Machien learning Homework 1

Gidon Abbas

Id 318482940

1 Question 1

The data matrix with the bias is

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

so $X^{\dagger} = X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 1 & 7 \end{bmatrix}$

We want to find θ vector $X^{\dagger}\theta = X^T Y$

$$\begin{bmatrix} 4 & 1 \\ 1 & 7 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \end{bmatrix} \Rightarrow \begin{pmatrix} 4\theta_0 + \theta_1 \\ \theta_0 + 7\theta_1 \end{pmatrix} = \begin{pmatrix} 5 \\ 8 \end{pmatrix}$$

$$\Downarrow$$

$$\boxed{\begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}}$$

So the minimum loss $J(\theta) = \sum_{i=1}^{n} \left(\theta^{\top}(1, x^{(i)}) - y^{(i)}\right)^2 = \sum_{i=1}^{n}((1 \quad 1)(1, x^{(i)})) - y^{(i)})^2 = ((1 \quad 1)\begin{pmatrix} 1 \\ -1 \end{pmatrix} + 1)^2 \quad + ((1 \quad 1)\begin{pmatrix} 1 \\ -1 \end{pmatrix} - 1)^2 \quad + ((1 \quad 1)\begin{pmatrix} 1 \\ 1 \end{pmatrix} - 2)^2 \quad + ((1 \quad 1)\begin{pmatrix} 1 \\ 2 \end{pmatrix} - 3)^2 = 1 + 1 + 0 + 0 = 2$

$$\Downarrow$$

$$\boxed{J(\theta) = 2}$$

2 Question 2

First we want to express the function $J(\theta)$

$$J(\theta) = \sum_{i=1}^{n} \left(\theta^{\top}(1, x^{(i)}) - y^{(i)}\right)^2 = \sum_{i=1}^{n}((\theta_0 \quad \theta_1)\begin{pmatrix} 1 \\ x^{(i)} \end{pmatrix} - y^{(i)})^2 = \sum_{i=1}^{n}(\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$

$$= \sum_{i=1}^{4} \theta_0^2 + 2\theta_0\theta_1 x^{(i)} + \theta_1^2 (x^{(i)})^2 - 2\theta_0 y^{(i)} - 2\theta_1 x^{(i)} y^{(i)} + (y^{(i)})^2$$

$$= 4\theta_0 + 2(-1 - 1 + 1 + 2)\theta_0\theta_1 + (1 + 1 + 1 + 4)\theta_1^2 - 2(-1 + 1 + 2 + 3)\theta_0 - 2(1 - 1 + 2 + 6)\theta_1 + (1 + 1 + 4 + 9)$$

$$\Downarrow$$

$$\boxed{J(\theta) = 4\theta_0^2 + 2\theta_0\theta_1 + 7\theta_1^2 - 10\theta_0 - 16\theta_1 + 15}$$

Now we need to find the Gradient $\nabla J(\theta)$: $\nabla_\theta J = \begin{bmatrix} \frac{\partial J}{\partial \theta_0} \\ \frac{\partial J}{\partial \theta_1} \end{bmatrix}$

$$J(\theta) = 4\theta_0^2 + 2\theta_0\theta_1 + 7\theta_1^2 - 10\theta_0 - 16\theta_1 + 15$$

$$\frac{\partial J}{\partial \theta_0} = \frac{\partial}{\partial \theta_0}\left(4\theta_0^2 + 2\theta_0\theta_1 + 7\theta_1^2 - 10\theta_0 - 16\theta_1 + 15\right) = 8\theta_0 + 2\theta_1 - 10$$

$$\frac{\partial J}{\partial \theta_1} = \frac{\partial}{\partial \theta_1}\left(4\theta_0^2 + 2\theta_0\theta_1 + 7\theta_1^2 - 10\theta_0 - 16\theta_1 + 15\right) = 2\theta_0 + 14\theta_1 - 16$$

$$\Downarrow$$

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

And lastly we want to show that the θ^* we found implemints that $\nabla J(\theta^*) = 0$
Recall: $\theta^* = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$$\boxed{\nabla J(\begin{pmatrix} 1 \\ 1 \end{pmatrix}) = \begin{bmatrix} 8+2-10 \\ 2+14-16 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}}$$

3 Question 3

3.1 part 1

After we found θ^*, the function to predict y is $\hat{y} = \theta_0 + \theta_1 x$ so at $x = 1.5$ we get that $\boxed{\hat{y} = 1 + 1.5 = 2.5}$

3.2 part 2

To use K-NN were K = 2 we seek for the two nearest neighbors of 1.5 and those are were $x = 1$ and $x = 2$
Recall K-NN $= \frac{1}{k} \sum_{i:x^{(i)} \in N_k(x)} y^{(i)}$

so $\boxed{\hat{y} = \frac{2+3}{2} = \frac{5}{2}}$

P.S. ** I began this assignment together with a partner, and we shared some of the initial code. Please take this into consideration during grading, as any similarities are due to early collaboration, and we completed the rest of the assignment independently.