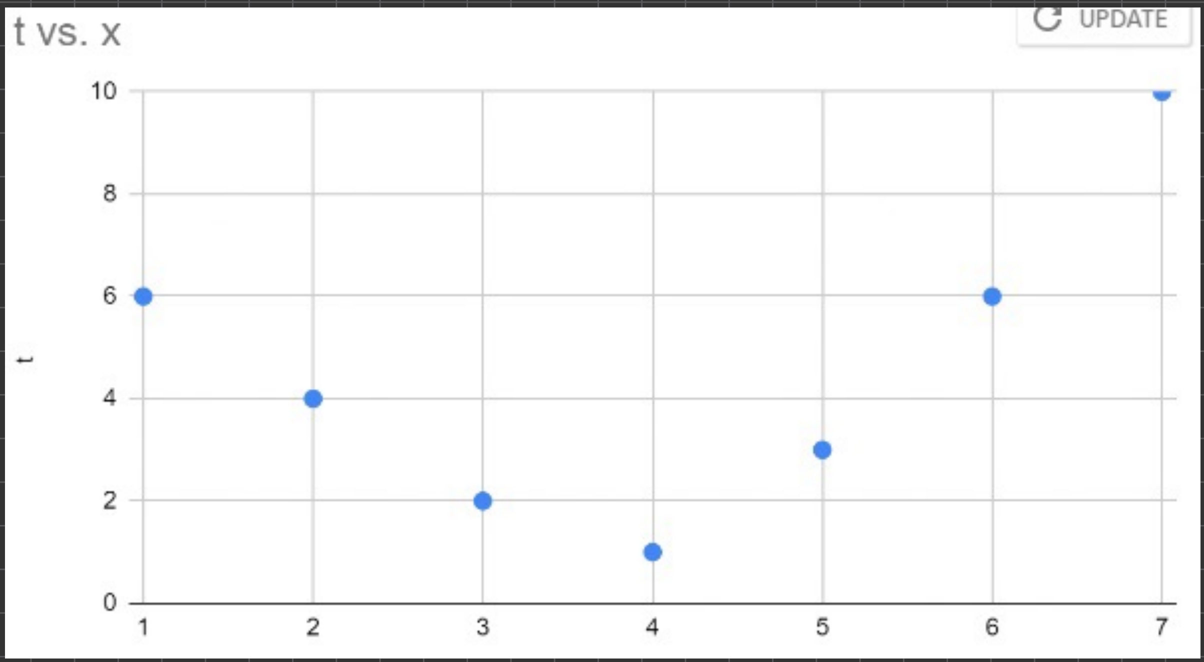
ECE421 Assignment #2

Problem 1.1



Problem 1.2

$$\mathcal{L} = \frac{1}{2N}\sum_{i=1}^{N}\left(wx^{(i)} + b - t^{(i)}\right)^2, \quad x = \sum_{i=1}^{N} x^{(i)}, \quad t = \sum_{i=1}^{N} t^{(i)}$$

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

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

$$\mathcal{L} = \frac{1}{2N}\sum_{i=1}^{N}\Big(w^2 \underbrace{x^{(i)^2}}_{A_i} \underbrace{-2x^{(i)}t^{(i)}}_{D_i} w + \underbrace{2x^{(i)}}_{C_i} wb + \underbrace{1}_{B_i} b^2 \underbrace{-2t^{(i)}}_{E_i} b + \underbrace{t^{(i)^2}}_{F_i} \Big)$$

Problem 1.3

$A = \sum_{i=1}^{N} x^{(i)^2}$ $\quad B = N$ $\quad C = 2x$ $\quad D = -2\sum_{i=1}^{N} x^{(i)}t^{(i)}$

$E = -2t$ $\quad F = \sum_{i=1}^{N} t^{(i)^2}$

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{1}{2N}\left(2wA + D + Cb\right) = 0$$

$$\rightarrow \quad w = \frac{-D - Cb}{2A} \quad \rightarrow \quad b = \frac{2Aw + D}{-C}$$

$$\frac{\partial \mathcal{L}}{\partial b} = \frac{1}{2N}\left(Cw + 2Bb + E\right) = 0$$

$$\rightarrow \quad w = \frac{-2Bb - E}{C} \quad \rightarrow \quad b = \frac{wC + E}{-2B}$$

$$\frac{\partial \mathcal{L}}{\partial b} = \frac{\partial \mathcal{L}}{\partial w}$$

$$\rightarrow \quad \frac{-2Bb}{C} - \frac{E}{C} = \frac{-D}{2A} - \frac{Cb}{2A}$$

$$\rightarrow \quad \frac{-2Bb}{C} + \frac{Cb}{2A} = \frac{E}{C} - \frac{D}{2A}$$

$$\rightarrow \quad b\left(\frac{C}{2A} - \frac{2B}{C}\right) = \frac{E}{C} - \frac{D}{2A}$$

$$b = \left(\frac{E}{C} - \frac{D}{2A}\right) \div \left(\frac{C}{2A} - \frac{2B}{C}\right)$$

$$\frac{2Aw + D}{-C} = \frac{wC + E}{-2B} \rightarrow -\frac{2Aw}{C} + \frac{Cw}{2B} = \frac{D}{C} - \frac{E}{2B}$$

Problem 1.4

$$\rightarrow w\left(\frac{C}{2B} - \frac{2A}{C}\right) = \frac{D}{C} - \frac{E}{2B}$$

$$w = \left(\frac{D}{C} - \frac{E}{2B}\right) \div \left(\frac{C}{2B} - \frac{2A}{C}\right)$$

$A = \sum_{i=1}^{N} x^{(i)^2}$ $\quad B = N$ $\quad C = 2x$ $\quad D = -2\sum_{i=1}^{N} x^{(i)} t^{(i)}$

$E = -2t$ $\quad F = \sum_{i=1}^{N} t^{(i)^2}$

$$w = \left(\frac{-\cancel{2}(145)}{\cancel{2}(28)} - \frac{-\cancel{2}(32)}{\cancel{2}(7)}\right) \div \left(\frac{\cancel{2}(28)}{\cancel{2}(7)} - \frac{2(140)}{\cancel{2}(28)}\right)$$

$$-0.607 \div -1$$

$$\underline{w = 0.607}$$

$$b = \left(\frac{E}{C} - \frac{D}{2A}\right) \div \left(\frac{C}{2A} - \frac{2B}{C}\right)$$

$$= \left(\frac{-2(32)}{2(28)} - \frac{-2(145)}{2(140)}\right) \div \left(\frac{2(28)}{2(140)} - \frac{2(7)}{2(28)}\right)$$

$$= 2.14$$

Problem 2

2.1

$$g_{\omega}(\vec{x}) = \vec{x}\,\vec{\omega} = \begin{bmatrix} x^{(i)} & 1 \end{bmatrix} \begin{bmatrix} \omega \\ b \end{bmatrix} \longrightarrow \underbrace{\omega x^{(i)} + b}_{\text{Same form as problem 1}} = g_{\omega, b}(x)$$

$$\therefore \vec{\omega} = \begin{bmatrix} \omega \\ b \end{bmatrix}$$

P2.2

$$\nabla_{\vec{\omega}} \| \mathcal{X}\vec{\omega} - \vec{t} \|^2 \qquad \vec{t} = \begin{bmatrix} t^{(1)} \\ \vdots \\ t^{(N)} \end{bmatrix} \quad \mathcal{X} = \begin{bmatrix} -\vec{x}^{(1)}- \\ \vdots \\ -\vec{x}^{(N)}- \end{bmatrix} = \begin{bmatrix} x^{(1)} & 1 \\ \vdots & \vdots \\ x^{(N)} & 1 \end{bmatrix} \quad \mathcal{X} \in \mathbb{R}^{N \times 2}$$

$$\vec{\omega} = \begin{bmatrix} \omega \\ b \end{bmatrix}$$

$$\mathcal{X}\vec{\omega} - \vec{t} = \begin{bmatrix} \omega x^{(1)} + b - t^{(1)} \\ \vdots \\ \omega x^{(N)} + b - t^{(N)} \end{bmatrix} \in \mathbb{R}^{N \times 1}$$

$$\| \mathcal{X}\vec{\omega} - \vec{t} \|^2 = \sum_{i=1}^{N} \left(\omega x^{(i)} + b - t^{(i)} \right)^2$$

$$\nabla \overbrace{\| \mathcal{X}\vec{\omega} - \vec{t} \|^2}^{\mathcal{L}} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial \omega} \\ \frac{\partial \mathcal{L}}{\partial b} \end{bmatrix}$$

because the summation is linear:

$$\sum_{i=1}^{N} \nabla \left(wx^{(i)} + b - t^{(i)} \right)^2$$

$$= \sum_{i=1}^{N} \begin{bmatrix} 2\left(wx^{(i)} + b - t^{(i)}\right)\left(x^{(i)}\right) \\ 2\left(wx^{(i)} + b - t^{(i)}\right)(1) \end{bmatrix}$$

$$\Longrightarrow \begin{bmatrix} \frac{\partial \mathcal{J}}{\partial w} \\ \frac{\partial \mathcal{J}}{\partial b} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} \left(2\left(wx^{(i)} + b - t^{(i)}\right)\left(x^{(i)}\right) \right) \\ \sum_{i=1}^{N} \left(2\left(wx^{(i)} + b - t^{(i)}\right) \right) \end{bmatrix}$$

P2.3

Take $\Sigma = \sum_{i=1}^{N}$

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

$(2 \times N)(N \times 2)$

2×2

$$\begin{array}{c} (1) \\ (2) \end{array} \begin{bmatrix} x^{(i)} & - & - & - \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x^{(i)} & 1 \\ - & 1 \\ - & 1 \\ - & 1 \end{bmatrix}$$

$$\begin{bmatrix} \overbrace{(1)\cdot(1)}^{\Sigma x^{(i)2}} & \overbrace{(1)\cdot(2)}^{\Sigma x^{(i)}} \\ \underbrace{(2)\cdot(1)}_{\Sigma x^{(i)}} & \underbrace{(2)\cdot(2)}_{N} \end{bmatrix} =$$

$$2\begin{bmatrix} \Sigma x^{(i)2} & \Sigma x^{(i)} \\ \Sigma x^{(i)} & N \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} - 2\begin{bmatrix} \Sigma x^{(i)} t^{(i)} \\ \Sigma t^{(i)} \end{bmatrix} = 0$$

Performing this multiplication gives us:

$$2\begin{bmatrix} w\Sigma x^{(i)2} + b\,\Sigma x^{(i)} \\ w\,\Sigma x^{(i)} + Nb \end{bmatrix} - 2\begin{bmatrix} \Sigma x^{(i)} t^{(i)} \\ \Sigma t^{(i)} \end{bmatrix} = 0$$

which leaves us with two equations

$$\begin{bmatrix} w\,\Sigma x^{(i)2} + b\,\Sigma x^{(i)} - \Sigma x^{(i)} t^{(i)} \\ w\,\Sigma x^{(i)} + Nb - \Sigma t^{(i)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Longrightarrow \quad w = \frac{\Sigma\left(x^{(i)} t^{(i)}\right) - b\,\Sigma\left(x^{(i)}\right)}{\Sigma\left(x^{(i)2}\right)} = \frac{\Sigma\left(t^{(i)}\right) - Nb}{\Sigma\left(x^{(i)}\right)}$$

$$\longrightarrow \quad b = \left(\frac{\Sigma\left(t^{(i)}\right)}{\Sigma\left(x^{(i)}\right)} - \frac{\Sigma\left(x^{(i)} t^{(i)}\right)}{\Sigma\left(x^{(i)2}\right)}\right) \div \left(\frac{N}{\Sigma\left(x^{(i)}\right)} - \frac{\Sigma\left(x^{(i)}\right)}{\Sigma\left(x^{(i)2}\right)}\right)$$

$$b = \left(\frac{t}{x} - \frac{\sum (x^{(i)} t^{(i)})}{\sum (x^{(i)^2})}\right) \div \left(\frac{N}{x} - \frac{x}{\sum (x^{(i)^2})}\right)$$

$A = \sum_{i=1}^{N} x^{(i)^2}$ $B = N$ $C = 2x$ $D = -2\sum_{i=1}^{N} x^{(i)} t^{(i)}$

$E = -2t$ $F = \sum_{i=1}^{N} t^{(i)^2}$

$$b = \left(\frac{E}{C} - \frac{D}{2A}\right) \div \left(\frac{C}{2A} - \frac{2B}{C}\right)$$

Subbing into expres.

$$b = \left(\frac{-t}{x} + \frac{\sum x^{(i)} t^{(i)}}{\sum x^{(i)^2}}\right) \div \left(\frac{x}{\sum x^{(i)^2}} - \frac{N}{x}\right)$$

We can see that solving for b gives us the same result as in did in Problem 1.

$\therefore$ The $\vec{w}^*$ that satisfies

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

is the $\vec{w}^*$ that minimizes our least squares loss.

P2.4

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

$$\rightarrow \cancel{2}X^T X\vec{w} = \cancel{2}X^T\vec{t}$$

By assuming invertibility there exists an inverse of $X^T X$ s.t.

$$(X^T X)^{-1}(X^T X) = I_d$$

$$\Rightarrow \underbrace{(X^T X)^{-1}(X^T X)}_{I_d}\vec{w} = (X^T X)^{-1} X^T \vec{t}$$

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

Problem 3

$$D = \left\{ \left(\left(x_j^{(i)} \right)_{j \in 1..d}, t^{(i)} \right) \right\}_{i \in 1 \dots N}$$

P 3.1

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

$$\vec{x}^{(i)} = \begin{bmatrix} x_1^{(i)} \\ \vdots \\ x_d^{(i)} \end{bmatrix} \qquad \vec{\omega} = \begin{bmatrix} \omega_0 \\ \vdots \\ \omega_d \end{bmatrix}$$

$$= \sum_{i=1}^{N} \begin{bmatrix} x_1^{(i)2} & \dots & x_1^{(i)} x_d^{(i)} \\ \vdots & \ddots & \vdots \\ x_d^{(i)} x_1^{(i)} & \dots & x_d^{(i)2} \end{bmatrix}$$

P 3.2

$$\vec{b} = \sum_{i=1}^{N} t^{(i)} \vec{x}^{(i)}$$

Prove $\nabla \mathcal{E}(\vec{\omega}, D) = \frac{1}{N} \left(A\vec{\omega} - \vec{b} \right) + \lambda \vec{\omega}$

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

$$g_\omega(\vec{x}) = \vec{x}\vec{\omega}$$

$$\Rightarrow \mathcal{E}(\vec{\omega}, D) = \frac{1}{2N} \sum_{i \in 1,N} \left(\vec{x}^{(i)} \vec{\omega} - t^{(i)} \right)^2 + \frac{\lambda}{2} \|\vec{\omega}\|_2^2$$

Now find $\nabla \mathcal{E}(\vec{\omega}, D)$

We can find the individual and gradients individually.

$$\Rightarrow \nabla \frac{\lambda}{2} \|\vec{\omega}\|_2^2 \Rightarrow \frac{\lambda}{2} \nabla \left(\sum_{k=0}^{d} \omega_k^2 \right)$$

We have that each component becomes

$$\omega_j^2 \longrightarrow 2\omega_j$$

$$\therefore \frac{\lambda}{2} \nabla \|\vec{\omega}\|_2^2 = \frac{\lambda}{2} 2 \vec{\omega}$$

$$= \lambda \vec{\omega}$$

For the first component: $\frac{1}{2N} \sum_{i \in 1,N} \left(\vec{x}^{(i)} \vec{\omega} - t^{(i)} \right)^2$

$$\nabla \left(\frac{1}{2N} \sum_{i \in 1,N} \left(\vec{x}^{(i)} \vec{\omega} - t^{(i)} \right)^2 \right) = \frac{1}{2N} \sum_{i \in 1,N} \nabla \left(\vec{x}^{(i)} \vec{\omega} - t^{(i)} \right)^2$$

$$= \frac{1}{2N} \sum_{i \in 1,N} \left(\vec{x}^{(i)} \vec{\omega} - t^{(i)} \right) \cdot 2 \cdot \left(\vec{x}^{(i)} \right)$$

$$= \frac{1}{N} \sum_{i \in 1,N} \left(\left(\vec{x}^{(i)} \vec{\omega} - t^{(i)} \right) \left(\vec{x}^{(i)} \right) \right)$$

$$= \frac{1}{N} \sum_{i \in 1,N} \left(\vec{x}^{(i)} \vec{x}^{(i)T} \vec{\omega} - t^{(i)} \vec{x}^{(i)} \right)$$

$$= \frac{1}{N} \sum_{i \in 1,N} \left(A \vec{\omega} - \vec{b} \right)$$

$$\therefore \nabla \mathcal{E}(\vec{\omega}, D) = \frac{1}{N} \left(A \vec{\omega} - \vec{b} \right) + \lambda \vec{\omega}$$

P3.3

$$\vec{w}^* = \arg\min_{\vec{w}} \mathcal{E}(\vec{w}, D)$$

To find $\vec{w}$ which minimizes the expression we simply set the equation from 3.2 to 0 and solve for $\vec{w}$:

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

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

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

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

This equation must be satisfied by $\vec{w}^*$

P3.4

Prove that the eigenvalues of A are all non-negative.

From 3.1

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

$$= \sum_{i=1}^{N} \begin{bmatrix} {x_1^{(i)}}^2 & \cdots & x_1^{(i)} x_d^{(i)} \\ \vdots & \ddots & \vdots \\ x_d^{(i)} x_1^{(i)} & \cdots & {x_d^{(i)}}^2 \end{bmatrix}$$

Check that A is a Positive Semi-definite

Must be

1) Symetrical.

Matrix is symmetrical as it is composed of a vector times it's transpose.

2) $v^T A v \geq 0 \quad \forall\, v \in \mathbb{R}^n$

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

$$v^T A = \begin{bmatrix} v_1 & \cdots & v_d \end{bmatrix} \underbrace{\begin{bmatrix} x_1^{(i)^2} & \cdots & x_1^{(i)} x_d^{(i)} \\ \vdots & \ddots & \vdots \\ x_d^{(i)} x_1^{(i)} & \cdots & x_d^{(i)^2} \end{bmatrix}}_{\text{All sums over } i}$$

$$(1 \times d)(d \times d) = (1 \times d)$$

To simplify notation we'll denote

$$\sum_{i=1}^{d} x_k^{(i)} x_j^{(i)} = x_k x_j \quad \text{where } k \,\&\, j \text{ are indices of } x^{(i)}$$

$$(v^T A)^T = \begin{bmatrix} v_1 x_1^2 + v_2 x_1 x_2 + \cdots + v_d x_1 x_d \\ v_1 x_2 x_1 + \cdots + \\ \vdots \\ v_1 x_d x_1 + \cdots + v_d x_d x_d \end{bmatrix}$$

Now: $v^T A v$:

$$\begin{bmatrix} v_1 x_1^2 + v_2 x_1 x_2 + \cdots + v_d x_1 x_d \\ v_1 x_2 x_1 + \cdots + \\ \vdots \\ v_1 x_d x_1 + \cdots + v_d x_d x_d \end{bmatrix}^T \begin{bmatrix} v_1 \\ \vdots \\ v_d \end{bmatrix}$$

From this we'll get a scalar value:

$$v_1(v_1 x_1^2 + v_2 x_1 x_2 + \dots + v_d x_1 x_d) +$$
$$v_2(v_1 x_2 x_1 + \dots + v_d x_2 x_d) + \qquad \geq 0?$$
$$\vdots \quad +$$
$$v_d(v_1 x_d x_1 + \dots + v_d x_d^2)$$

because all our x components are positive this will be ≥ 0.

$\therefore$ A is a positive semi-definite matrix

$\therefore$ A has all non-negative eigen values.

P3.5

$(A + \lambda NI)$ remains symetric because we only adjust the main diagonal.

Proving strictly positive eigenvalues:

where λ is the eigenvalue of $A + \lambda NI$

$$\langle \lambda v, v \rangle = \langle (A + \lambda NI) v, v \rangle \geq 0$$

$$\rightarrow \quad \lambda \langle v, v \rangle = \langle (A + \lambda NI) v, v \rangle$$

$$= \langle Av, v \rangle + \langle \lambda NIv, v \rangle$$

$$= \langle Av, v \rangle + \lambda N \langle v, v \rangle$$

We know $\langle Av, v \rangle \geq 0$ and $\lambda N \langle v, v \rangle > 0$
unless trivial solution

$$\therefore \langle Av, v \rangle + \lambda N \langle v, v \rangle > 0$$

$\therefore$ eigenvalues are all positive.

P3.6

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

$$\Longrightarrow \nabla \mathcal{E}(\vec{\omega}, D) = \frac{1}{N}(A\vec{\omega} - \vec{b}) + \lambda \vec{\omega}$$

Continuing from P3.3:

We had: $(A + \lambda N I_d)\vec{\omega}^* = \vec{b}$

And due to the invertibility of $A + \lambda N I_d$ we have

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

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