

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

an dataset A assüsum, known  $\theta_0, \theta_1$   
 in Normal Equation  $\theta = (X^T X)^{-1} X^T Y$

anhand z.B.  $X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 1 \\ 1 & 8 \end{bmatrix}, Y = \begin{bmatrix} 12 \\ 9 \\ 6 \\ 7 \end{bmatrix}$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 1 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 1 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 16 \\ 16 & 94 \end{bmatrix}$$

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

$$(X^T X)^{-1} X^T Y = \begin{bmatrix} 94/120 & -16/120 \\ -16/120 & 4/120 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 1 & 8 \end{bmatrix} \cdot \begin{bmatrix} 62/120 & 14/120 & 98/120 & -34/120 \\ -8/120 & 4/120 & -12/120 & 16/120 \end{bmatrix}$$

$$\theta = (X^T X)^{-1} X^T Y = \begin{bmatrix} 62/120 & 14/120 & 98/120 & -34/120 \\ -8/120 & 4/120 & -12/120 & 16/120 \end{bmatrix} \begin{bmatrix} 12 \\ 9 \\ 6 \\ 7 \end{bmatrix} = \begin{bmatrix} 1160/120 \\ -20/120 \end{bmatrix} = \begin{bmatrix} 9.67 \\ -0.167 \end{bmatrix}$$

also  $\theta_0 = 9.67 \quad \theta_1 = -0.167$

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

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

2. Given the total iteration = 3, learning rate = 0.05, and initial gradient's value  $\theta_0=1$ ,  $\theta_1=1$ ,  $\theta_2=1$ , and  $\theta_3=1$ , respectively. From the training dataset B, use **batch gradient descent** to find  $\theta_0$ ,  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$  for each iteration step.

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

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

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

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

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

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

update  $\theta_j = \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h(x^{(i)}) - y^{(i)}) x_j^{(i)}$

for  $\theta_0 = 1 - 0.05 \frac{1}{4} ((12-12)(1) + (22-9)(1) + (8-6)(1) + (17-7)(1))$   
 $= 1 - \frac{0.05}{4} (25) = 0.69$

$\theta_1 = 1 - 0.05 \frac{1}{4} ((12-12)(2) + (22-9)(5) + (8-6)(1) + (17-7)(8))$   
 $= -0.84$

$\theta_2 = 1 - 0.05 \frac{1}{4} ((12-12)(3) + (22-9)(9) + (8-6)(4) + (17-7)(5))$   
 $= -1.19$

$\theta_3 = 1 - 0.05 \frac{1}{4} ((12-12)(6) + (22-9)(7) + (8-6)(2) + (17-7)(3))$   
 $= -0.56$

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

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

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

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

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

$$\tilde{J}_0 \quad \theta_0 = 0.69 - \frac{0.05}{4} [ (-7.92-12)(1) + (-18.14-9)(1) + (-6.03-6)(1) + (-13.66-7)(1) ]$$

$$= 1.69$$

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

$$= 3.59$$

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

$$= 4.5$$

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

$$= 4.38$$

iteration 3  $\theta_0 = 1.69$   $\theta_1 = 3.59$   $\theta_2 = 4.5$   $\theta_3 = 4.38$

$$h_{\theta}(x^{(1)}) = 1.69(1) + 3.59(2) + 4.5(3) + 4.38(6) = 48.61$$

$$h_{\theta}(x^{(2)}) = 1.69(1) + 3.59(5) + 4.5(9) + 4.38(9) = 90.7$$

$$h_{\theta}(x^{(3)}) = 1.69(1) + 3.59(1) + 4.5(4) + 4.38(2) = 32.02$$

$$h_{\theta}(x^{(4)}) = 1.69(1) + 3.59(8) + 4.5(5) + 4.38(3) = 65.89$$

$$\theta_0 = 1.69 - \frac{0.05}{4} [ (48.61-12)(1) + (90.7-9)(1) + (32.02-6)(1) + (65.89-7)(1) ] = -0.85$$

$$\theta_1 = 3.59 - \frac{0.05}{4} [ (48.61-12)(2) + (90.7-9)(5) + (32.02-6)(1) + (65.89-7)(8) ] = -8.67$$

$$\theta_2 = 4.5 - \frac{0.05}{4} [ (48.61-12)(3) + (90.7-9)(9) + (32.02-6)(4) + (65.89-7)(5) ] = -11.05$$

$$\theta_3 = 4.38 - \frac{0.05}{4} [ (48.61-12)(6) + (90.7-9)(9) + (32.02-6)(2) + (65.89-7)(3) ] = -8.37$$

### 3. random $\alpha = 0.05$ stochastic

iteration 0  $\theta_0 = \theta_1 = \theta_2 = \theta_3 = 1$  select row = 3

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

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

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

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

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

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

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

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

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

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

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

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

$$i=4 \quad h_\theta(x^{(4)}) = 1.035(4) + 1.17(8) + 1.005(5) + 1.61(3) = 20.25$$

$$\theta_0 = 1.035 - 0.05(20.25-7)(1) = 0.39$$

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

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

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

4. extend  $d=0.05$ ,  $b=2$  Mini batch

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

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

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

$$\text{w.r.t } \theta_j: \theta_j = \frac{1}{L} \sum_{i=1}^L (h(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

$$\text{w.r.t } \theta_0: 1 - \frac{0.05}{2} [(12-12)(1) + (22-9)(1)] = 0.675$$

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

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

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

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

$$h_{\theta}(x^{(3)}) = 0.675(1) - 0.625(1) - 1.925(4) - 1.275(2) = -10.1$$

$$h_{\theta}(x^{(4)}) = 0.675(1) - 0.625(8) - 1.925(5) - 1.275(3) = -17.75$$

$$\text{w.r.t } \theta_0: 0.675 - \frac{0.05}{2} [(10.2-6)(1) + (-17.78-7)(1)] = 1.70$$

$$\theta_1 = -0.625 - \frac{0.05}{2} [(10.2-6)(1) + (-17.78-7)(9)] = 4.74$$

$$\theta_2 = -1.925 - \frac{0.05}{2} [(10.2-6)(4) + (-17.78-7)(5)] = 2.79$$

$$\theta_3 = -1.275 - \frac{0.05}{2} [(10.2-6)(2) + (-17.78-7)(3)] = 1.39$$

iteration 3  $\theta_0 = 1.7$   $\theta_1 = 4.74$   $\theta_2 = 2.79$   $\theta_3 = 1.39$  select row 2 & 3

$$h_{\theta}(x^{(2)}) = 1.7(1) + 4.74(5) + 2.79(9) + 1.39(7) = 60.24$$

$$h_{\theta}(x^{(3)}) = 1.7(1) + 4.74(1) + 2.79(4) + 1.39(2) = 20.38$$

$$\theta_0 = 1.7 - \frac{0.05}{2} [(60.24-9)(1) + (20.38-6)(1)] = 0.06$$

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

$$\theta_2 = 2.79 - \frac{0.05}{2} [(60.24-9)(9) + (20.38-6)(4)] = -6.18$$

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