Part 1: Theoretical Exercises (16 points)

We have the following data:

x	y
-1	-1
-1	1
1	2
2	3

1. We would like to fit a linear regression model to this data for the purpose of predicting future values of y from x.
   - Write the data matrix X for this regression. Make sure to include the bias term.
   - Write the pseudo inverse $X^\dagger$ of X.
   - Use $X^\dagger$ to find the vector $\theta^* \in \mathbb{R}^2$ that minimizes the sum of squares loss:

$$J(\theta) = \sum_{i=1}^{n} \left(\theta^\top (1, x^{(i)}) - y^{(i)} \right)^2$$

   - Compute the minimum loss $J(\theta^*)$.
2. Confirm that this is the minimum loss using calculus.
   - Exprss the loss in the form $J(\theta) = A\theta_0^2 + B\theta_1\theta_0 + C\theta_1^2 + D\theta_0 + E\theta_1 + F$, for some A, B, C, D, E, and F that depend on x and y.
   - Find an expression for the gradient $\nabla J(\theta) \in \mathbb{R}^2$ for aritrary $\theta \in \mathbb{R}^2$.
   - Show that $\nabla J(\theta^*) = 0$.
3. Consider the prediction of y at a test point $x = 1.5$.
   - What is the predicted value of y at this point based on linear regression with θ^*?
   - What is the predicted value of y at this point based on K-NN with $K = 2$?

Home Work #1 – Romi & Ophere

Data:

x	y
−1	−1
−1	1
1	2
2	3

← future values of y from x

1. A. Data Matrix - X

$$X = \begin{array}{cc} x_0 & x \\ \end{array} \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \qquad Y = \begin{bmatrix} -1 \\ 1 \\ 2 \\ 3 \end{bmatrix}$$

B. $X^+ = (X^T X)^{-1} \cdot X^T \rightarrow \theta^* \in \mathbb{R}^2 = X^+ y$

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 2 \end{bmatrix}$$

$$X^T \cdot X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 2 \end{bmatrix} \times \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1\cdot1 + 1\cdot1 + 1\cdot1 + 1\cdot1 & (-1)\cdot1 + (-1) + 1 + 1\cdot2 \\ 1\cdot(-1) + 1\cdot(-1) + 1\cdot1 + 1\cdot2 & (-1)(-1) + (-1)(-1) + 1 + 4 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 1 & 7 \end{bmatrix}$$

inverse

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$(X^T \cdot X)^{-1} = \frac{1}{4\cdot7 - 1\cdot1} \cdot \begin{bmatrix} 7 & -1 \\ -1 & 4 \end{bmatrix} = \frac{1}{27} \begin{bmatrix} 7 & -1 \\ -1 & 4 \end{bmatrix}$$

$$X^+ = (X^T \cdot X)^{-1} \cdot X^T = \frac{1}{27} \begin{bmatrix} 7 & -1 \\ -1 & 4 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 2 \end{bmatrix}$$

c. $\theta^* = (X^T \cdot X)^{-1} \cdot X^T \cdot y$

$$X^T \cdot y = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 2 \end{bmatrix} \times \begin{bmatrix} -1 \\ -1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} \cancel{-1+1}+2+3 \\ 1-1+2+6 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$$

$$\theta^* = \frac{1}{27} \begin{bmatrix} 7 & -1 \\ -1 & 4 \end{bmatrix} \times \begin{bmatrix} 5 \\ 8 \end{bmatrix} = \frac{1}{27} \begin{bmatrix} 35-8 \\ -5+32 \end{bmatrix} = \frac{1}{27} \begin{bmatrix} 27 \\ 27 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\hat{y} = 1 + 1 \cdot x$$

Computing the minimus loss - $J(\theta) = \sum_{i=1}^{n} \left(\theta^\top (1, x^{(i)}) - y^{(i)} \right)^2$

$$J(\theta^*) = \sum_{i=1}^{4} \left((1,1) \cdot \begin{pmatrix} 1 \\ x^i \end{pmatrix} - y^i \right)^2 = \sum_{i=1}^{4} \left((1 + x^i) - y^i \right)^2 =$$

$$\begin{bmatrix} -1 \\ -1 \\ 2 \\ 3 \end{bmatrix}$$

$$((1+(-1)-(-1))^2 + (1+(-1)-1)^2 + (1+1-2)^2 + ((1+2)-3)^2 =$$

$$1^2 + 1^2 + 0^2 + 0^2 = 2 = J(\theta^*)$$

2. Confirm that this is the minimum loss using calculus.

- Exprss the loss in the form $J(\theta) = A\theta_0^2 + B\theta_1\theta_0 + C\theta_1^2 + D\theta_0 + E\theta_1 + F$, for some A, B, C, D, E, and F that depend on x and y.
- Find an expression for the gradient $\nabla J(\theta) \in \mathbb{R}^2$ for aritrary $\theta \in \mathbb{R}^2$.
- Show that $\nabla J(\theta^*) = 0$.

plug in

$$J(\theta) = \sum_{i=1}^{4} \left((\theta_0, \theta_1) \begin{pmatrix} n \\ x_i \end{pmatrix} - y_i \right)^2 = \sum_{i=1}^{4} (\theta_0 + \theta_1 x_i - y_i)^2 =$$

based on formula

$$(\theta_0 + \theta_1(-1) - (-1))^2 + (\theta_0 + \theta_1(-1) - 1)^2 + (\theta_0 + \theta_1(1) - 2)^2 + (\theta_0 + \theta_1(2) - 3)^2 =$$

$$(\theta_0 - \theta_1 + 1)^2 + (\theta_0 - \theta_1 - 1)^2 + (\theta_0 + \theta_1 - 2)^2 + (\theta_0 + 2\theta_1 - 3)^2 =$$

* $(\theta_0 - \theta_1 + 1)(\theta_0 - \theta_1 + 1) = \theta_0^2 - \theta_0\theta_1 + \theta_0 - \theta_0\theta_1 - \theta_1 + \theta_0 - \theta_1 + 1 = \theta_0^2 - 2\theta_0\theta_1 - 2\theta_0 - 2\theta_1 + 1 + \theta_1^2$

* $(\theta_0 - \theta_1 - 1)(\theta_0 - \theta_1 - 1) = \theta_0^2 - \theta_0\theta_1 - \theta_0 - \theta_1\theta_0 + \theta_1^2 + \theta_1 - \theta_0 + \theta_1 + 1 = \theta_0^2 - 2\theta_0\theta_1 + 2\theta_0 + \theta_1^2 + 2\theta_1 + 1 =$

* + * $= 2\theta_0^2 - 4\theta_0\theta_1 + 2\theta_1^2 + 2$

* $(\theta_0 + \theta_1 - 2)(\theta_0 + \theta_1 - 2) = \theta_0^2 + \theta_0\theta_1 - 2\theta_0 + \theta_0\theta_1 + \theta_1^2 - 2\theta_1 - 2\theta_0 - 2\theta_1 + 4 = \theta_0^2 + 2\theta_0\theta_1 + \theta_1^2 - 4\theta_0 - 4\theta_1 + 4$

* $(\theta_0 + 2\theta_1 - 3)(\theta_0 + 2\theta_1 - 3) = \theta_0^2 + 2\theta_0\theta_1 - 3\theta_0 + 2\theta_0\theta_1 + 4\theta_1^2 - 6\theta_1 - 3\theta_0 - 6\theta_1 + 9 = \theta_0^2 + 4\theta_0\theta_1 + 4\theta_1^2 - 6\theta_0 - 12\theta_1 + 9$

* + * $= 2\theta_0^2 + 6\theta_0\theta_1 + 5\theta_1^2 - 10\theta_0 - 16\theta_1 + 13$

Total: $2\theta_0^2 - 4\theta_0\theta_1 + 2\theta_1^2 + 2 + 2\theta_0^2 + 6\theta_0\theta_1 + 5\theta_1^2 - 10\theta_0 - 16\theta_1 + 13 = 4\theta_0^2 + 2\theta_0\theta_1 + 7\theta_1^2 - 10\theta_0 - 16\theta_1 + 15$

- Find an expression for the gradient $\nabla J(\theta) \in \mathbb{R}^2$ for aritrary $\theta \in \mathbb{R}^2$.
- Show that $\nabla J(\theta^*) = 0$.

$$\nabla J(\theta) = \left(\frac{\partial J}{\partial \theta_0}(\theta), \frac{\partial J}{\partial \theta_1}(\theta) \right)$$

$$\left.\begin{aligned} \frac{\partial J}{\partial \theta_0}(\theta) &= 8\theta_0 + 2\theta_1 - 10 \\ \frac{\partial J}{\partial \theta_1}(\theta) &= 2\theta_0 + 14\theta_1 - 16 \end{aligned}\right\} \quad \theta^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\nabla J(\theta^*) = \begin{bmatrix} 8 + 2 - 10 \\ 2 + 14 - 16 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

3. Consider the prediction of y at a test point $x = 1.5$.

- What is the predicted value of y at this point based on linear regression with θ^*?
- What is the predicted value of y at this point based on K-NN with $K = 2$?

(1) $\hat{y}$ predicted value of y $= \theta_0 + \theta_1 x = 1 + 1 \cdot x = 1 + x = 1 + 1\frac{1}{2} = 2\frac{1}{2}$

$\theta^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

(2) Based on K Nearest Neighbors –

K=2 – Choose the two nearest

$x_1 = 1 \quad x_2 = 2$

$\underbrace{y_1 = 2 \quad y_2 = 3}$

$\hat{y} = \frac{2+3}{2} = 2\frac{1}{2}$