ECE421 Assignment

Problem 1)

1) See excel sheet for scatter plot of dataset (attached at end of pdf file).

2) Let $\{y^{(i)}\}_{i=1,\dots,7} = (y^{(1)}, y^{(2)}, \dots y^{(7)})$ be the models predictions for inputs $\{x^{(i)}\}_{i=1,\dots,7} = (1,2,3,4,5,6,7)$, so that $y^{(i)} = g_{w,b}(x^{(i)}) = wx^{(i)}+b$. Then, we can write our loss as:

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

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

$$\boxed{\mathcal{E}(w,b) = \frac{1}{2N}\sum_{i=1}^{N}\left(w^2(x^{(i)})^2+b^2+(t^{(i)})^2+2bwx^{(i)}-2wx^{(i)}t^{(i)}-2bt^{(i)}\right)}$$

Matching coefficients with the expression $\frac{1}{2N}\sum_{i=1}^{N} A_iw^2+B_ib^2+C_iwb+D_iw+E_ib+F_i$ to verify our result above gives:

$A_iw^2 = (x^{(i)})^2w^2 \Rightarrow A_i = (x^{(i)})^2$

$B_ib^2 = b^2 \Rightarrow B_i = 1$

$C_iwb = 2x^{(i)}wb \Rightarrow C_i = 2x^{(i)}$

$D_iw = -2x^{(i)}t^{(i)}w \Rightarrow D_i = -2x^{(i)}t^{(i)}$

$E_ib = -2t^{(i)}b \Rightarrow E_i = -2t^{(i)}$

$F_i = (t^{(i)})^2 \Rightarrow F_i = (t^{(i)})^2$

3) $\mathcal{E}(w,b)$ is a continuous function of w and b. We seek to find w,b that minimize $\mathcal{E}(w,b)$. Now, it was covered in Video 6 that the minimum of smooth function occurs at a critical point; i.e where derivative is 0. Since we are dealing with a function of 'two' variables; we can find the critical point by setting partial derivatives to 0.

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

$$= \frac{1}{2N}\left(w^2\sum A_i + b^2\sum B_i + wb\sum C_i + w\sum D_i + b\sum E_i + \sum F_i\right)$$

$$= \frac{1}{2N}\left(Aw^2+Bb^2+Cwb+Dw+Eb+\sum F_i\right)$$

where $A=\sum_{i=1}^{N}A_i$, $B=\sum_{i=1}^{N}B_i$, $C=\sum_{i=1}^{N}C_i$, $D=\sum_{i=1}^{N}D_i$ and $E=\sum_{i=1}^{N}E_i$

$$\frac{\partial\mathcal{E}(w,b)}{\partial w}=0 \Rightarrow \frac{\partial}{\partial w}\left[\frac{1}{2N}\left(Aw^2+Bb^2+Cwb+Dw+Eb+\sum F_i\right)\right]=0 \qquad \frac{\partial\mathcal{E}(w,b)}{\partial b}=0 \Rightarrow \frac{\partial}{\partial b}\left[\frac{1}{2N}\left(Aw^2+Bb^2+Cwb+Dw+Eb+\sum F_i\right)\right]=0$$

$$\frac{1}{2N}(2Aw+Cb+D)=0 \quad (1) \qquad \frac{1}{2N}(2Bb+Cw+E)=0 \quad (2)$$

Solving system of equations defined by (1) and (2):

① × 2NC : $2ACw+C^2b+CD=0$

② × 4AN : $4ABb+2ACw+2AE=0$

$C^2b+CD-4ABb-2AE=0$

$b(C^2-4AB)=2AE-CD$

$$\boxed{b=\frac{2AE-CD}{C^2-4AB}}$$

$$w=\frac{-Cb-D}{2A}=\frac{-C\left(\frac{2AE-CD}{C^2-4AB}\right)-D}{2A}=\left(\frac{C^2D-2ACE}{C^2-4AB}-\frac{C^2D-4ABD}{C^2-4AB}\right)/2A$$

$$=\frac{4ABD-2ACE}{2A(C^2-4AB)}=\frac{2BD-CE}{C^2-4AB}$$

$$\Rightarrow\boxed{w=\frac{2BD-CE}{C^2-4AB}}$$

4) For this part, we will simply compute the values of b and w found in part (3), using the dataset given.

$A = \sum_{i=1}^{N} A_i = \sum (x^{(i)})^2 = \sum i^2 = 1^2 + 2^2 \cdots + 7^2 = \frac{7(8)(15)}{6} = 140$

$B = \sum B_i = \sum 1 = 7$

$C = \sum_i C_i = \sum_i 2x^{(i)} = 2\sum_{i=1}^{N} x^{(i)} = 2\,\frac{7(8)}{2} = 56$

$D = \sum D_i = \sum -2x^{(i)}t^{(i)} = -2\sum x^{(i)}t^{(i)} = -2\left((6+8)+(2)(3)+(1)(4)+(3)(5)+(6)(6)+(10)(7)\right) = -290$

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

Now, we just plug in to find w, b that minimize loss:

$$b = \frac{2AE - CD}{C^2 - 4AB} = \frac{2(140)(-64) - (56)(-290)}{56^2 - 4(140)(7)} = 15/7 \Rightarrow \boxed{b = 15/7}$$

$$w = \frac{2(7)(-290) - (56)(-64)}{56^2 - 4(140)(7)} = 17/28 \Rightarrow \boxed{w = 17/28}$$

Problem 2

1) We have:

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

$$g_{w,b}(x) = \begin{bmatrix} x & 1 \end{bmatrix}\begin{bmatrix} w \\ b \end{bmatrix}$$

$$g_{w,b}(x) = \vec{x}\,\vec{w}$$

Since the right-hand side depends only on $\vec{w}$ and $\vec{x}$, we have $g_{w,b}(x) = g_w(\vec{x}) = \vec{x}\,\vec{w}$, as desired. For a specific input-output pair $(x^{(i)}, g_w(x^{(i)}))$, we have $g_{w,b}(x^{(i)}) = wx^{(i)} + b = \begin{bmatrix} x^{(i)} & 1 \end{bmatrix}\begin{bmatrix} w \\ b \end{bmatrix} = \vec{x}\,\vec{w} = g_w(\vec{x}^{(i)})$

2) We must derive analytically $\nabla_{\vec{w}} \|X\vec{w} - \vec{t}\|^2$

Let $\vec{y} = X\vec{w} - \vec{t}$

Then $y_i = \vec{x}^{(i)}\vec{w} - t^{(i)}$

This gives:

$$\|X\vec{w} - \vec{t}\|^2 = \|\vec{y}\|^2 = \sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} (\vec{x}^{(i)}\vec{w} - t^{(i)})^2 = \sum_{i=1}^{N} \left(\begin{bmatrix} x^{(i)} & 1 \end{bmatrix}\begin{bmatrix} w \\ b \end{bmatrix} - t^{(i)}\right)^2 = \sum_{i=1}^{N} (x^{(i)}w + b - t^{(i)})^2$$

Now, we will compute $\frac{\partial \|\vec{y}\|^2}{\partial w}$ and $\frac{\partial \|\vec{y}\|^2}{\partial b}$

$$\frac{\partial \|\vec{y}\|^2}{\partial w} = \frac{\partial}{\partial w}\left(\sum_{i=1}^{N} (x^{(i)}w + b - t^{(i)})^2\right) = \sum_{i=1}^{N} \frac{\partial}{\partial w}(x^{(i)}w + b - t^{(i)})^2 = \sum_{i=1}^{N} 2(x^{(i)}w + b - t^{(i)})(x^{(i)})$$

$$= \sum_{i=1}^{N} 2(\vec{x}^{(i)}\vec{w} - t^{(i)})(x^{(i)}) = 2(X\vec{w} - \vec{t}) \cdot \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(N)} \end{bmatrix}$$

$$\frac{\partial \|\vec{y}\|^2}{\partial b} = \frac{\partial}{\partial b}\left(\sum_{i=1}^{N} (x^{(i)}w + b - t^{(i)})^2\right) = \sum_{i=1}^{N} \frac{\partial}{\partial b}(x^{(i)}w + b - t^{(i)})^2 = \sum_{i=1}^{N} 2(x^{(i)}w + b - t^{(i)})(1)$$

$$= \sum_{i=1}^{N} 2(\vec{x}^{(i)}\vec{w} - t^{(i)})(1) = 2(X\vec{w} - \vec{t}) \cdot \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \text{ where } \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^{N \times 1}$$

It follows that:

$$\frac{\partial \|\vec{y}\|^2}{\partial w} = 2\begin{bmatrix} x^{(1)} \\ \vdots \\ x^{(N)} \end{bmatrix}^T (X\vec{w} - \vec{t}) = 2\begin{bmatrix} x^{(1)}, & \ldots, & x^{(N)} \end{bmatrix}(X\vec{w} - \vec{t})$$

$$\frac{\partial \|\vec{y}\|^2}{\partial b} = 2\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}^T (X\vec{w} - \vec{t}) = 2\begin{bmatrix} 1 & \ldots & 1 \end{bmatrix}(X\vec{w} - \vec{t})$$

So:

$$\begin{bmatrix} \frac{\partial \|\vec{y}\|^2}{\partial w} \\ \frac{\partial \|\vec{y}\|^2}{\partial b} \end{bmatrix} = 2\begin{bmatrix} x^{(1)} & x^{(2)} & \ldots & x^{(N)} \\ 1 & 1 & \ldots & 1 \end{bmatrix}(X\vec{w} - \vec{t})$$

$$= 2X^T(X\vec{w} - \vec{t}), \text{ where we note } X^T = \begin{bmatrix} \vec{x}^{(1)} \\ \vec{x}^{(2)} \\ \vdots \\ \vec{x}^{(N)} \end{bmatrix}^T = \begin{bmatrix} (\vec{x}^{(1)})^T & (\vec{x}^{(2)})^T & \ldots & (\vec{x}^{(N)})^T \end{bmatrix} = \begin{bmatrix} x^{(1)} & x^{(2)} & \ldots & x^{(N)} \\ 1 & 1 & \ldots & 1 \end{bmatrix}$$

$$\boxed{\nabla_{\vec{w}} \|X\vec{w} - \vec{t}\|^2 = \begin{bmatrix} \frac{\partial \|\vec{y}\|^2}{\partial w} \\ \frac{\partial \|\vec{y}\|^2}{\partial b} \end{bmatrix} = 2X^T(X\vec{w} - \vec{t})}$$

3) The minimum value of smooth function occurs at a critical point.

The derivative of a function at a critical point is 0.

Since our least squares expression is a function of 2 variables (w and b), it is sufficient to set the partial derivatives to 0:

$$\frac{\partial}{\partial w} \|X\vec{w}-\vec{t}\|^2 = 0$$

$$\frac{\partial}{\partial b} \|X\vec{w}-\vec{t}\|^2 = 0$$

This is equivalent to setting:

$$\begin{bmatrix} \frac{\partial}{\partial w} \|X\vec{w}-\vec{t}\|^2 \\ \frac{\partial}{\partial b} \|X\vec{w}-\vec{t}\|^2 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

or: $\nabla_{\vec{w}} \|X\vec{w}-\vec{t}\|^2 = 0$

But we have already computed:

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

So, the models weight value $\vec{w}^*$ which minimizes the least squares loss must satisfy:

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

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

4) We simply solve equation above:

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

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

$\boxed{\vec{w}^* = (X^T X)^{-1} X^T\vec{t}}$, where $(X^T X)^{-1}$ is inverse of $X^T X$

Problem 3

1) Let $c^{(i)} = x^{(i)}(x^{(i)})^T$. By definition of the matrix multiplication $c^{(i)}_{mn} = x^{(i)}_m x^{(i)}_n$

Now, $A = \sum_{i=1}^{N} c^{(i)}$. So $A_{mn} = \sum_{i=1}^{N} c^{(i)}_{mn} = \sum_{i=1}^{N} x_m^{(i)} x_n^{(i)}$.

So, a simple analytic expression for the components of A is:

$$\boxed{A_{mn} = \sum_{i=1}^{N} \vec{x}_m^{(i)} \vec{x}_n^{(i)}}$$

2) We follow a similar procedure as in Problem 2.2 to compute the gradient:

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

$$= \frac{1}{2N} \sum_{i=1}^{N} \left(\sum_{k=1}^{d} w_k \vec{x}_k^{(i)} - t^{(i)}\right)^2 + \frac{\lambda}{2} \|\vec{w}\|_2^2$$

$$= \frac{1}{2N} \sum_{i=1}^{N} \left(\sum_{k=1}^{d} w_k \vec{x}_k^{(i)} - t^{(i)}\right)^2 + \frac{\lambda}{2} \sum_{k=1}^{d} w_k^2$$

Now, for an arbitrary w_j, $1 \le j \le d$:

$$\frac{\partial \mathcal{E}(w,b)}{\partial w_j} = \frac{1}{2N} \sum_{i=1}^{N} \frac{\partial}{\partial w_j} \left(\sum_{k=1}^{d} w_k \vec{x}_k^{(i)} - t^{(i)}\right)^2 + \frac{\lambda}{2} \sum_{k=1}^{d} \frac{\partial}{\partial w_j} w_k^2$$

$$= \frac{1}{2N} \sum_{i=1}^{N} 2\left(\sum_{k=1}^{d} w_k \vec{x}_k^{(i)} - t^{(i)}\right) \vec{x}_j^{(i)} + \frac{\lambda}{2}(2w_j)$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{k=1}^{d} w_k \vec{x}_k^{(i)} - t^{(i)}\right) \vec{x}_j^{(i)} + \lambda w_j$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left(\vec{x}_j^{(i)} \sum_{k=1}^{d} w_k \vec{x}_k^{(i)} - t^{(i)} \vec{x}_j^{(i)}\right) + \lambda w_j$$

$$= \frac{1}{N} \left(\sum_{i=1}^{N} \vec{x}_j^{(i)} \sum_{k=1}^{d} w_k \vec{x}_k^{(i)} - \sum_{i=1}^{N} t^{(i)} \vec{x}_j^{(i)}\right) + \lambda w_j$$

$$= \frac{1}{N} \left(\left(\sum_{k=1}^{d} w_k \left(\sum_{i=1}^{N} \vec{x}_j^{(i)} \vec{x}_k^{(i)}\right)\right) - \sum_{i=1}^{N} t^{(i)} \vec{x}_j^{(i)}\right) + \lambda w_j$$

From part (1) we know:

$$A_{jk} = \sum_{i=1}^{N} \vec{x}_j^{(i)} \vec{x}_k^{(i)}$$

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

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

So:

$$\frac{\partial \mathcal{E}(w,b)}{\partial w_j} = \frac{1}{N} \left(\sum_{k=1}^{d} w_k A_{jk} - b_j\right) + \lambda w_j$$

Now, clearly, $\sum_{k=1}^{d} w_k A_{jk}$ is the j^{th} row of $A\vec{w}$. So:

$$\frac{\partial \mathcal{E}(w,b)}{\partial w_j} = \frac{1}{N}\left((A\vec{w})_j - b_j\right) + \lambda w_j$$

Ardroy

Combining these partial derivatives to obtain the gradient vector gives:

$$\nabla \mathcal{E}(\vec{w}, D) = \begin{bmatrix} \frac{1}{N}((A\vec{w})_1 - \vec{b}_1) + \lambda w_1 \\ \frac{1}{N}((A\vec{w})_2 - \vec{b}_2) + \lambda w_2 \\ \vdots \\ \frac{1}{N}((A\vec{w})_d - \vec{b}_d) + \lambda w_d \end{bmatrix}$$

$$= \frac{1}{N}\begin{bmatrix} (A\vec{w})_1 - \vec{b}_1 \\ (A\vec{w})_2 - \vec{b}_2 \\ \vdots \\ (A\vec{w})_d - \vec{b}_d \end{bmatrix} + \lambda \begin{bmatrix} w_1 \\ \vdots \\ w_d \end{bmatrix}$$

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

3) The minimum value of a smooth function occurs at a critical point.

The derivative of a function at a critical point is 0

Since our least squares expression is a function of d variables $(w_1, w_2, \dots, w_d)$, it is sufficient to set the partial derivatives to 0. This is equivalent to setting the gradient to 0.

$$\nabla \mathcal{E}(\vec{w}, D) = 0$$

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

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

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

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

4) Assume for sake of contradiction there exists a negative eigenvalue of A, that is, there exists $\lambda < 0$ such that $Aq = \lambda q$ for some $\vec{q} \neq 0$ and $A = \sum_{i=1}^{N} \vec{x}^{(i)}(\vec{x}^{(i)})^T$

Now:

$$\langle Aq, q\rangle = (Aq)^T q = q^T A q = q^T \left(\sum_{i=1}^{N} \vec{x}^{(i)}(\vec{x}^{(i)})^T\right) q = \sum_{i=1}^{N} q^T \vec{x}^{(i)}(\vec{x}^{(i)})^T q = \sum_{i=1}^{N} ((\vec{x}^{(i)})^T q)^T ((\vec{x}^{(i)})^T q)$$

$$= \sum_{i=1}^{N} \|(\vec{x}^{(i)})^T q\|^2 = \sum_{i=1}^{N} ((\vec{x}^{(i)})^T q)^2, \text{ since each } (\vec{x}^{(i)})^T q \text{ is a scalar}$$

$\Rightarrow \langle Aq, q\rangle \geq 0$, since sum of squares is always positive or zero.

We also have:

$\langle Aq, q\rangle = \langle \lambda q, q\rangle$, where $\lambda < 0$ (by our assumption)

$= \lambda \langle q, q\rangle$

$= \lambda \|q\|^2$

This gives:

$$\lambda = \frac{\langle Aq, q\rangle}{\|q\|^2}$$

We know $\langle Aq, q\rangle \geq 0$. Also $\|q\|^2 > 0$, since (1) norm squared must always be positive or zero and (2) $\|x\|^2 = 0$ if and only if $x = 0$ in general, but here we must have $q \neq 0$. So, it is a strict "greater than" relationship. This gives:

$$\lambda = \frac{\langle Aq, q\rangle}{\|q\|^2} \geq 0$$

But $\lambda < 0$ is a contradiction. So, there is no negative eigenvalues. Obviously then, all eigenvalues are non-negative.

5) We follow a similar procedure as (4). Let λ be any eigenvalue of $A + \lambda N I_d$. We then have:

$(A + \lambda N I_d) q = \lambda q$, for some $q \neq 0$

$\langle (A + \lambda N I_d) q, q\rangle = \langle Aq + \lambda N q, q\rangle$

$= \langle Aq, q\rangle + \langle \lambda N q, q\rangle$

$= \langle Aq, q\rangle + \lambda N \langle q, q\rangle$

$= \langle Aq, q\rangle + \lambda N \|q\|^2$

Before, in part 4), we established that $\langle Aq, q\rangle \geq 0$. Now we have a new term, $\lambda N \|q\|^2$, which is strictly greater than 0, since $\|q\|^2 \geq 0$ for any $q \in \mathbb{R}^d$. Given q is an eigenvector, $q \neq 0$. Now, since $\|q\| = 0$ if and only if $q = 0$, we must strictly have $\|q\|^2 > 0$. This changes our inequality from part 4 by making it stricter:

$$\langle (A + \lambda N I_d) q, q\rangle = \langle Aq, q\rangle + \lambda N \|q\|^2 > 0$$

Again, we have:

$\langle (A + \lambda N I_d) q, q\rangle = \langle \lambda q, q\rangle = \lambda \langle q, q\rangle = \lambda \|q\|^2$

So:

$$\lambda = \frac{\langle (A + \lambda N I_d) q, q\rangle}{\|q\|^2}$$

Since both numerator and denominator are strictly greater than 0:

$$\boxed{\lambda > 0}$$

Thus, all eigenvalues are strictly positive, so there can be no eigenvalues that are 0.

6) The equation stated in (3) is:

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

Using the invertibility of matrix $A + \lambda N I_d$:

$$\vec{w}^* = (A + \lambda N I_d)^{-1}\vec{b}$$, where $(A + \lambda N I_d)^{-1}$ is inverse of $A + \lambda N I_d$

This is the analytic solution for $\vec{w}^*$

	A	B
1	1	6
2	2	4
3	3	2
4	4	1
5	5	3
6	6	6
7	7	10