Problem 1.

2. $LSE = \frac{1}{2N}\sum_{i=1}^{N}(y^{(i)}-t^{(i)})^2$ where $N=7$, $y^{(i)} \equiv g_{w,b}(x) = wx^{(i)}+b$ $x \in 1..7$

$$\Rightarrow \frac{1}{2N}\sum_{i=1}^{N}(wx^{(i)}+b-t^{(i)})(wx^{(i)}+b-t^{(i)})$$

$$= \frac{1}{2N}\sum_{i=1}^{N}\Big[w^2(x^{(i)})^2 + wbx^{(i)} - wx^{(i)}t^{(i)} + wbx^{(i)} + b^2 - bt^{(i)} - wx^{(i)}t^{(i)} - bt^{(i)} + (t^{(i)})^2\Big]$$

$$= \frac{1}{2N}\sum_{i=1}^{N}\Big[w^2(x^{(i)})^2 + 2wbx^{(i)} - 2wx^{(i)}t^{(i)} - 2bt^{(i)} + b^2 + (t^{(i)})^2\Big]$$

rearranging in the order requested:

$$\boxed{LSE = \frac{1}{2N}\sum_{i=1}^{N}\Big[(x^{(i)})^2 w^2 + b^2 + 2x^{(i)}wb - 2x^{(i)}t^{(i)}w - 2t^{(i)}b + (t^{(i)})^2\Big]}$$

3. Let $A_i = (x^{(i)})^2$, $B_i = 1$, $C_i = 2x^{(i)}$, $D_i = -2x^{(i)}t^{(i)}$, $E_i = -2t^{(i)}$, $F = +1$.

w, b which minimize the MSE occur at the critical points of the MSE function

$$\frac{\partial \mathcal{E}(w,b)}{\partial w} = \frac{\partial}{\partial w}\Big[\frac{1}{2N}\sum_{i=1}^{N} A_i w^2 + B_i b^2 + C_i wb + D_i w + E_i b + F\Big]$$

$$= \frac{1}{2N}\sum_{i=1}^{N}(2A_i w + C_i b + D_i)$$

$$\frac{\partial \mathcal{E}(w,b)}{\partial b} = \frac{\partial}{\partial b}\Big[\frac{1}{2N}\sum_{i=1}^{N} A_i w^2 + B_i b^2 + C_i wb + D_i w + E_i b + F\Big]$$

$$= \frac{1}{2N}\sum_{i=1}^{N}(2B_i b + C_i w + E_i)$$

Critical points - values for which the first derivative is 0.

$$\frac{\partial \mathcal{E}}{\partial w} = 0 \Rightarrow 0 = \frac{1}{2N}\sum_{i=1}^{N} 2A_i w + C_i b + D_i$$

$$0 = 2w\sum_{i=1}^{N} A_i + b\sum_{i=1}^{N} C_i + \sum_{i=1}^{N} D_i$$

substitute $A = \sum_{i=1}^{N} A_i$, $C = \sum_{i=1}^{N} C_i$, $D = \sum_{i=1}^{N} D_i$

$$0 = 2Aw + Cb + D$$

$$\boxed{w = \frac{-(Cb + D)}{2A}}$$

$$\frac{\partial \mathcal{E}}{\partial b} = 0 \Rightarrow 0 = \frac{1}{2N}\sum_{i=1}^{N}(2B_i b + C_i w + E_i)$$

$$0 = \frac{1}{2N}\, 2b\sum_{i=1}^{N} B_i + w\sum_{i=1}^{N} C_i + \sum_{i=1}^{N} E_i$$

sub $B = \sum_{i=1}^{N} B_i$, $C = \sum_{i=1}^{N} C_i$, $E = \sum_{i=1}^{N} E_i$

$$0 = 2Bb + Cw + E$$

$$\boxed{b = \frac{-(Cw+E)}{2B}}$$

4. $A = \sum_{i=1}^{7} A_i \Rightarrow \sum_{i=1}^{7} (x^{(i)})^2 = 1^2+2^2+3^2+4^2+5^2+6^2+7^2 = 140$

$B = \sum_{i=1}^{7} B_i \Rightarrow \sum_{i=1}^{7} (1) = 1+1+1+1+1+1+1 = 7$

$C = \sum_{i=1}^{7} C_i \Rightarrow \sum_{i=1}^{7} 2x^{(i)} = 2(1+2+3+4+5+6+7) = 56$

$D = \sum_{i=1}^{7} D_i \Rightarrow \sum_{i=1}^{7} -2x^{(i)}t^{(i)} = -2(1\times6 + 2\times4 + 3\times2 + 4\times1 + 5\times3 + 6\times6 + 7\times10)$
$= -290$

$E = \sum_{i=1}^{7} E_i \Rightarrow \sum_{i=1}^{7} -2t^{(i)} = -2(6+4+2+1+3+6+10) = -64$

$F = \sum_{i=1}^{7} F_i \Rightarrow \sum_{i=1}^{7} (+1) = +1+1+1+1+1+1+1 = +7$

$w = \frac{-(56b - 290)}{2(140)}$ $\qquad$ $b = \frac{-(56w - 64)}{2(7)}$

$280w + 56b = 290$ $\qquad$ $56w + 14b = 64$

solving system of equations. $\boxed{w = 17/28 = 0.607 \quad b = 15/7 = 2.143}$

$g_{w,b}(x) = 0.607x + 2.143$

Problem 2

1. $g_{\vec{w}}(\vec{x}) = \vec{x}\vec{w}$ where $\vec{x} = [x^{(i)}, 1]$

want $g_{\vec{w}}(\vec{x}) = \vec{x}\vec{w} \iff g_{w,b}(x) = wx + b = wx + b \cdot 1.$
(vector) (scalar).

If $\vec{x} = [x^{(i)}, 1]$ and $w = \begin{bmatrix} w \\ b \end{bmatrix}$, then $\vec{x}\vec{w} = [x^{(i)} \ 1]\begin{bmatrix} w \\ b \end{bmatrix}$

$= wx^{(i)} + b \cdot 1 = wx + b$ $\square$.

2. $\nabla_{\vec{w}} \|X\vec{w} - \vec{t}\|^2$. Let $h(\vec{w}) = X\vec{w} - \vec{t}$
$\Rightarrow \nabla_{\vec{w}} \|h(\vec{w})\|^2$
$= 2\nabla h(\vec{w})$
$\nabla h(\vec{w}) = X^T(X\vec{w} - \vec{t})$
$\Rightarrow 2[X^T(X\vec{w} - \vec{t})]$
$= 2(X^T X\vec{w} - X^T\vec{t})$
$= \boxed{2X^T X\vec{w} - 2X^T\vec{t}}$

3. Least squares loss: $L(\vec{y}, \vec{t}) = (\vec{y} - \vec{t})^2$
$\Rightarrow (X\vec{w} - \vec{t})^2$ where y = prediction output $= g_w(\vec{x}) = \vec{x}\vec{w}$

To minimize least squares loss, solve for $\vec{w}$ at the critical point.

$LSE_{min} = \underset{\vec{w}}{\operatorname{argmin}}(\|X\vec{w} - \vec{t}\|^2) \Rightarrow \nabla_{\vec{w}} \|X\vec{w} - \vec{t}\|^2 = 0.$

From (2), $\nabla_{\vec{w}} \|X\vec{w} - \vec{t}\|^2 = 2X^T X\vec{w} - 2X^T\vec{t} = 0$

The value of $\vec{w}$ that minimizes the loss ($\vec{w}^*$) satisfies $2X^T X\vec{w}^* - 2X^T\vec{t} = 0$ $\square$.

4. $2X^T X\vec{w}^* - 2X^T\vec{t} = 0.$
$2X^T X\vec{w}^* = 2X^T\vec{t}.$
$X^T X\vec{w}^* = X^T\vec{t}.$
If $X^T X$ is invertible, then $(X^T X)^{-1}(X^T X) = I$, where I is the identity matrix.
$(X^T X)^{-1}(X^T X)\vec{w}^* = (X^T X)^{-1}(X^T\vec{t})$

$\boxed{\vec{w}^* = (X^T X)^{-1} X^T\vec{t}}$

Problem 3

1. $A = \sum_{i=1}^{N} \vec{x}^{(i)} \vec{x}^{(i)T}$ where $\vec{x}^{(i)}$ is a $j \times 1$ column vector
$\vec{x}^{(i)T}$ is a $1 \times j$ row vector

$$= \sum_{i=1}^{N} \begin{bmatrix} x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_d^{(i)} \end{bmatrix} \begin{bmatrix} x_1^{(i)} & x_2^{(i)} & \dots & x_d^{(i)} \end{bmatrix}$$

$$= \sum_{i=1}^{N} \begin{bmatrix} x_1^{(i)} * x_1^{(i)} & x_1^{(i)} * x_2^{(i)} & \dots & x_1^{(i)} * x_d^{(i)} \\ x_2^{(i)} * x_1^{(i)} & x_2^{(i)} * x_2^{(i)} & \dots & x_2^{(i)} * x_d^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ x_d^{(i)} * x_1^{(i)} & x_d^{(i)} * x_2^{(i)} & \dots & x_d^{(i)} * x_d^{(i)} \end{bmatrix}$$

so $\boxed{A_{a,b} = \sum_{i=1}^{N} x_a^{(i)} * x_b^{(i)}}$

2. $\nabla \mathcal{E}(\vec{w}, D) = \nabla \left[\frac{1}{2N} \sum_{i=1}^{N} (g_{\vec{w}}(\vec{x}^{(i)}) - t^{(i)})^2 + \frac{\lambda}{2} \|\vec{w}\|^2 \right]$

$$= \frac{1}{2N} \sum_{i=1}^{N} \underbrace{\nabla [g_{\vec{w}}(\vec{x}^{(i)}) - t^{(i)}]^2}_{①} + \frac{\lambda}{2} \underbrace{\nabla \|\vec{w}\|^2}_{(2)}$$

Solving ①:

$\nabla [g_{\vec{w}}(\vec{x}^{(i)}) - t^{(i)}]^2 = \nabla [(\vec{x}^{(i)})^T \vec{w} - t^{(i)}]^2$ where $g_{\vec{w}}(\vec{x}^{(i)}) = (\vec{x}^{(i)})^T \vec{w}$
and $\vec{x}^{(i)}$ is a $j \times 1$ column vector.

$$\nabla [(\vec{x}^{(i)})^T \vec{w} - t^{(i)}]^2 = 2 [(\vec{x}^{(i)T})^T (\vec{x}^{(i)T} \vec{w} - t^{(i)})]$$
$$= 2 (\vec{x}^{(i)} [(\vec{x}^{(i)})^T \vec{w} - t^{(i)}]$$
$$= 2 (\vec{x}^{(i)} \vec{x}^{(i)T} \vec{w} - \vec{x}^{(i)} t^{(i)})$$
$$= 2 (\vec{x}^{(i)} \vec{x}^{(i)T} \vec{w} - t^{(i)} \vec{x}^{(i)})$$ since $t^{(i)}$ is a scalar.

Solving ②:

$\|\vec{w}\|^2 = (w_1^2 + w_2^2 + \dots + w_d^2)$

$$\nabla \|\vec{w}\|^2 = \begin{bmatrix} 2w_1 \\ \vdots \\ 2w_d \end{bmatrix} = 2 \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix} = 2\vec{w}.$$

Combining: $\nabla \mathcal{E}(\vec{w}, D) = \frac{1}{2N} \sum_{i=1}^{N} [2 (\vec{x}^{(i)} \vec{x}^{(i)T} \vec{w} - t^{(i)} \vec{x}^{(i)})] + \frac{\lambda}{2} (2\vec{w})$

$$= \frac{1}{N} \sum_{i=1}^{N} \left[\vec{x}^{(i)} \vec{x}^{(i)T} \vec{w} - t^{(i)} x^{(i)} \right] + \lambda \vec{w}$$

Substituting A from previous question and B:

$$\nabla \mathcal{E}(\vec{w}, D) = \frac{1}{N} \left(A\vec{w} - \vec{b} \right) + \lambda \vec{w}. \quad \square$$

3. $\vec{w}^* = \underset{\vec{w}}{\operatorname{argmin}} \left(\mathcal{E}(\vec{w}, D) \right).$

$$\underset{\vec{w}}{\operatorname{argmin}}\, \mathcal{E}(\vec{w}, D) \Rightarrow \frac{1}{N} \left(A\vec{w}^* - \vec{b} \right) + \lambda \vec{w}^* = 0.$$

$$\left(A\vec{w}^* - \vec{b} \right) + \lambda N \vec{w}^* = 0$$

$$A\vec{w}^* + \lambda N \vec{w}^* = \vec{b}$$

$$\boxed{(A + \lambda N)\vec{w}^* = \vec{b}}$$

4. Let $\vec{v}$ be the eigenvalues of A.

Then $A\vec{v} = \alpha \vec{v}$ for some eigenvalue α.

$$\rightarrow \vec{v}^T A \vec{v} = \vec{v}^T \alpha \vec{v}$$

$$\Rightarrow \vec{v}^T A \vec{v} = \alpha \underbrace{\vec{v}^T \vec{v}}$$

$$= \|\vec{v}\|^2 \geq 0.$$

Substitute $A = \sum_{i=1}^{N} x^{(i)} x^{(i)T}$ from part 1.

$$\Rightarrow v^T \sum_{i=1}^{N} x^{(i)} x^{(i)T} \vec{v} = \alpha \|\vec{v}\|^2$$

$$= \sum_{i=1}^{N} v^T x^{(i)} x^{(i)T} \vec{v} = \alpha \|\vec{v}\|^2.$$

Let $y^{(i)} = v^T x^{(i)}$.

$$\Rightarrow \sum_{i=1}^{N} \underbrace{y^{(i)} y^{(i)T}} = \alpha \|\vec{v}\|^2$$

$$= \|y^{(i)}\|^2 \geq 0.$$

$$\Rightarrow \sum_{i=1}^{N} \|y^{(i)}\|^2 = \alpha \|\vec{v}\|^2.$$

Since LHS ≥ 0, RHS ≥ 0 too.

RHS ≥ 0 if either
- $\rightarrow \alpha, \|\vec{v}\|^2 > 0$
- $\rightarrow \alpha, \|\vec{v}\|^2 < 0$

But by definition, $\|\vec{v}\|^2 \geq 0$. So $\alpha \geq 0$ and the eigenvalues of A are non-negative $\square$

5. The eigenvalues of $(A + \lambda N I)$ are values β that satisfy $(A + \lambda N I)\vec{v} = \beta \vec{v}$ for an eigenvector $\vec{v}$. We need $\beta > 0$.

$(A + \lambda N I)\vec{v} = A\vec{v} + \lambda N I \vec{v}$.

From part (4), we know $A\vec{v} = \alpha \vec{v}$ where α is a non-negative eigenvalue of A.

$\Rightarrow \alpha \vec{v} + \lambda N I \vec{v}$.

$\Rightarrow \alpha \vec{v} + \lambda N \vec{v}$ drop I with no consequence on value $\lambda N I \vec{v}$.

$= \underbrace{(\alpha + \lambda N)}_{\text{eigenvalues of } A + \lambda N I, \text{ need eigenvalues} > 0} \vec{v}$

Need $\alpha + \lambda N > 0$

$\rightarrow$ from part (4), we know $\alpha \geq 0$.

$\rightarrow$ need $\lambda N > 0$.

$\rightarrow$ from the problem definition, $\lambda > 0$ and $N > 0$ so $\lambda N > 0$.

Thus $\alpha + \lambda N > 0$.

Since $(A + \lambda N I)$ is a square matrix with non-zero eigenvalues, it is invertible.

6. From part (3): $(A + \lambda N)\vec{w}^* = \vec{b}$

$\rightarrow (A + \lambda N I_d)\vec{w}^* = \vec{b}$

Since we know $(A + \lambda N I_d)$ is invertible, we can solve for $\vec{w}^*$.

$\underbrace{(A + \lambda N I_d)^{-1}(A + \lambda N I_d)}_{I_d}\vec{w}^* = (A + \lambda N I_d)\vec{b}$

$\boxed{\vec{w}^* = (A + \lambda N I_d)^{-1}\vec{b}}$