SML Assignment 1

Dongju Ma
A1942340

June 29, 2025
Trimester 2

Question 1a: Closed Form of Linear Regression: derivation

$$\nabla_W(J(W)) = \frac{\partial}{\partial W}\frac{1}{2m}||\hat{Y} - Y||_2^2 = \frac{\partial}{\partial W}\frac{1}{2m}(XW - Y)^T(XW - Y).$$

Denote that $X \in \mathbb{R}^{m\times n}, Y \in \mathbb{R}^{m\times 1}$ and $W \in \mathbb{R}^{n\times 1}$,
hence $(W^T X^T Y)^T = Y^T XW = W^T X^T Y$ due to each term is scalar.
Using the identify $(AB)^T = B^T A^T$ and expanding $J(W)$

$$\begin{aligned} J(W) &= \frac{1}{2m}(XW - Y)^T(XW - Y) \\ &= \frac{1}{2m}(W^T X^T - Y^T)(XW - Y) \\ &= \frac{1}{2m}\left(W^T X^T XW - 2W^T X^T Y + Y^T Y\right). \end{aligned}$$

Apply partial derivatives to $J(W)$

$$\begin{aligned} \nabla_W(J(W)) &= \frac{\partial}{\partial W}\frac{1}{2m}(W^T X^T XW - 2W^T X^T Y + Y^T Y) \\ &= \frac{1}{2m}\left(\frac{\partial}{\partial W}W^T X^T XW - \frac{\partial}{\partial W}2W^T X^T Y\right) \\ &= \frac{1}{2m}(2X^T XW - 2X^T Y) \\ &= \frac{1}{m}(X^T XW - X^T Y) \end{aligned}$$

To minimize $J(W)$, we should compare $\nabla_W(J(W))$ to 0, therefore

$$\nabla_W(J(W)) = 0 \quad \Rightarrow \quad X^T XW = X^T Y$$

multiply both sides of the equation by $(X^T X)^{-1}$, thus

$$W = (X^T X)^{-1} X^T Y$$

Question 1b: Matrix Alignment Verification

Given that $X \in \mathbb{R}^{5\times 3}$ and $Y \in \mathbb{R}^{5\times 1}$, we can compute the dimension of W by the following steps, Since

$$W = (X^T X)^{-1} X^T Y,$$

$$\begin{aligned} X^T \in \mathbb{R}^{3\times 5} \Rightarrow X^T X \in \mathbb{R}^{3\times 3} &\Rightarrow (X^T X)^{-1} \in \mathbb{R}^{3\times 3} \\ &\Rightarrow (X^T X)^{-1} X^T \in \mathbb{R}^{3\times 5} \\ &\Rightarrow (X^T X)^{-1} X^T Y = W \in \mathbb{R}^{3\times 1} \end{aligned}$$

Thus $\hat{Y}$ has the dimension of that

$$\hat{Y} = XW \Rightarrow \hat{Y} \in \mathbb{R}^{5\times 1}$$

Which makes $\hat{Y}$ has the same dimension with Y, the alignment is correct.

Question 4a: Machine Learning Matrix Operations

- $\langle \mathbf{w}, \mathbf{x}_i \rangle$
  Using the definiation of inner products, $\langle \mathbf{w}, \mathbf{x}_i \rangle = \mathbf{x}_i^\top \cdot \mathbf{w} = \sum_j^d \mathbf{x}_{ij} \mathbf{w}_j$, hence that,
  $$\sum_i \langle \mathbf{w}, \mathbf{x}_i \rangle = \sum_i \mathbf{x}_i{}^\top \mathbf{w},$$
  which is equivelent to Eq1.

- $(\sum_i \mathbf{x}_i)\mathbf{w}$
  As $\mathbf{x}_i \in \mathbb{R}^{d\times 1}$, therefore $\sum_i \mathbf{x}_i \in \mathbb{R}^{d\times 1}$.
  Thus $\sum_i \mathbf{x}_i$ cannot dot with $\mathbf{w}$ due to that $\sum_i \mathbf{x}_i$ has a dimension of d rows and 1 column which is the same with $\mathbf{w}$, hence $(\sum_i \mathbf{x}_i)\mathbf{w}$ is not equivelent to Eq1.

- $\sum_i \mathrm{Tr}(\mathbf{x}_i \mathbf{w}^\top)$
  Given that $\mathbf{x}_i \in \mathbb{R}^{d\times 1}$, $\mathbf{w} \in \mathbb{R}^{d\times 1}$, therefore $\mathbf{x}_i^\top \mathbf{w}$ is scalar, which means that
  $$\mathbf{x}_i^\top \mathbf{w} = \mathbf{w}^\top \mathbf{x}_i$$
  Using the identify of trace, $\mathrm{Tr}(\mathbf{x}_i \mathbf{w}^\top) = \mathrm{Tr}(\mathbf{w}^\top \mathbf{x}_i)$, which is equivelent to $\mathbf{w}^\top \mathbf{x}_i$. Therefore
  $$\sum_i \mathrm{Tr}(\mathbf{x}_i \mathbf{w}^\top) = \sum_i \mathrm{Tr}(\mathbf{w}^\top \mathbf{x}_i) = \sum_i \mathbf{w}^\top \mathbf{x}_i = \sum_i \mathbf{x}_i^\top \mathbf{w},$$
  which is equivelent to Eq1.

- $\mathbf{w}^T \sum_i \mathbf{x}_i$
  As $\mathbf{x}_i^\top \mathbf{w}$ is scalar, therefore
  $$\sum_i \mathbf{x}_i^\top \mathbf{w} = \sum_i \mathbf{w}^\top \mathbf{x}_i.$$
  Using the identify of sum,
  $$\sum_i \mathbf{w}^\top \mathbf{x}_i = \mathbf{w}^\top \sum_i \mathbf{x}_i,$$
  which means $\mathbf{w}^\top \sum_i \mathbf{x}_i$ is equivelent to Eq1.