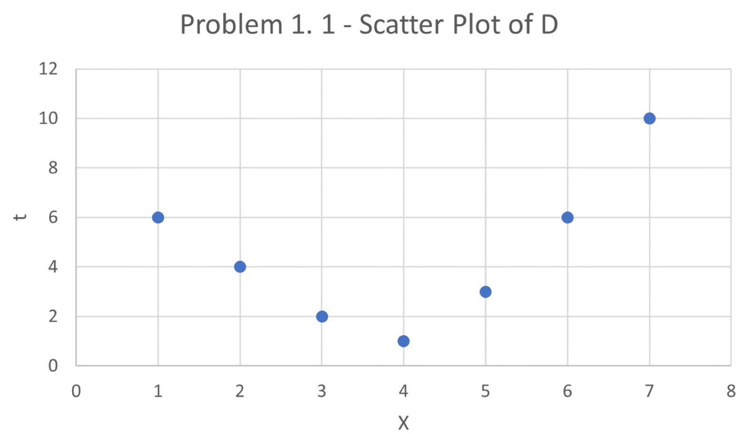
ECE421 Assignment 2

Problem 1

1.



2. $g_{w,b}(x) = wx + b \quad , N = 7$

$$\mathcal{E}(w,b) = \frac{1}{2N} \sum_{i=1}^{7} (g^{(i)} - t^{(i)})^2$$

$$\Rightarrow \mathcal{E}(w,b) = \frac{1}{2N} \sum_{i=1}^{7} (wx^{(i)} + b - t^{(i)})^2$$

$$= \frac{1}{2N} \sum_{i=1}^{7} w^2 x^{(i)^2} + 2wb\,x^{(i)} - 2wt^{(i)}x^{(i)} + b^2 - 2bt^{(i)} + t^{(i)^2}$$

$$= \frac{1}{2N} \sum_{i=1}^{7} x^{(i)^2} w^2 + b^2 + 2x^{(i)} wb - 2t^{(i)}x^{(i)} w - 2t^{(i)} b + t^{(i)^2}$$

where $A_i = x^{(i)^2}$, $B_i = 1$, $C_i = 2x^{(i)}$, $D_i = (-2)t^{(i)}x^{(i)}$, $E_i = (-2)t^{(i)}$, $F_i = t^{(i)^2}$

3. To minimize the mean squared loss ($\mathcal{E}(w,b)$), calculate $\frac{d\mathcal{E}}{dw}$ and $\frac{d\mathcal{E}}{db}$

$$\mathcal{E}(w,b) = \frac{1}{2\times 7}(Aw^2 + Bb^2 + Cwb + Dw + Eb + F)$$

$$= \frac{1}{14}(Aw^2 + Bb^2 + Cwb + Dw + Eb + F)$$

$$\frac{d\mathcal{E}}{dw} = 0$$

$$\Rightarrow \frac{1}{14}(2Aw + Cb + D) = 0 \quad \Rightarrow 2Aw + Cb + D = 0 \quad \Rightarrow \boxed{w = \frac{-(D + Cb)}{2A}}$$

$$\frac{d\mathcal{E}}{db} = 0$$

$$\Rightarrow \frac{1}{14}(2Bb + Cw + E) = 0 \quad \Rightarrow 2Bb + Cw + E = 0 \quad \Rightarrow \boxed{b = \frac{-(E + Cw)}{2B}}$$

4. $A = \sum_{i=1}^{7} x^{(i)^2} = 140$ $\qquad B = \sum_{i=1}^{7} 1 = 7$ $\qquad C = \sum_{i=1}^{7} 2x^{(i)} = 2 \times 28 = 56$

$D = \sum_{i=1}^{7} (-2)t^{(i)}x^{(i)} = -2 \times 145 = -290$ $\qquad E = \sum_{i=1}^{7} (-2)t^{(i)} = -2 \times 32 = -64$

So

$$\begin{cases} w = \dfrac{-(-290 + 56b)}{2 \times 140} = \dfrac{290 - 56b}{280} \\ b = \dfrac{-(-64 + 56w)}{2 \times 7} = \dfrac{64 - 56w}{14} \end{cases}$$

$$\Rightarrow \begin{cases} w = \dfrac{17}{28} \approx 0.6071 \\ b = \dfrac{15}{7} \approx 2.1429 \end{cases}$$

Problem 2

1. x is a 2-dimensional row vector $\Rightarrow \vec{x} = [x, 1]$ where $\vec{x}^{(i)} = (x^{(i)}, 1)$

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

To eliminate b:

when $\vec{w} = \begin{bmatrix} w \\ b \end{bmatrix}$,

$$\vec{x}\vec{w} = g_w(\vec{x}) = [x, 1]\begin{bmatrix} w \\ b \end{bmatrix} = xw + b = g_{w,b}(x)$$

2.

$$X = \begin{bmatrix} x^1 & 1 \\ x^2 & 1 \\ \vdots & \\ x^7 & 1 \end{bmatrix} \quad \vec{w} = \begin{bmatrix} w \\ b \end{bmatrix} \quad \vec{t} = \begin{bmatrix} t^1 \\ t^2 \\ \vdots \\ t^7 \end{bmatrix}$$

$$X\vec{w} - \vec{t} = \begin{bmatrix} x^1 & 1 \\ x^2 & 1 \\ \vdots & \\ x^7 & 1 \end{bmatrix} \cdot \begin{bmatrix} w \\ b \end{bmatrix} - \begin{bmatrix} t^1 \\ t^2 \\ \vdots \\ t^7 \end{bmatrix} = \begin{bmatrix} x^1 w + b - t^1 \\ x^2 w + b - t^2 \\ \vdots \\ x^7 w + b - t^7 \end{bmatrix}$$

$$\Rightarrow \|X\vec{w} - \vec{t}\|^2 = \sum_{i=1}^{7} (x^{(i)} w + b - t^{(i)})^2$$

$$\frac{d}{dw}(\|X\vec{w} - \vec{t}\|^2) = \frac{d}{dw}\left[\sum_{i=1}^{7} x^{(i)^2} w^2 + 2(b - t^{(i)}) x^{(i)} w + (b - t^{(i)})^2\right]$$

$$= 2 \cdot \sum_{i=1}^{7} x^{(i)^2} w + 2 \sum_{i=1}^{7} (b - t^{(i)}) x^{(i)}$$

$$= 2 \cdot 140 \cdot w + 2 \cdot 28 \cdot b - 2 \cdot 145$$

$$= 280w + 56b - 290$$

$$\frac{d}{db}\left(\|Xw-\vec{t}\|^2\right) = \frac{d}{db}\left[\sum_{i=1}^{7} x^{(i)^2}w^2 + 2(b-t^{(i)})x^{(i)}w + (b-t^{(i)})^2\right]$$

$$= 2\sum_{i=1}^{7} 1\cdot b + \sum_{i=1}^{7} 2x^{(i)}w - \sum_{i=1}^{7} 2t^{(i)}$$

$$= 14b + 56w - 64$$

$$\Rightarrow \nabla_{\vec{w}}\|X\vec{w}-\vec{t}\|^2 = \begin{bmatrix} 280w + 56b - 290 \\ 14b + 56w - 64 \end{bmatrix}$$

*-2: Derive the gradient analytically with ve

3.

When $\frac{d\varepsilon}{dw} = 0$ $\frac{d\varepsilon}{db} = 0$, we can find $\vec{w}^*$ minimize $\varepsilon(w, b)$

$\Rightarrow \vec{w}^*$ must satisfy $\nabla_{\vec{w}}\|X\vec{w}-\vec{t}\|^2 = 0$

$$\|x^{(i)}\vec{w} - t^{(i)}\|^2 = x^{(i)^2}\vec{w}^2 + 2x^{(i)}\vec{w}\,t^{(i)} + t^{(i)^2}$$

When X is a $N\times 2$ matrix, we can use X^TX to calculate $x^{(i)^2}$

$$\|X\vec{w}-\vec{t}\|^2 = \boxed{X^TX\vec{w}^2} - 2X\vec{w}\vec{t} + \vec{t}^2$$

$$\Rightarrow \nabla_{\vec{w}}\left(X^TX\vec{w}^2 - 2X\vec{w}\vec{t} + \vec{t}^2\right) = 2X^TX\vec{w} - 2X\vec{t}$$

$\Rightarrow \vec{w}^*$ must satisfy $2X^TX\vec{w}^* - 2X\vec{t} = 0$

4.

$$2X^TX\vec{w}^* - 2X^T\vec{t} = 0$$

$$\Rightarrow 2X^TX\vec{w}^* = 2X^T\vec{t}$$

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

When $X = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \\ 6 & 1 \\ 7 & 1 \end{bmatrix}$ $\vec{t} = \begin{bmatrix} 6 \\ 4 \\ 2 \\ 1 \\ 3 \\ 6 \\ 10 \end{bmatrix}$, $\vec{w}^* = \begin{bmatrix} \frac{17}{28} \\ \frac{15}{7} \end{bmatrix}$

Problem 3

1. $\vec{x}^{(i)} = \{x_j^{(i)}\}_{j \in 1 \cdots d} = \begin{bmatrix} x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_d^{(i)} \end{bmatrix}$

$x^{(i)}$ is a $d \times 1$ matrix $\Rightarrow \vec{x}^{(i)} \cdot \vec{x}^{(i)T}$ is a $d \times d$ matrix

for $j \in 1 \cdots d$, jth row in the matrix is $x_j^{(i)} \cdot [x_1^{(i)} \cdots x_d^{(1)}]$

$$= x_j^{(i)} (x_k^{(i)})_{k \in 1 \cdots d}$$

So $A = \sum_{i=1}^{N} \left[x_j^{(i)} (x_k^{(i)})_{k \in 1 \cdots d} \right]_{j \in 1 \cdots d}$ where $x_j^{(i)} = x_k^{(i)}$ when $j = k$

For example, the component in 2nd raw and 3rd column is $\sum_{i=1}^{N} x_2^{(i)} \cdot x_3^{(i)}$

2. $\vec{x}^{(i)}$ is a column vector

$g_w(\vec{x}) = \vec{x}\vec{w} \Rightarrow g_{\vec{w}}(\vec{x}^{(i)}) = \vec{x}^{(i)T}\vec{w}$

$\mathcal{E}(\vec{w}, D) = \frac{1}{2N} \sum_{i=1}^{N} (g_{\vec{w}}(x^{(i)}) - t^{(i)})^2 + \frac{\lambda}{2} \|\vec{w}\|_2^2$

$\Rightarrow \mathcal{E}(\vec{w}, D) = \frac{1}{2N} \sum_{i=1}^{N} (\vec{x}^{(i)T}\vec{w} - t^{(i)})^2 + \frac{\lambda}{2} \|\vec{w}\|^2$

$\sum_{i=1}^{N} (\vec{x}^{(i)T}\vec{w} - t^{(i)})^2 = \sum_{i=1}^{N} \boxed{\vec{x}^{(i)T} \cdot \vec{x}^{(i)}} \vec{w}^2 - 2\vec{x}^{(i)T}\vec{w}\, t^{(i)} + t^{(i)2}$

$= A\vec{w}^2 - 2b\vec{w} + \sum_{i=1}^{N} t^{(i)2}$

-3 incorrect manipulation. Does n

$\Rightarrow \mathcal{E}(\vec{w}, D) = \frac{1}{2N}(A\vec{w}^2 - 2b\vec{w} + \sum_{i=1}^{N} t^{(i)2}) + \frac{\lambda}{2}\|\vec{w}\|^2$

$\nabla \mathcal{E}(\vec{w}, D) = \frac{d}{d\vec{w}} \mathcal{E}(\vec{w}, D)$

$= \frac{1}{2N} \cdot 2A\vec{w} - \frac{1}{2N} \times 2b + \frac{\lambda}{2} \cdot 2\vec{w}$

$= \frac{1}{N} A\vec{w} - \frac{b}{N} + \lambda\vec{w}$

$= \frac{1}{N}(A\vec{w} - b) + \lambda\vec{w}$

3. $\vec{w}^* = \underset{\vec{w}}{\arg\min}\ \mathcal{E}(\vec{w}, D)$

$\Rightarrow$ Find $\vec{w}^*$ at $\nabla \mathcal{E}(\vec{w}, D) = 0$

$$\Rightarrow \frac{1}{N}(A\vec{w}^* - \vec{b}) + \lambda \vec{w}^* = 0$$

$$\frac{1}{N} A\vec{w}^* + \lambda \vec{w}^* - \frac{1}{N}\vec{b} = 0$$

4.

$$A = \sum_{i=1}^{N} x^{(i)} x^{(i)T}$$

Let $\vec{v}$ be the eigenvector, λ_a be the eigenvalues

$$A\vec{v} = \lambda_a \vec{v}$$

$$\vec{v}^T \cdot A\vec{v} = \sum_{i=1}^{N} \vec{v}^T x^{(i)} x^{(i)T} \vec{v}$$

the inner part $\vec{v}^T x^{(i)} x^{(i)T} \vec{v} = (x^{(i)} \vec{v})^T \cdot (x^{(i)} \vec{v})$

$$= \| x^{(i)} \vec{v} \|^2 \geq 0$$

$$\sum_{i=1}^{N} \vec{v}^T x^{(i)} x^{(i)T} \vec{v} \geq 0$$

$\Rightarrow A$ is a positive semidefinite matrix, $\vec{v}^T A \vec{v} \geq 0$

$$\vec{v}^T A \vec{v} = \vec{v}^T (\lambda_a \vec{v})$$

$$= \vec{v}^T \vec{v} \cdot \lambda_a \geq 0$$

$\vec{v}^T \vec{v}$ must $\geq 0 \Rightarrow \lambda_a \geq 0$, all eigenvalues of A are non-negative

5. $\lambda > 0$, $N > 0$, I_d is the identity $d \times d$ matrix $\begin{bmatrix} 1 & \cdots & & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & 1 \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & 1 \end{bmatrix}$

$\Rightarrow$ all elements in $\lambda N I_d$ must be greater or equal to 0

A is a positive matrix

$\Rightarrow A + \lambda N I_d$ is also a positive matrix ---- ① What do you mean by positive ma

Let $\vec{v}$ be the eigenvector

If eigenvalue $\lambda_a = 0$, $(A + \lambda N I_d)\vec{v} = \lambda_a \vec{v} = 0$

$\vec{v}$ is by definition non-zero

$\Rightarrow A + \lambda N I_d = 0$, which contradicts to ①

Hence, $A + \lambda N I_d$ has no eigenvalue equal to 0 This proof does not work. It is possible to have non-z

6. $\frac{1}{N} A \vec{w} + \lambda \vec{w} - \frac{1}{N} \vec{b} = 0$

$$\Rightarrow \quad \frac{1}{N}(A + \lambda N I_d)\vec{w} - \frac{1}{N}\vec{b} = 0$$

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

$$= \vec{b}\,(A + \lambda N I_d)^{-1}$$ inverse should be on left side

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