3.] Given :- We have a third degree polynomial

$$y = ax^3 + bx^2 + cx + d + v$$

$$\Rightarrow y = W^T b(x) + v$$

where, $W = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$ $b(x) = \begin{bmatrix} x^3 \\ x^2 \\ x \\ 1 \end{bmatrix}$

and $v \sim \mathcal{N}(0, \sigma^2)$

Prior of $W \Rightarrow W \sim \mathcal{N}(0, \gamma^2 I)$

W_{True} are the true parameters of y.

$\rightarrow$ Estimate the MAP function parameter W.

Solution :- $y = W^T b(x) + v$

From the linearity property $x = Az + b$,

we know $y \sim \mathcal{N}(W^T b(x), \sigma^2)$

To estimate W, we have our MAP estimator:

$$W_{MAP} = \underset{W}{\text{argmax}}\ p(W; y \mid x)$$

Applying Bayes Rule and disregarding the evidence;

$$W_{MAP} = \underset{W}{\text{argmax}}\ \underbrace{p(y \mid x ; W)}_{\text{likelihood}} * \underbrace{p(W)}_{\text{prior}}$$

Taking log on both sides:

$$\ln W_{MAP} = \underset{W}{\operatorname{argmax}} \sum_{i=1}^{K} \ln p(y_i | x_i ; W) + \ln p(W)$$

(Considering i samples)

Solving each term independently :-

$$\ln p(y_i | x_i ; W) = \ln \frac{1}{\sqrt{2\pi}\,\sigma} \times e^{-\frac{1}{2}\frac{(y_i - W^T b(x_i))^2}{\sigma^2}}$$

$$= \underbrace{\ln \frac{\sigma^{-1}}{\sqrt{2\pi}}}_{\text{Constant}} + \frac{1}{2} \frac{(y_i - W^T b(x_i))^2}{\sigma^2}$$

$$\ln p(W) = \ln \frac{|\gamma^2 I|^{-1/2}}{2\pi} \times e^{-\frac{1}{2} W^T |\gamma^2 I|^{-1} W}$$

$$= \underbrace{\ln \frac{|\gamma^2 I|^{-1/2}}{2\pi}}_{\text{Constant}} + \left(-\frac{1}{2}\right) W^T |\gamma^2 I|^{-1} W$$

$$\therefore \ln W_{MAP} = \underset{W}{\operatorname{argmax}} \sum_{i=1}^{K} \left[-\frac{1}{2} \frac{(y_i - W^T b(x_i))^2}{\sigma^2} \right] - \frac{1}{2} W^T |\gamma^2 I|^{-1} W$$

$$\Rightarrow \ln W_{MAP} = \underset{W}{\operatorname{argmin}} \sum_{i=1}^{K} \frac{(y_i - W^T b(x_i))^2}{\sigma^2} - W^T |\gamma^2 I|^{-1} W$$

To minimize the RHS, we take a gradient at W_{MAP}.

$$\nabla_{W=W_{MAP}} \Rightarrow \frac{2}{\sigma^2}\sum_{i=1}^{K} (y_i - W_{MAP}^T b(x_i))\, b(x_i)^T + (-2)(\gamma^2 I)^{-1} W_{MAP} = 0$$

$$\Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{K} b(x_i)\, y_i - \frac{1}{\sigma^2}\sum_{i=1}^{K} W_{MAP}^T\, b(x_i)\, b(x_i)^T +$$

$$(-1)(\gamma^2 I)^{-1} W_{MAP} = 0$$

Multiplying by σ^2:

$$\sum_{i=1}^{K} b(x_i)\, y_i = \sum_{i=1}^{K} b(x_i)\, b(x_i)^T W_{MAP} + \sigma^2 (\gamma^2 I)^{-1} W_{MAP}$$

$$\therefore \boxed{W_{MAP} = \left[\sum_{i=1}^{K} b(x_i)\, b(x_i)^T \cancel{W_{MAP}} + \sigma^2 (\gamma^2 I)^{-1}\right]^{-1} \times \sum_{i=1}^{K} b(x_i)\, y_i}$$

This is the MAP estimate of the ~~pa~~parameter W