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}$

then w

becomes $\boxed{w = \left(\phi(x)^T \phi(x) + \frac{NI}{\sigma^2 \beta}\right)^{-1} \phi(x)^T \cdot T}$

ii) Given $\phi(0) = (1,0)^T$

$\phi(1) = (0,1)^T$

$y = w^T \phi(x)$

x	y
0	$(-1,-1)^T$
0	$(-1,-2)^T$
0	$(-2,-1)^T$
1	$(1,1)^T$
1	$(1,2)^T$
1	$(2,1)^T$

Required to find MLE $w = [w_1 \; w_2]$

$$w = (\phi^T \phi)^{-1} \phi^T T$$

$$\phi(x) = \begin{bmatrix} \phi(0) \\ \phi(0) \\ \phi(0) \\ \phi(1) \\ \phi(1) \\ \phi(1) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

$$T = \begin{bmatrix} -1 & -1 \\ -1 & -2 \\ -2 & -1 \\ 1 & 1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}$$

$$\phi^T \phi = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \Rightarrow (\phi^T \phi)^{-1} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/3 \end{bmatrix}$$

$$(\phi^T \phi)^{-1} \phi^T = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \end{bmatrix}$$

$$(\phi^T \phi)^{-1} \phi^T . T = \begin{bmatrix} -4/3 & -4/3 \\ 4/3 & 4/3 \end{bmatrix}$$

$$\boxed{\begin{gathered} \therefore\ W = [w_1\ w_2] = \begin{bmatrix} -4/3 & -4/3 \\ 4/3 & 4/3 \end{bmatrix} \\ \text{i.e } w_1 = \begin{bmatrix} -4/3 \\ 4/3 \end{bmatrix} \text{ and } w_2 = \begin{bmatrix} -4/3 \\ 4/3 \end{bmatrix} \end{gathered}}$$