ש.ב 1 - תשובות.

20700 2544
206860967

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 $J(\theta^*)$.

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

• $X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & 7 \end{pmatrix}$

השורות במטריצה בת״ל ולכן הפיכה.

$$(X^T X)^{-1} = \frac{1}{|X^T X|} \cdot Adj(X^T X)$$

$$\det(X^T X) = 4 \cdot 7 - 1 \cdot 1 = 27$$

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

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

$$= \begin{pmatrix} \frac{8}{27} & \frac{8}{27} & \frac{2}{9} & \frac{5}{27} \\ \frac{-5}{27} & \frac{-8}{27} & \frac{1}{9} & \frac{7}{27} \end{pmatrix}$$

$$\bullet\ \Theta^* = X^+ \cdot y = \begin{pmatrix} \frac{8}{27} & \frac{8}{27} & \frac{2}{9} & \frac{5}{27} \\ \frac{-5}{27} & \frac{-8}{27} & \frac{1}{9} & \frac{7}{27} \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

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

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

$$\left((1,1) \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} - (-1) \right)^2 + \left((1,1) \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} - 1 \right)^2$$

$$+ \left((1,1) \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 2 \right)^2 + \left((1,1) \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} - 3 \right)^2 =$$

$$1 + 1 + 0 + 0 = 2$$

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 $\nabla J(\theta^*) = 0$.

$$J(\Theta) = \sum_{i=1}^{4} \left((\Theta_0, \Theta_1) \begin{pmatrix} 1 \\ x^{(i)} \end{pmatrix} - y_i \right)^2 = \sum_{i=1}^{4} (\Theta_0 + \Theta_1 x^{(i)} - y^{(i)})^2$$

$$= (\Theta_0 + \Theta_1(-1) - (-1))^2 + (\Theta_0 + \Theta_1(-1) - 1)^2 + (\Theta_0 + \Theta_1(1) - 2)^2$$
$$+ (\Theta_0 + \Theta_1(2) - 3)^2$$

$$= (\Theta_0 - \Theta_1 + 1)^2 + (\Theta_0 - \Theta_1 - 1)^2 + (\Theta_0 + \Theta_1 - 2)^2 + (\Theta_0 + 2\Theta_1 - 3)^2$$

$$= \Theta_0^2 - 2\Theta_0\Theta_1 + \Theta_1^2 + 2\Theta_0 - 2\Theta_1 + 1 + \Theta_0^2 - 2\Theta_0\Theta_1 + \Theta_1^2 - 2\Theta_0 + 2\Theta_1 + 1$$
$$+ \Theta_0^2 + 2\Theta_0\Theta_1 + \Theta_1^2 - 4\Theta_0 - 4\Theta_1 + 4 + \Theta_0^2 + 4\Theta_0\Theta_1 + 4\Theta_1^2 - 6\Theta_0 - 12\Theta_1 + 9$$

$$= 4\Theta_0^2 + 2\Theta_0\Theta_1 + 7\Theta_1^2 - 10\Theta_0 - 16\Theta_1 + 15$$

$$\nabla J(\Theta) = \left(\frac{\partial J}{\partial \Theta_0}(\Theta), \frac{\partial J}{\partial \Theta_1}(\Theta) \right)$$

$$\frac{\partial J}{\partial \Theta_0}(\Theta) = \frac{\partial}{\partial \Theta_0}(\Theta) \cdot (4\Theta_0^2 + 2\Theta_0\Theta_1 + 7\Theta_1^2 - 10\Theta_0 - 16\Theta_1 + 15)$$

$$= 8\Theta_0 + 2\Theta_1 - 10$$

$$\frac{\partial J}{\partial \Theta_1}(\Theta) = \frac{\partial}{\partial \Theta_1}(\Theta) \, (4\Theta_0^2 + 2\Theta_0\Theta_1 + 7\Theta_1^2 - 10\Theta_0 - 16\Theta_1 + 15)$$

$$= 2\Theta_0 + 14\Theta_1 - 16$$

$$\nabla J(\theta) = \begin{pmatrix} 8\theta_0 + 2\theta_1 - 10 \\ 2\theta_0 + 14\theta_1 - 16 \end{pmatrix}$$ סך,

נראה כי $\nabla J(\theta^*) = 0$

$$\nabla J(\theta^*) = \begin{pmatrix} 8 \cdot 1 + 2 \cdot 1 - 10 \\ 2 \cdot 1 + 14 \cdot 1 - 16 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

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

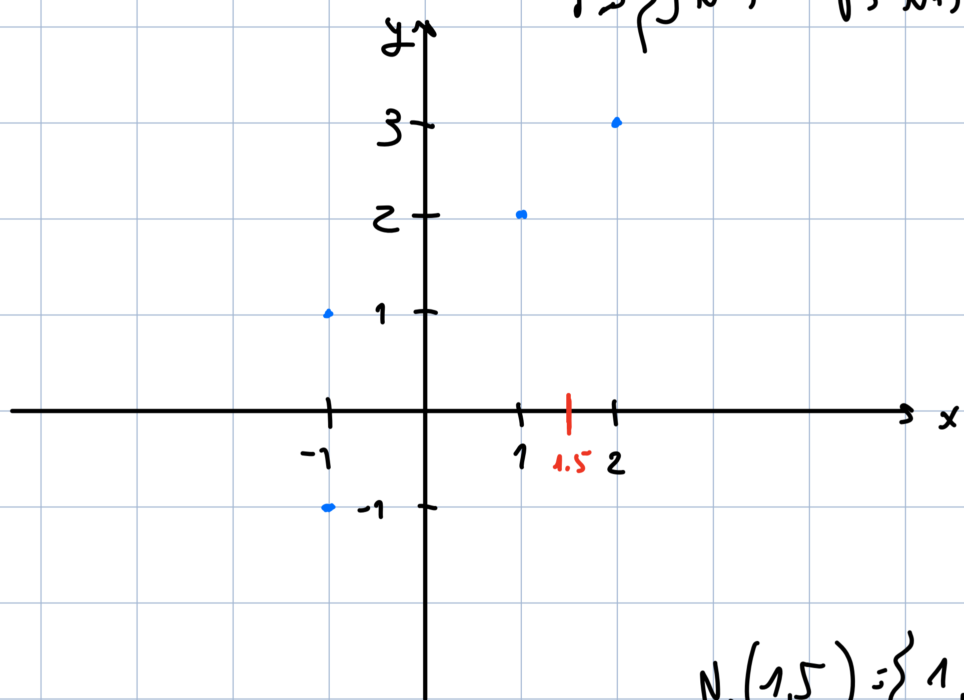
- What is the predicted value of y at this point based on linear regression with θ^*?
- What is the predicted value of y at this point based on K-NN with $K = 2$?

- נזכיר כי $\theta^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ סך עבור θ^* נקבל

$$\theta_0 + \theta_1 x = y \Rightarrow 1 + x = \hat{y}$$

ולכן עבור $x = 1.5$ נקבל $\hat{y} = 2.5$ (התוצאה הסופית)

- נביט בגרף המייצג המתקבל



כחול - דטא קיים

אדום - test

עבור $k=2$ $N_k(1.5) = \{1, 2\}$

ולכן $\hat{y} = \frac{1}{2}(2+3) = 2.5$