ECE421 Problem Set 2

Micol Altomare

September 27, 2023

1 Linear Regression

Given data:

$x^{(i)})$	$t^{(i)})$
1	6
2	4
3	2
4	1
5	3
6	6
7	10

Table 1: Given data

1.1

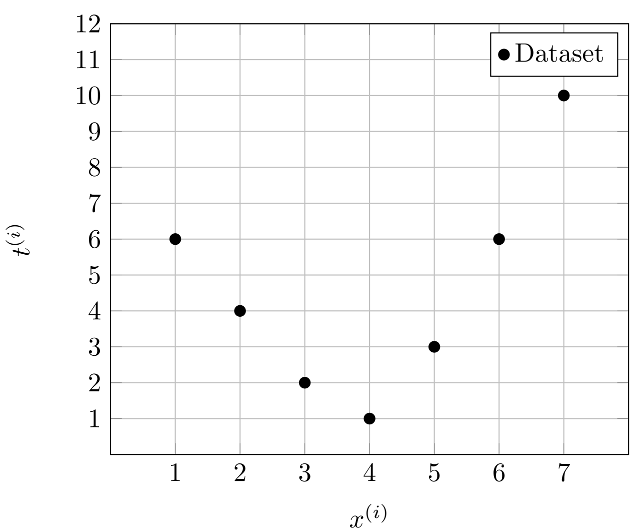

Figure 1: Scatter plot of $x^{(i)}$ vs. $t^{(i)}$

1.2

$$
\begin{aligned}
\mathcal{E}(w,b) &= \frac{1}{2N}\sum_{i=1}^{N}(y^{(i)}-t^{(i)})^2\\
&= \frac{1}{2N}\sum_{i=1}^{N}(wx^{(i)}-t^{(i)})^2\\
&= \frac{1}{2N}\sum_{i=1}^{N}\left[(wx^{(i)})^2+wx^{(i)}b-wx^{(i)}t^{(i)}+wx^{(i)}b+b^2-bt^{(i)}-wx^{(i)}t^{(i)}-t^{(i)}b+(t^{(i)})^2\right]\\
&= \frac{1}{2N}\sum_{i=1}^{N}\left[(w^2x^{(i)})^2+2wx^{(i)}b-2wx^{(i)}t^{(i)}-2bt^{(i)}+b^2+(t^{(i)})^2\right]\\
&= \boxed{\frac{1}{2N}\sum_{i=1}^{N}\left[(x^{(i)})^2w^2+1b^2+2x^{(i)}wb-2t^{(i)}x^{(i)}w-2t^{(i)}b+(t^{(i)})^2\right]}\\
\implies A_i &= (x^{(i)})^2, B_i=1, C_i=2x^{(i)}, D_i=-2t^{(i)}x^{(i)}, E_i=-2t^{(i)}, F_i=(t^{(i)})^2
\end{aligned}
$$

in the form: $\mathcal{E}(w,b) = \frac{1}{2N}\sum_{i=1}^{N}A_iw^2+B_ib^2+C_iwb+D_iw+E_ib+F_i$

1.3

The loss function is minimized when $\frac{\partial\mathcal{E}}{\partial w}=0$ and $\frac{\partial\mathcal{E}}{\partial b}=0$. Where $A=\sum_i A_i$, $B=\sum_i B_i, C=\sum_i C_i, D=\sum_i D_i, E=\sum_i E_i$:

$$
\begin{aligned}
\frac{\partial\mathcal{E}}{\partial w} &= \frac{1}{2N}\sum_{i=1}^{N}2wA_i+C_ib+D_i\\
&= 2wA+Cb+D=0\\
\implies w &= \frac{-Cb-D}{2A}\\
\frac{\partial\mathcal{E}}{\partial b} &= \frac{1}{2N}\sum_{i=1}^{N}2B_ib+C_iw+E_i\\
&= 2Bb+Cw+E=0\\
\implies b &= \frac{-Cw-E}{2B}\\
\implies w &= \frac{-2Bb-E}{C}\\
\implies & \boxed{b=\frac{2AE-CD}{C^2-4AB}; w=\frac{2BD-CE}{C^2-4AB}}
\end{aligned}
$$

1.4

By plugging in numerical values from the dataset D (Table 1) and using the results in **1.2** and **1.3**, the values are found to be approximately: $b=2.1429$ and $w=0.6071$.

1.5

Using Excel's linear regression tool, it is found that $b=2.1429$ and $w=0.6071$.

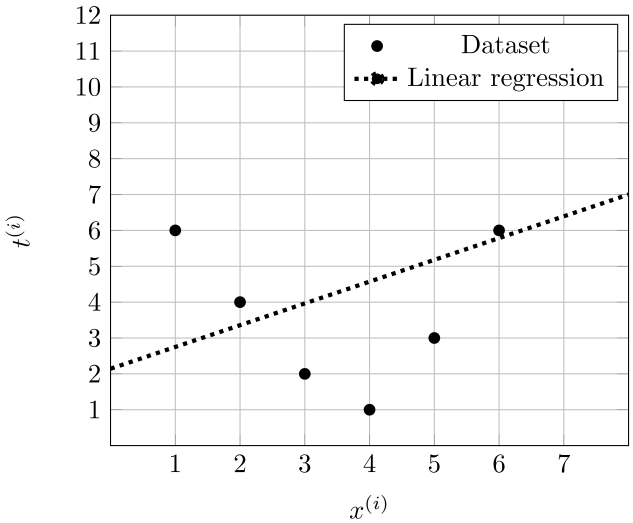

Figure 2: Scatter plot of $x^{(i)}$ vs. $t^{(i)}$

2 Least Squares

2.1

$$
\begin{aligned}
g_w(\vec{x}) &= \vec{x}\vec{w} \\
&= \begin{bmatrix} x^{(i)} & 1 \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} \\
&= wx + (1)b \\
&= g_{w,b}(x) \\
\implies & \boxed{\vec{w} = \begin{bmatrix} w \\ b \end{bmatrix}}
\end{aligned}
$$

2.2

$$
\begin{aligned}
X\vec{w} - \vec{t} &= \begin{bmatrix} x^{(1)} & 1 \\ x^{(2)} & 1 \\ \vdots & \vdots \\ x^{(N)} & 1 \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} - \begin{bmatrix} t^{(1)} \\ t^{(2)} \\ \vdots \\ t^{(N)} \end{bmatrix} \\
&= \sum_{i=1}^{N} \begin{bmatrix} x^{(i)} & 1 \end{bmatrix} \vec{w} - t^{(i)} \\
&= \sum_{i=1}^{N} x^{(i)} w + b - t^{(i)} \\
\implies ||X\vec{w} - \vec{t}||^2 &= \sum_{i=1}^{N} (x^{(i)} w + b - t^{(i)})^2 \\
\nabla_{\vec{w}} ||X\vec{w} - \vec{t}||^2 &= \begin{bmatrix} \frac{\partial ||X\vec{w} - \vec{t}||^2}{\partial w} \\ \\ \frac{\partial ||X\vec{w} - \vec{t}||^2}{\partial b} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^{N} \frac{\partial}{\partial w} (x^{(i)} w + b - t^{(i)})^2 \\ \\ \sum_{i=1}^{N} \frac{\partial}{\partial b} (x^{(i)} w + b - t^{(i)})^2 \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^{N} 2(x^{(i)} w + b - t^{(i)}) \cdot (x^{(i)}) \\ \sum_{i=1}^{N} 2(x^{(i)} w + b - t^{(i)}) \end{bmatrix} \\
&= 2 \begin{bmatrix} x^{(i)} \\ 1 \end{bmatrix} (X\vec{w} - \vec{t}) \\
&= \boxed{2X^T (X\vec{w} - \vec{t})}
\end{aligned}
$$

2.3

Setting the derived loss function to zero,

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

Thus $\vec{w^*}$ must satisfy:

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

2.4

Assuming that $X^T X$ is invertible,

$$
\begin{aligned}
0 &= 2X^T X\vec{w^*} - 2X^T \vec{t} \\
2X^T \vec{t} &= 2X^T X\vec{w^*} \\
X^T \vec{t} &= X^T X\vec{w^*} \\
\vec{w^*} &= (X^T X)^{-1} X^T \vec{t}
\end{aligned}
$$

3 Regularized Linear Regression Model

3.1

$$
\begin{aligned}
A &= \sum_{i=1}^{N} \vec{x}^{(i)} \vec{x}^{(i)^T} \\
&= \sum_{i=1}^{N} \begin{bmatrix} x_1^{(i)} \\ \vdots \\ x_d^{(i)} \end{bmatrix} \begin{bmatrix} x_1^{(i)} & \cdots & x_d^{(i)} \end{bmatrix} \\
&= \sum_{i=1}^{N} \begin{bmatrix} x_1^{(i)} x_1^{(i)} & \cdots & x_1^{(i)} x_d^{(i)} \\ \vdots & \ddots & \vdots \\ x_d^{(i)} x_1^{(i)} & \cdots & x_d^{(i)} x_d^{(i)} \end{bmatrix} \\
&= \begin{bmatrix} x_1^{(1)} x_1^{(1)} & \cdots & x_1^{(1)} x_d^{(1)} \\ \vdots & \ddots & \vdots \\ x_d^{(1)} x_1^{(1)} & \cdots & x_d^{(1)} x_d^{(1)} \end{bmatrix} + \begin{bmatrix} x_1^{(2)} x_1^{(2)} & \cdots & x_1^{(2)} x_d^{(2)} \\ \vdots & \ddots & \vdots \\ x_d^{(2)} x_1^{(2)} & \cdots & x_d^{(2)} x_d^{(2)} \end{bmatrix} + \cdots + \begin{bmatrix} x_1^{(N)} x_1^{(N)} & \cdots & x_1^{(N)} x_d^{(N)} \\ \vdots & \ddots & \vdots \\ x_d^{(N)} x_1^{(N)} & \cdots & x_d^{(N)} x_d^{(N)} \end{bmatrix} \\
\implies A &= \begin{bmatrix} \sum_{i=1}^{N} x_1^{(i)} x_1^{(i)} & \cdots & \sum_{i=1}^{N} x_1^{(i)} x_d^{(i)} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} x_d^{(i)} x_1^{(i)} & \cdots & \sum_{i=1}^{N} x_d^{(i)} x_d^{(i)} \end{bmatrix}
\end{aligned}
$$

3.2

Given:

$$
\begin{aligned}
\mathcal{E}(\vec{w}, D) &= \frac{1}{2N} \sum_{i=1}^{N} (g_{\vec{w}}(\vec{x}^{(i)}) - t^{(i)})^2 + \frac{\lambda}{2} ||\vec{w}||_2^2 \\
g_{\vec{w}} &= x^{\vec{(i)}^T} \vec{w} \\
A &= \sum_{i=1}^{N} \vec{x}^{(i)} \vec{x}^{(i)^T} \\
\vec{b} &= \sum_{i=1}^{N} t^{(i)} \vec{x}^{(i)}, \\
\nabla \mathcal{E}(\vec{w}, D) &= \nabla \left(\frac{1}{2N} \sum_{i=1}^{N} g_{\vec{w}}(\vec{x}^{(i)}) - t^{(i)})^2 + \frac{\lambda}{2} ||\vec{w}||_2^2 \right) \\
&= \nabla \left(\frac{1}{2N} \sum_{i=1}^{N} g_{\vec{w}}(\vec{x}^{(i)}) - t^{(i)})^2 \right) + \nabla \left(\frac{\lambda}{2} ||\vec{w}||_2^2 \right) \\
&= \frac{1}{2N} \sum_{i=1}^{N} \nabla \left((x^{\vec{(i)}^T} \vec{w} - t^{(i)})^2 \right) + \frac{\lambda}{2} 2w \\
&= \frac{1}{2N} \sum_{i=1}^{N} \nabla \left(((x^{\vec{(i)}^T} \vec{w})^2 - 2x^{\vec{(i)}^T} \vec{w} t^{(i)} + (t^{(i)})^2)^2 \right) + \lambda w \\
&= \frac{1}{N} (\sum_{i=1}^{N} \vec{x}^{(i)} \vec{x}^{(i)^T} \vec{w} - \sum_{i=1}^{N} t^{(i)} \vec{x}^{(i)}) + \lambda w \\
&= \frac{1}{N} (A\vec{w} - \vec{b}) + \lambda w
\end{aligned}
$$

3.3

Setting $\nabla\mathcal{E}(\vec{w}, D)$ to zero and using $\vec{w}^*$ which minimizes the loss,

$$\begin{aligned}
\nabla\mathcal{E}(\vec{w}, D) &= \frac{1}{N}(A\vec{w}^* - \vec{b}) + \lambda\vec{w}^* = 0 \\
0 &= \frac{1}{N}A\vec{w}^* - \frac{1}{N}\vec{b} + \lambda\vec{w}^* \\
&= A\vec{w} - \vec{b} + \lambda N I\vec{w}^* \\
\implies \vec{b} &= (A + \lambda N I)\vec{w}^*
\end{aligned}$$

3.4

We will prove that all eigenvalues of A are non-negative. If A is positive semi-definite, then all its eigenvalues will be non-negative (ECE421 tutorial #1 notes).

$$\vec{v}^T A\vec{v} \geq 0 \forall \vec{v} \epsilon \mathbb{R}^n \implies A \text{ is positive semi-definite.}$$

$$\text{Recall that: } A = \sum_{i=1}^{N} \vec{x}^{(i)}\vec{x}^{(i)^T}.$$

$$\begin{aligned}
\vec{v}^T A\vec{v} &= \vec{v}^T \left(\sum_{i=1}^{N} \vec{x}^{(i)}\vec{x}^{(i)^T}\right)\vec{v} \\
&= \sum_{i=1}^{N} (\vec{v}^T\vec{x}^{(i)})(\vec{x}^{(i)^T}\vec{v}) \\
&= \sum_{i=1}^{N} (\vec{v}^T\vec{x}^{(i)})^2 \geq 0 \qquad \implies \text{A's eigenvalues are non-negative.}
\end{aligned}$$

3.5

Let $A\vec{v} = \alpha\vec{v}$ where α is a non-negative eigenvalue of A (as proved in **3.4**) and $\vec{v}$ be an eigenvector of A.

$$\begin{aligned}
(A + \lambda N I_d)\vec{v} &= A\vec{v} + \lambda N\vec{v} \\
&= \alpha\vec{v} + \lambda N\vec{v} \\
&= (\alpha + \lambda N)\vec{v}
\end{aligned}$$

$\alpha \geq 0$ from **3.4** and $\lambda N > 0$. Thus, none of the eigenvalues $(\alpha + \lambda N)$ of the matrix $(A + \lambda N I_d)$ are zero, so the matrix $(A + \lambda N I_d)$ is invertible.

3.6

Using the result of **3.3**,

$$\begin{aligned}
(A + \lambda N I_d)\vec{w}^* &= \vec{b} \\
\implies \vec{w}^* &= (A + \lambda N I_d)^{-1}\vec{b}
\end{aligned}$$