

* سوال دوم

$$h_{\theta}(x) = b + w_1 x + w_2 x^2$$

x	1	2	1	-2
y	2	0	4	1

$$X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & -2 \end{bmatrix}, y = \begin{bmatrix} 2 \\ 0 \\ 4 \\ 1 \end{bmatrix}, \theta = \begin{bmatrix} b \\ w_1 \\ w_2 \end{bmatrix}, \alpha = 0.2, \lambda = 0.1$$

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m |h_{\theta}(x^i) - y^i| + \frac{\lambda}{m} \sum_{j=1}^n |\theta_j|$$

$$\text{هدف: } \min_{\theta} \left[\frac{1}{m} \sum_{i=1}^m |y_i - x_i^T \theta| + \frac{\lambda}{m} \sum_{j=1}^n |\theta_j| \right]$$

برای تحقق شدن این هدف کمترین از عبارت فوق نسبت به θ گزینیم و سپس با روش نزول گزینیم (subgradient): چون MAE در بعضی نقاط مشتق ندارد. θ^* بهینه را پیدا کنیم.

$$J(\theta) = \begin{cases} \frac{1}{m} \sum_{i=1}^m (y_i - x_i^T \theta) + \frac{\lambda}{m} \sum_{j=1}^n |\theta_j| & ; \text{if } y_i \geq x_i^T \theta \\ \frac{1}{m} \sum_{i=1}^m (x_i^T \theta - y_i) + \frac{\lambda}{m} \sum_{j=1}^n |\theta_j| & ; \text{if } y_i < x_i^T \theta \end{cases}$$

$$\frac{\partial}{\partial \theta_j} \left[\frac{\lambda}{m} \sum_{j=1}^n |\theta_j| \right] = \begin{cases} \frac{\lambda}{m} & ; \text{if } \theta_j \geq 0 \\ -\frac{\lambda}{m} & ; \text{if } \theta_j < 0 \end{cases}$$

$$\frac{\partial J(\theta)}{\partial \theta} = \begin{cases} \sum_{i=1}^m -\frac{1}{m} x_i^T + \frac{\lambda}{m} [1 \ 1]^T & ; \text{if } y_i \geq x_i^T \theta \text{ and } \theta_j \geq 0 \\ \sum_{i=1}^m -\frac{1}{m} x_i^T - \frac{\lambda}{m} [1 \ 1]^T & ; \text{if } y_i \geq x_i^T \theta \text{ and } \theta_j < 0 \\ \sum_{i=1}^m \frac{1}{m} x_i^T + \frac{\lambda}{m} [1 \ 1]^T & ; \text{if } y_i < x_i^T \theta \text{ and } \theta_j \geq 0 \\ \sum_{i=1}^m \frac{1}{m} x_i^T - \frac{\lambda}{m} [1 \ 1]^T & ; \text{if } y_i < x_i^T \theta \text{ and } \theta_j < 0 \end{cases}$$

حالا برنامه شده رو با یک بار run کنیم . پس داریم :

* توجه : اگر درستم لهه : $j = 0, 1, \dots, n$; $\theta_j = \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta}$

* توجه : مقدار دهی اولیه به ضرایب $\theta = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ بوده است .

$$h_{\theta}(x) = x\theta = \begin{bmatrix} 1 & 1 & 2 & 4 \\ 1 & 1 & -2 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \\ 3 \end{bmatrix} = y_{\text{pred}}$$

* از آنجایی که مقدار دهی اولیه θ تمام یک بوده پس در گزینش داریم (بایک) :

$$\frac{\partial J(\theta)}{\partial \theta} = \begin{cases} -\frac{1}{m} \sum_{i=1}^m x_i^T + \frac{\lambda}{m} [1 \ 1 \ 1]^T ; \text{ if } y_{\text{pred}} \geq y_{\text{true}} \\ \frac{1}{m} \sum_{i=1}^m x_i^T + \frac{\lambda}{m} [1 \ 1 \ 1]^T ; \text{ if } y_{\text{pred}} < y_{\text{true}} \end{cases}$$

$$\theta^* = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 0.2 \times \frac{1}{4} \begin{bmatrix} -2 \\ 0 \\ -8 \end{bmatrix} + \begin{bmatrix} 0.025 \\ 0.025 \\ 0.025 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -0.1 \\ 0 \\ -0.4 \end{bmatrix} + \begin{bmatrix} 0.025 \\ 0.025 \\ 0.025 \end{bmatrix}$$

$$= \begin{bmatrix} 1.125 \\ 1.025 \\ 1.425 \end{bmatrix} \quad \leftarrow \text{ضرایب به روز شده بعد از یک ایپاک}$$