Answer to the Question No 2

Part (a):

Data Matrix: $Z = \begin{bmatrix} 1 & x^{(1)1} & x^{(1)2} & \cdots & x^{(1)m} \\ 1 & x^{(2)1} & x^{(2)2} & \cdots & x^{(2)m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x^{(n)1} & x^{(n)2} & \cdots & x^{(n)m} \end{bmatrix}$

The weight vector: $w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_m \end{bmatrix}$

Then, $Zw = \begin{bmatrix} 1 & x^{(1)1} & x^{(1)2} & \cdots & x^{(1)m} \\ 1 & x^{(2)1} & x^{(2)2} & \cdots & x^{(2)m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x^{(n)1} & x^{(n)2} & \cdots & x^{(n)m} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_m \end{bmatrix}$

$$= \begin{bmatrix} w_0 + w_1 x^{(1)1} + w_2 x^{(1)2} + \cdots + w_m x^{(1)m} \\ w_0 + w_1 x^{(2)1} + w_2 x^{(2)2} + \cdots + w_m x^{(2)m} \\ \vdots \\ w_0 + w_1 x^{(n)1} + w_2 x^{(n)2} + \cdots + w_m x^{(n)m} \end{bmatrix}$$

$$= \begin{bmatrix} y(x^{(1)}) \\ y(x^{(2)}) \\ \vdots \\ y(x^{(n)}) \end{bmatrix} = \hat{y}$$

Therefore, we have proved that $\hat{y} = Zw$.

Answer to the Question No: 2

Part (b):

$\| Zw - t \|^2 = \| \hat{y} - t \|^2$ # from part a we know $\hat{y} = Zw$

$$= \left\| \begin{bmatrix} y(x^{(1)}) \\ y(x^{(2)}) \\ \vdots \\ y(x^{(n)}) \end{bmatrix} - \begin{bmatrix} t^{(1)} \\ t^{(2)} \\ \vdots \\ t^{(3)} \end{bmatrix} \right\|^2$$

$$= \left\| \begin{bmatrix} y(x^{(1)}) - t^{(1)} \\ y(x^{(2)}) - t^{(2)} \\ \vdots \\ y(x^{(n)}) - t^{(n)} \end{bmatrix} \right\|^2$$

$= \sum_n (y(x^{(n)}) - t^{(n)})^2$ # as stated in the question $\|v\|^2 = \sum_n v_n^2$

$= \sum_n (t^{(n)} - y(x^{(n)}))^2$ # $(t^{(i)} - y(x^{(i)}))^2 = (y(x^{(i)}) - t^{(i)})^2$

$= \ell(w)$ # definition of $\ell(w)$

Therefore proved

Answer to the Question No: 2

Part (c)

$$l(w) = \sum_n [t^{(n)} - y(x^{(n)})]^2$$

$$\frac{\partial l(w)}{\partial w_m} = \frac{\partial}{\partial w_m}\left[(t^{(1)} - y(x^{(1)}))^2 + (t^{(2)} - y(x^{(2)}))^2 + \dots + (t^{(n)} - y(x^{(n)}))^2\right]$$

$$= 2(t^{(1)} - y(x^{(1)}))\frac{\partial}{\partial w_m}(t^{(1)} - y(x^{(1)})) + \dots + 2(t^{(n)} - y(x^{(n)}))\frac{\partial}{\partial w_m}(t^{(n)} - y(x^{(n)}))$$

$$= 2(t^{(1)} - y(x^{(1)}))\frac{\partial}{\partial w_m}(t^{(1)} - w_0 - w_1 x^{(1)1} - \dots - w_m x^{(1)m}) + \dots +$$

$$2(t^{(n)} - y(x^{(n)}))\frac{\partial}{\partial w_m}(t^{(n)} - w_0 - w_1 x^{(n)} - \dots - w_m x^{(n)m})$$

$$= 2(t^{(1)} - y(x^{(1)}))(-x^{(1)m}) + \dots + 2(t^{(n)} - y(x^{(n)}))(-x^{(n)m})$$

$$= -2\left[(t^{(1)} - y(x^{(1)}))Z_{1m} + \dots + (t^{(n)} - y(x^{(n)}))Z_{nm}\right]$$

$$= -2\sum_n (t^{(n)} - y(x^{(n)}))Z_{nm}$$

Therefore proved.

Answer to the Question No 2

Part d

$$\frac{\partial l(w)}{\partial(w)} = \begin{bmatrix} \frac{\partial l(w)}{\partial w_1} \\ \frac{\partial l(w)}{\partial w_2} \\ \vdots \\ \frac{\partial l(w)}{\partial w_m} \end{bmatrix} = \begin{bmatrix} -2\sum_n (t^{(n)} - y(x^{(n)}))\, z_{n1} \\ -2\sum_n (t^{(n)} - y(x^{(n)}))\, z_{n2} \\ \vdots \\ -2\sum_n (t^{(n)} - y(x^{(n)}))\, z_{nm} \end{bmatrix} = -2 \begin{bmatrix} \sum_n (t^{(n)} - y(x^{(n)}))\, z_{n1} \\ \sum_n t^{(n)} - y(x^{(n)})\, z_{n2} \\ \vdots \\ \sum_n t^{(n)} - y(x^{(n)})\, z_{nm} \end{bmatrix}$$

$$= -2 \begin{bmatrix} z_{11} & z_{21} & \cdots & z_{m1} \\ z_{12} & z_{22} & \cdots & z_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ z_{1n} & z_{2n} & \cdots & z_{mn} \end{bmatrix} \cdot \begin{bmatrix} t^{(1)} - y(x^{(1)}) \\ t^{(2)} - y(x^{(2)}) \\ \vdots \\ t^{(n)} - y(x^{(n)}) \end{bmatrix}$$

by definition of matrix multiplication

$$= -2\, z^T \cdot \left(\begin{bmatrix} t^{(1)} \\ t^{(2)} \\ \vdots \\ t^{(n)} \end{bmatrix} - \begin{bmatrix} y(x^{(1)}) \\ y(x^{(2)}) \\ \vdots \\ y(x^{(n)}) \end{bmatrix} \right)$$

$$= -2\, z^T (t - \hat{y})$$

Answer to the Question No:2

Part(e)

$$\frac{\partial l(w)}{\partial w} = 0$$

$-2Z^T(t - \hat{y}) = 0$ # from part(d), we know $\frac{\partial l(w)}{\partial w} = -2Z^T(t-\hat{y})$

$Z^T(t - \hat{y}) = 0$

$Z^T t - Z^T \hat{y} = 0$

$Z^T t = Z^T \hat{y}$

$Z^T t = Z^T Z w$ # from part(a), we know $\hat{y} = Zw$

$(Z^T Z)^{-1} Z^T t = (Z^T Z)^{-1} (Z^T Z) w$ # multiply both side by $(Z^T Z)^{-1}$

$(Z^T Z)^{-1} Z^T t = w$

Therefore pooved

Answer to the Question No : 5

Part(a):

We know, $l(w) = \sum_n [t^{(n)} - y(x^{(n)})]^2 + \alpha \sum_{j=1}^{M} w_j^2$

$$\frac{\partial l(w)}{\partial w_m} = \frac{\partial}{\partial w_m} \sum_n [t^{(n)} - y(x^{(n)})]^2 + \alpha \frac{\partial}{\partial w_m} \sum_{j=1}^{M} w_j^2$$

$$= -2 \sum_n \left[\left(t^{(n)} - y(x^{(n)})\right) \cdot z_{nm} \right] + \alpha \frac{\partial}{\partial m} \left(w_1^2 + w_2^2 + \cdots + w_m^2\right)$$

from 2(c)

$$= -2 \sum_n \left[\left(t^{(n)} - y(x^{(n)})\right) \cdot z_{nm} \right] + 2\alpha w_m \; ; \quad m \geq 1$$

$$= 2\alpha w_m - 2 \sum \left[\left(t^{(n)} - y(x^{(n)})\right) \cdot z_{nm} \right].$$

Therefore, proved for $m \geq 1$

Part(b):

$$\frac{\partial L(w)}{\partial w_0} = -2 \sum \left[\left(t^{(n)} - y(x^{(n)})\right)^2 \right] + 0.$$

This is because, $w_0 = 0$ in regularization term, thus $\frac{\partial}{\partial w_0} w_0 = 0$.

Hence, only $\frac{\partial}{\partial w_m} \sum_n [t^{(n)} - y(x^{(n)})]^2$ gets evaluated.