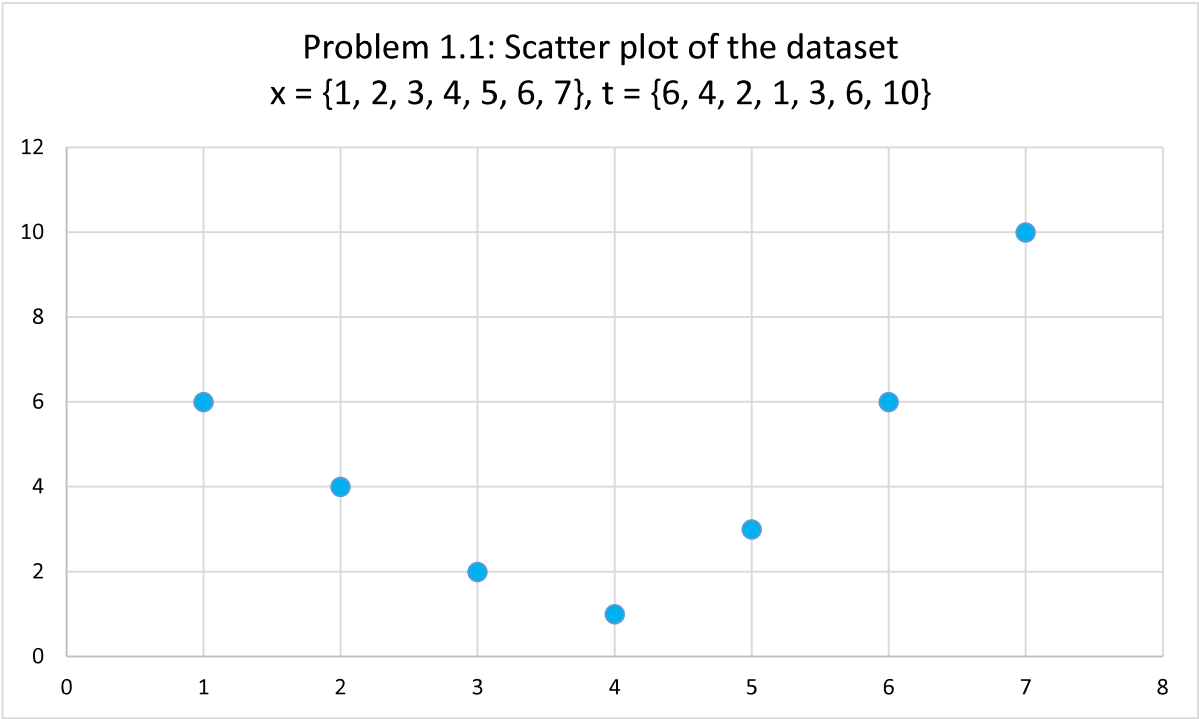


Problem 1

i	x(i)	t(i)	y(i)
1	1	6	2.75
2	2	4	3.357143
3	3	2	3.964286
4	4	1	4.571429
5	5	3	5.178571
6	6	6	5.785714
7	7	10	6.392857



Problem 1.

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

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

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

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

rearranging in the order requested:

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

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

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

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

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

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

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

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

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

$$0 = 2w \sum_{i=1}^N A_i + b \sum_{i=1}^N C_i + \sum_{i=1}^N D_i \quad \left. \begin{array}{l} \text{substitute } A = \sum_{i=1}^N A_i, C = \sum_{i=1}^N C_i \\ D = \sum_{i=1}^N D_i \end{array} \right\}$$

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

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

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

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

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

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

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

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

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

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

$$D = \sum_{i=1}^7 D_i \Rightarrow \sum_{i=1}^7 -2x^{(i)} t^{(i)} = -2(1 \times 6 + 2 \times 4 + 3 \times 2 + 4 \times 1 + 5 \times 3 + 6 \times 6 + 7 \times 10) = -290$$

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

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

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

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

$$280w + 56b = 290$$

$$56w + 14b = 64$$

solving system of equations.

$$w = 17/28 = 0.607 \quad b = 15/7 = 2.143$$

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

Problem 2

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

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

If $\vec{x} = [x^{(i)}, 1]$ and $w \in \begin{bmatrix} w \\ b \end{bmatrix}$, then $\vec{x} \vec{w} = [x^{(i)} \ 1] \begin{bmatrix} w \\ b \end{bmatrix}$
 $= wx^{(i)} + b \cdot 1 = wx + b \quad \square$

2. $\nabla_{\vec{w}} \|\vec{x} \vec{w} - \vec{e}\|^2$. Let $h(\vec{w}) = \vec{x} \vec{w} - \vec{e}$

$\Rightarrow \nabla_{\vec{w}} \|h(\vec{w})\|^2$

$= 2 \nabla h(\vec{w})$

$\nabla h(\vec{w}) = x^T (\vec{x} \vec{w} - \vec{e})$

$\Rightarrow 2 [x^T (\vec{x} \vec{w} - \vec{e})]$

$= 2 (x^T \vec{x} \vec{w} - x^T \vec{e})$

$= \boxed{2 x^T \vec{x} \vec{w} - 2 x^T \vec{e}}$

3. Least squares loss: $l(\vec{y}, \vec{e}) = (\vec{y} - \vec{e})^2$

$\Rightarrow (\vec{x} \vec{w} - \vec{e})^2$ where $y = \text{prediction output} = g_w(\vec{x}) = \vec{x} \vec{w}$

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

$l_{SEmin} = \underset{\vec{w}}{\operatorname{argmin}} (\|\vec{x} \vec{w} - \vec{e}\|^2) \Rightarrow \nabla_{\vec{w}} \|\vec{x} \vec{w} - \vec{e}\|^2 = 0$

From (2) $\nabla_{\vec{w}} \|\vec{x} \vec{w} - \vec{e}\|^2 = 2 x^T \vec{x} \vec{w} - 2 x^T \vec{e} = 0$

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

4. $2 x^T \vec{x} \vec{w}^* - 2 x^T \vec{e} = 0$

$2 x^T \vec{x} \vec{w}^* = 2 x^T \vec{e}$

$x^T \vec{x} \vec{w}^* = x^T \vec{e}$

If $x^T x$ is invertible, then $(x^T x)^{-1} (x^T x) = I$, where I is the identity matrix

$(x^T x)^{-1} (x^T x) \vec{w}^* = (x^T x)^{-1} (x^T \vec{e})$

$\boxed{\vec{w}^* = (x^T x)^{-1} x^T \vec{e}}$

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

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

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

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

2. $\nabla \mathcal{E}(\vec{w}, 0) = \nabla \left[\frac{1}{2N} \sum_{i=1}^N (g_{\vec{w}}(\vec{x}^{(i)}) - t^{(i)})^2 + \frac{\lambda}{2} \|\vec{w}\|^2 \right]$

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

Solving (1):

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

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

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: $\nabla \mathcal{E}(\vec{w}, 0) = \frac{1}{2N} \sum_{i=1}^N [2 (\vec{x}^{(i)} \vec{x}^{(i)T} \vec{w} - t^{(i)} \vec{x}^{(i)})] + \frac{\lambda}{2} (2\vec{w})$

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

Substituting A from previous question and B:

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

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

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

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

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

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

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

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

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

$$\Rightarrow \vec{v}^T A \vec{v} = \alpha \underbrace{\vec{v}^T \vec{v}}_{= \|\vec{v}\|^2} = \alpha \|\vec{v}\|^2 \geq 0.$$

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

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

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

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

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

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

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

Since LHS ≥ 0 , RHS ≥ 0 too.

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

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 $(A + \lambda NI)$ are values β that satisfy $(A + \lambda NI)\vec{v} = \beta\vec{v}$ for an eigenvector \vec{v} . We need $\beta > 0$.

$$(A + \lambda NI)\vec{v} = A\vec{v} + \lambda NI\vec{v}$$

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

$$\Rightarrow \alpha\vec{v} + \lambda NI\vec{v}$$

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

drop I with no consequence on value $\lambda NI\vec{v}$.

$$= (\alpha + \lambda N)\vec{v}$$

eigenvalues of $A + \lambda NI$, need eigenvalues > 0 .

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

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

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

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

$$\underbrace{(A + \lambda NI_d)^{-1}(A + \lambda NI_d)}_{Id.} \vec{w}^* = (A + \lambda NI_d)^{-1} \vec{b}$$

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