x_0	x_1	y
1	2	12
1	5	9
1	1	6
1	8	7

1. From the training dataset A, write the normal equations to find θ_0 and θ_1

จาก Normal equation $\theta = (X^TX)^{-1}X^TY$

โจทย์กำหนดให้ จะได้ $X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 1 \\ 1 & 8 \end{bmatrix}$, $Y = \begin{bmatrix} 12 \\ 9 \\ 6 \\ 7 \end{bmatrix}$

ขั้นตอนการคำนวณหาค่า θ ได้ดังนี้

$$X^TX = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 1 & 8 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 1 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 1+1+1+1 & 2+5+1+8 \\ 2+5+1+8 & 4+25+1+64 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 16 \\ 16 & 94 \end{bmatrix}$$

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

จะได้ $(X^TX)^{-1} = \frac{1}{376-256}\begin{bmatrix} 94 & -16 \\ -16 & 4 \end{bmatrix} = \begin{bmatrix} 94/120 & -16/120 \\ -16/120 & 4/120 \end{bmatrix}$

$$(X^TX)^{-1}X^T = \begin{bmatrix} 94/120 & -16/120 \\ -16/120 & 4/120 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 1 & 8 \end{bmatrix}$$

$$= \begin{bmatrix} 94/120 - 32/120 & 94/120 - 80/120 & 94/120 - 16/120 & 94/120 - 128/120 \\ -16/120 + 8/120 & -16/120 + 20/120 & -16/120 + 4/120 & -16/120 + 32/120 \end{bmatrix}$$

$$= \begin{bmatrix} 62/120 & 14/120 & 78/120 & -34/120 \\ -8/120 & 4/120 & -12/120 & 16/120 \end{bmatrix}$$

ดังนั้น $\theta = (X^TX)^{-1}X^TY = \begin{bmatrix} 62/120 & 14/120 & 78/120 & -34/120 \\ -8/120 & 4/120 & -12/120 & 16/120 \end{bmatrix} \begin{bmatrix} 12 \\ 9 \\ 6 \\ 7 \end{bmatrix}$

$= \begin{bmatrix} (744 + 126 + 468 - 238)/120 \\ (-96 + 36 - 72 + 112)/120 \end{bmatrix}$

$= \begin{bmatrix} 1100/120 \\ -20/120 \end{bmatrix} = \begin{bmatrix} 9.167 \\ -0.167 \end{bmatrix}$

ดังนั้น $\theta_0 = 9.167$ และ $\theta_1 = -0.167$

Given a training dataset B of three features X and one output Y, as shown below.

x_0	x_1	x_2	x_3	y
1	2	3	6	12
1	5	9	7	9
1	1	4	2	6
1	8	5	3	7

2. Given the total iteration = 3, learning rate = 0.05, and initial gradient's value θ_0=1, θ_1=1, θ_2=1, and θ_3=1, respectively.
From the training dataset B, use **batch gradient descent** to find θ_0, θ_1, θ_2, and θ_3 for each iteration step.

3. Given the total iteration = 3, learning rate = 0.05, and initial gradient's value θ_0=1, θ_1=1, θ_2=1, and θ_3=1, respectively.
From the training dataset B, use **stochastic gradient descent** to find θ_0, θ_1, θ_2, and θ_3 for each iteration step.
Let iteration 1 randomly picks row = 3
Let iteration 2 randomly picks row = 1
Let iteration 3 randomly picks row = 4

4. Given the total iteration = 3, learning rate = 0.05, batch size = 2, and initial gradient's value θ_0 =1, θ_1 =1, θ_2 =1, and θ_3 =1, respectively.
From the training dataset B, use **mini-batch gradient descent** to find θ_0, θ_1, θ_2, and θ_3 for each iteration step.

2. ถ้ากำหนด $\alpha = 0.05$, $m = 4$, Batch

iteration 1 $\theta_0 = 1$, $\theta_1 = 1$, $\theta_2 = 1$, $\theta_3 = 1$

$h_\theta(x^{(1)}) = 1(1) + 1(2) + 1(3) + 1(6) = 12$

$h_\theta(x^{(2)}) = 1(1) + 1(5) + 1(9) + 1(7) = 22$

$h_\theta(x^{(3)}) = 1(1) + 1(1) + 1(4) + 1(2) = 8$

$h_\theta(x^{(4)}) = 1(1) + 1(8) + 1(5) + 1(3) = 17$

สมการ $\theta_j = \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$

แทนค่า $\theta_0 = 1 - 0.05 \frac{1}{4}\left((12-12)(1) + (22-9)(1) + (8-6)(1) + (17-7)(1)\right)$

$= 1 - \frac{0.05}{4}(25) = \boxed{0.69}$

$\theta_1 = 1 - 0.05 \frac{1}{4}\left((12-12)(2) + (22-9)(5) + (8-6)(1) + (17-9)(8)\right)$

$= 1 - \frac{0.05}{4}(147) = \boxed{-0.84}$

$\theta_2 = 1 - \frac{0.05}{4}\left((12-12)(3) + (22-9)(9) + (8-6)(4) + (17-7)(5)\right)$

$= 1 - \frac{0.05}{4}(175) = \boxed{-1.19}$

$\theta_3 = 1 - \frac{0.05}{4}\left((12-12)(6) + (22-9)(7) + (8-6)(2) + (17-7)(3)\right)$

$= 1 - \frac{0.05}{4}(125) = \boxed{-0.56}$

iteration 2 $\theta_0 = 0.69,\ \theta_1 = -0.84,\ \theta_2 = -1.19,\ \theta_3 = -0.56$

$h_\theta(x^{(1)}) = 0.69(1) - 0.84(2) - 1.19(3) - 0.56(6) = -7.92$

$h_\theta(x^{(2)}) = 0.69(1) - 0.84(5) - 1.19(9) - 0.56(7) = -18.14$

$h_\theta(x^{(3)}) = 0.69(1) - 0.84(1) - 1.19(4) - 0.56(2) = -6.03$

$h_\theta(x^{(4)}) = 0.69(1) - 0.84(8) - 1.19(5) - 0.56(3) = -13.66$

update $\theta_0 = 0.69 - \frac{0.05}{4}\big((-7.92-12)(1) + (-18.14-9)(1) + (-6.03-6)(1) + (-13.66-7)(1)\big)$

$= 0.69 - \frac{0.05}{4}(-79.75) = \boxed{1.69}$

$\theta_1 = -0.84 - \frac{0.05}{4}\big((-7.92-12)(2) + (-18.14-9)(5) + (-6.03-6)(1) + (-13.66-7)(8)\big)$

$= -0.84 - \frac{0.05}{4}(-352.85) = \boxed{3.57}$

$\theta_2 = -1.19 - \frac{0.05}{4}\big((-7.92-12)(3) + (-18.14-9)(9) + (-6.03-6)(4) + (-13.66-7)(5)\big)$

$= -1.19 - \frac{0.05}{4}(-455.44) = \boxed{4.50}$

$\theta_3 = -0.56 - \frac{0.05}{4}\big((-7.92-12)(6) + (-18.14-9)(7) + (-6.03-6)(2) + (-13.66-7)(3)\big)$

$= -0.56 - \frac{0.05}{4}(-395.54) = \boxed{4.38}$

iteration 3 $\theta_0 = 1.69,\ \theta_1 = 3.57,\ \theta_2 = 4.50,\ \theta_3 = 4.38$

$$h_\theta(x^{(1)}) = 1.69(1) + 3.57(2) + 4.5(3) + 4.38(1) = 48.61$$

$$h_\theta(x^{(2)}) = 1.69(1) + 3.57(5) + 4.5(9) + 4.38(7) = 90.7$$

$$h_\theta(x^{(3)}) = 1.69(1) + 3.57(1) + 4.5(4) + 4.38(2) = 32.02$$

$$h_\theta(x^{(4)}) = 1.69(1) + 3.57(8) + 4.5(5) + 4.38(3) = 65.89$$

and so

$$\theta_0 = 1.69 - \frac{0.05}{4}\Big((48.61-12)(1) + (90.7-9)(1) + (32.02-6)(1) + (65.89-9)(1)\Big)$$

$$= 1.69 - \frac{0.05}{4}(203.21) = \boxed{-0.85}$$

$$\theta_1 = 3.57 - \frac{0.05}{4}\Big((48.61-12)(2) + (90.7-9)(5) + (32.02-6)(1) + (65.89-7)(8)\Big)$$

$$= 3.57 - \frac{0.05}{4}(998.86) = \boxed{-8.69}$$

$$\theta_2 = 4.5 - \frac{0.05}{4}\Big((48.61-12)(3) + (90.7-9)(9) + (32.02-6)(4) + (65.89-7)(5)\Big)$$

$$= 4.5 - \frac{0.05}{4}(1243.66) = \boxed{-11.05}$$

$$\theta_3 = 4.38 - \frac{0.05}{4}\Big((48.61-12)(6) + (90.7-9)(7) + (32.02-6)(2) + (65.89-7)(3)\Big)$$

$$= 4.38 - \frac{0.05}{4}(1020.27) = \boxed{-8.37}$$

3. กำหนด $\alpha = 0.05$, Stochastic

iteration 1 $\theta_0 = 1, \theta_1 = 1, \theta_2 = 1, \theta_3 = 1$, select row 3

กฎ $\theta_j = \theta_j - \alpha (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$

$i = 3$ จะได้ $h_\theta(x^{(3)}) = 1(1) + 1(1) + 1(4) + 1(2) = 8$

$\theta_0 = 1 - 0.05(8-6)(1) = \boxed{0.9}$

$\theta_1 = 1 - 0.05(8-6)(1) = \boxed{0.9}$

$\theta_2 = 1 - 0.05(8-6)(4) = \boxed{0.6}$

$\theta_3 = 1 - 0.05(8-6)(2) = \boxed{0.8}$

iteration 2 $\theta_0 = 0.9, \theta_1 = 0.9, \theta_2 = 0.6, \theta_3 = 0.8$, select row 1

$i = 1$ จะได้ $h_\theta(x^{(1)}) = 0.9(1) + 0.9(2) + 0.6(3) + 0.8(6) = 9.3$

$\theta_0 = 0.9 - 0.05(9.3-12)(1) = \boxed{1.035}$

$\theta_1 = 0.9 - 0.05(9.3-12)(2) = \boxed{1.17}$

$\theta_2 = 0.6 - 0.05(9.3-12)(3) = \boxed{1.005}$

$\theta_3 = 0.8 - 0.05(9.3-12)(6) = \boxed{1.61}$

iteration 3 $\theta_0 = 1.035, \theta_1 = 1.17, \theta_2 = 1.005, \theta_3 = 1.61$, select row 4

$i = 4$ จะได้ $h_\theta(x^{(4)}) = 1.035(1) + 1.17(8) + 1.005(5) + 1.61(3)$

$= 20.25$

$\theta_0 = 1.035 - 0.05(20.25-7)(1) = \boxed{0.37}$

$\theta_1 = 1.17 - 0.05(20.25-7)(8) = \boxed{-4.13}$

$\theta_2 = 1.005 - 0.05(20.25-7)(5) = \boxed{-2.31}$

$\theta_3 = 1.61 - 0.05(20.25-7)(3) = \boxed{-0.38}$

4. กำหนด $\alpha = 0.05$, $b = 2$, Mini-batch

iteration 1 $\theta_0 = 1, \theta_1 = 1, \theta_2 = 1, \theta_3 = 1$, select row 1 & 2

$$h_\theta(x^{(1)}) = 1(1) + 1(2) + 1(3) + 1(6) = 12$$

$$h_\theta(x^{(2)}) = 1(1) + 1(5) + 1(9) + 1(7) = 22$$

จาก $\theta_j = \theta_j - \alpha \frac{1}{b} \sum (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$

จะได้ $\theta_0 = 1 - \frac{0.05}{2}((12-12)(1) + (22-9)(1)) = \boxed{0.675}$

$\theta_1 = 1 - \frac{0.05}{2}((12-12)(2) + (22-9)(5)) = \boxed{-0.625}$

$\theta_2 = 1 - \frac{0.05}{2}((12-12)(3) + (22-9)(9)) = \boxed{-1.925}$

$\theta_3 = 1 - \frac{0.05}{2}((12-12)(6) + (22-9)(7)) = \boxed{-1.275}$

iteration 2 $\theta_0 = 0.675, \theta_1 = -0.625, \theta_2 = -1.925, \theta_3 = -1.275$, row 3 & 4

$$h_\theta(x^{(3)}) = 0.675(1) - 0.625(1) - 1.925(4) - 1.275(2) = -10.2$$

$$h_\theta(x^{(4)}) = 0.675(1) - 0.625(8) - 1.925(5) - 1.275(3) = -17.78$$

$\theta_0 = 0.675 - \frac{0.05}{2}((-10.2-6)(1) + (-17.78-9)(1)) = \boxed{1.70}$

$\theta_1 = -0.625 - \frac{0.05}{2}((-10.2-6)(1) + (-17.78-9)(8)) = \boxed{4.74}$

$\theta_2 = -1.925 - \frac{0.05}{2}((-10.2-6)(4) + (-17.78-9)(5)) = \boxed{2.79}$

$\theta_3 = -1.275 - \frac{0.05}{2}((-10.2-6)(2) + (-17.78-9)(3)) = \boxed{1.39}$

iteration 3 $\theta_0 = 1.70, \theta_1 = 4.74, \theta_2 = 2.79, \theta_3 = 1.39$, row 2 & 3

$$h_\theta(x^{(2)}) = 1.70(1) + 4.74(5) + 2.79(9) + 1.39(7) = 60.24$$

$$h_\theta(x^{(3)}) = 1.70(1) + 4.74(1) + 2.79(4) + 1.39(2) = 20.38$$

$\theta_0 = 1.70 - \frac{0.05}{2}((60.24-9)(1) + (20.38-6)(1)) = \boxed{0.06}$

$\theta_1 = 4.74 - \frac{0.05}{2}((60.24-9)(5) + (20.38-6)(1)) = \boxed{-2.03}$

$\theta_2 = 2.79 - \frac{0.05}{2}((60.24-9)(9) + (20.38-6)(4)) = \boxed{-10.18}$

$\theta_3 = 1.39 - \frac{0.05}{2}((60.24-9)(7) + (20.38-6)(2)) = \boxed{-8.30}$