MFML Assignment 2

Shubhangi Agarwal
2025ab05196

Ans Given linear regression Model

$$\hat{y}_i = w_0 + w_1 x_i$$

Shop Number	Operating hrs (x)	Consumption
1	1	3
2	2	5
3	4	9
4	6	13
5	8	17

~~Given~~ Now Squared error loss function $J(w_0, w_1) = \dfrac{\sum_{i=1}^{5} (\hat{y}_i - y_i)^2}{2}$

$$J(w_0, w_1) = \frac{1}{2}\Big[(w_0 + w_1 - 3)^2 + (w_0 + 2w_1 - 5)^2 + (w_0 + 4w_1 - 9)^2 + (w_0 + 6w_1 - 13)^2 + (w_0 + 8w_1 - 17)^2\Big]$$

$$= \frac{1}{2}\Big[(w_0 + w_1)^2 + 9 - 6(w_0 + w_1) + (w_0 + 2w_1)^2 + 25 - 10(w_0 + 2w_1) + (w_0 + 4w_1)^2 + 81 - 18(w_0 + 4w_1) + (w_0 + 6w_1)^2$$

$$= \frac{1}{2}\Big[w_0^2 + w_1^2 + 2w_0 w_1 - \cancel{6w_0} - 6w_1$$

Shubhang Agrawal
2025ab05196

$$= \frac{1}{2} \Big[w_0^2 + w_1^2 + 9 + 2w_0w_1 - 6w_0 - 6w_1$$
$$+ w_0^2 + 4w_1^2 + 25 + 4w_0w_1 - 10w_0 - 20w_1$$
$$+ w_0^2 + 16w_1^2 + 81 + 8w_0w_1 - 18w_0 - 72w_1$$
$$+ w_0^2 + 36w_1^2 + 169 + 12w_0w_1 - 26w_0 - 156w_1$$
$$+ w_0^2 + 64w_1^2 + 289 + 16w_0w_1 - 34w_0 - 272w_1 \Big]$$

$$= \frac{1}{2} \left[5w_0^2 + 121w_1^2 + 573 + 42w_0w_1 - 94w_0 - 526w_1 \right]$$

$$\frac{\partial J(w_0, w_1)}{\partial w_0} = \frac{1}{2} \left[10w_0 + 42w_1 - 94 \right]$$

$$= 5w_0 + 21w_1 - 47$$

$$\frac{\partial J(w_0, w_1)}{\partial w_1} = \frac{1}{2} \left[242w_1 + 42w_0 - 526 \right]$$

$$\frac{\partial^2 J(w_0, w_1)}{\partial w_0^2} = 5 \qquad \frac{\partial^2 J(w_0, w_1)}{\partial w_0 \partial w_1} = 21$$

$$\frac{\partial^2 J(w_0, w_1)}{\partial w_1 \partial w_0} = 21 \qquad \frac{\partial^2 J(w_0, w_1)}{\partial w_1^2} = 121$$

$$H_{J(w_0, w_1)} = \begin{bmatrix} \frac{\partial^2 J(w_0, w_1)}{\partial w_0^2} & \frac{\partial^2 J(w_0, w_1)}{\partial w_0 \partial w_1} \\ \frac{\partial^2 J}{\partial w_1 \partial w_0} & \frac{\partial^2 J}{\partial w_1^2} \end{bmatrix} = \begin{bmatrix} 5 & 21 \\ 21 & 121 \end{bmatrix}$$

$$= 164,$$

$$> 0.$$

Shubham Agrawal
2025-60091

Since $H_{J(w_0,w_1)} > 0$ $\therefore$ $J(w_0,w_1)$ is a convex function.

b) $$\nabla J = \begin{pmatrix} \dfrac{\partial J}{\partial w_0} \\[2ex] \dfrac{\partial J}{\partial w_1} \end{pmatrix} = \begin{bmatrix} 5w_0 + 21w_1 - 47 \\[2ex] 121w_1 + 21w_0 - \overset{263}{\cancel{269}} \end{bmatrix}$$

(as calculated in part (a))

Now Given $w_0^{(0)} = 2$ $\quad w_1^{(0)} = 2$

c) 1. $\eta = 0.05$

$$w^{(0)} = \begin{bmatrix} w_0^{(0)} \\ w_1^{(0)} \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\cancel{\nabla J \Big|_{w_0^{(0)}} = \frac{w_1^{(0)}}{w}}$$

$$\nabla J \Big|_{w^{(0)}} = \begin{bmatrix} \dfrac{\partial J}{\partial w_0} \\[2ex] \dfrac{\partial J}{\partial w_1} \end{bmatrix}_{w^{(0)}} = \begin{bmatrix} 5\times 2 + 21\times 2 - 47 \\[2ex] 121\times 2 + 21\times 2 - 263 \end{bmatrix}$$

$$\cancel{w^{(1)} : w_0^{(0)}} \qquad = \begin{bmatrix} 5 \\ 21 \end{bmatrix}$$

$$w^{(1)} = w^{(0)} - \eta_1 \nabla J \Big|_{w^{(0)}}$$

$$= \begin{bmatrix} 2 \\ 2 \end{bmatrix} - 0.05 \begin{bmatrix} 5 \\ 21 \end{bmatrix}$$

Shubhang Agarwal
2025 ad 05196

$$= \begin{bmatrix} 2 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.25 \\ 1.05 \end{bmatrix}$$

$$= \begin{bmatrix} 1.75 \\ 0.95 \end{bmatrix}$$

$$w_2^{(2)} = w^{(1)} - \eta \nabla J |_{w_2^{(1)}}$$

$$= \begin{bmatrix} 1.75 \\ 0.95 \end{bmatrix} - 0.05 \begin{bmatrix} 5 \times 1.75 + 21 \times 0.95 - 47 \\ 121 \times 0.95 + 21 \times \cancel{0.95}^{1.75} - \cancel{258}^{263} \end{bmatrix}$$

$$= \begin{bmatrix} 1.75 \\ 0.7 \end{bmatrix} - \cancel{0.05 \begin{bmatrix} -18.05 \\ -136.55 \end{bmatrix}} \quad 0.05 \begin{bmatrix} -18.3 \\ -111.3 \end{bmatrix}$$

$$= \begin{bmatrix} 1.75 \\ 0.7 \end{bmatrix} + \cancel{\begin{bmatrix} 1.1775 \\ 6.8275 \end{bmatrix}} \quad \begin{bmatrix} 0.915 \\ 5.565 \end{bmatrix}$$

$$= \cancel{\begin{bmatrix} 2.8675 \\ 6.5275 \end{bmatrix}} \quad \begin{bmatrix} 2.665 \\ 6.515 \end{bmatrix}$$

2. $\eta_t = \eta_0 e^{-kt}$ $\quad \eta_0 = 0.1 \quad k = 0.4 \quad t \in \mathbb{N}$

$$\eta_1 = 0.1 \, e^{-0.4 \times 1}$$
$$= 0.067$$

$$w^{(1)} = w^{(0)} - \eta \nabla J |_{w^{(0)}}$$

Shubhang Agarwal
2025ab05156

$$w^{(1)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} - 0.067 \begin{bmatrix} 5 \\ 21 \end{bmatrix}$$

$$= \begin{bmatrix} 1.665 \\ 0.593 \end{bmatrix}$$

$$w^{(2)} = w^{(1)} - \eta_2 \nabla J \big|_{w^{(1)}}$$

$$\eta_2 = \eta_0 e^{-kt}$$

$$= 0.1 \times e^{-0.4 \times 2}$$

$$= 0.045$$

$$w^{(2)} = \begin{bmatrix} 1.665 \\ 0.258 \end{bmatrix} - 0.045 \begin{bmatrix} 5 \times 1.665 + 21 \times 0.258 - 47 \\ 121 \times 0.258 + 21 \times 1.665 - 258 \end{bmatrix}$$

$$= \begin{bmatrix} 1.665 \\ 0.258 \end{bmatrix} - 0.045 \begin{bmatrix} -33.257 \\ -191.817 \end{bmatrix}$$

$$= \begin{bmatrix} 1.665 \\ 0.258 \end{bmatrix} + \begin{bmatrix} 1.4966 \\ 8.6318 \end{bmatrix}$$

$$= \begin{bmatrix} 3.1616 \\ 8.8898 \end{bmatrix}$$

Shubhangi Agarwal
2015 ch 05196

$$w^{(2)} = w^{(1)} - \eta_2 \nabla J \big|_{w^{(1)}}$$

$$= \begin{bmatrix} 1.665 \\ 0.593 \end{bmatrix} - 0.045 \begin{bmatrix} 5 \times 1.665 + 21 \times 0.593 - 47 \\ 121 \times 0.593 + 21 \times 1.665 - 263 \end{bmatrix}$$

$$= \begin{bmatrix} 1.665 \\ 0.593 \end{bmatrix} - 0.045 \begin{bmatrix} -26.222 \\ -156.282 \end{bmatrix}$$

$$= \begin{bmatrix} 2.845 \\ 7.62569 \end{bmatrix}$$

d) The convergence is more stable in case of constant η.

The convergence is faster in case of exponentially decaying learning rate.

e) fixed descent direction, $d = -\nabla J(w)$

step size $= \eta$

$$\phi(\eta) = J(w + \eta d)$$

Shubhangi Agarwal
2025 ds05196

~~1. Binary Search~~

$$\phi(\eta) = J(w + \eta d)$$

$$w + \eta d = \begin{bmatrix} 2 \\ 2 \end{bmatrix} + \eta \begin{bmatrix} -5 \\ -21 \end{bmatrix}$$

$$= \begin{bmatrix} 2 - 5\eta \\ 2 - 21\eta \end{bmatrix}$$

$$\hat{y}_i = (2 - 5\eta) + (2 - 21\eta) \times x_i$$

$$e_i(\eta) = \hat{y}_i - y_i = (2 - 5\eta) + (2 - 21\eta) x_i - y_i$$

$$\phi(\eta) = \sum_{i=1}^{5} e_i(\eta)$$

Solving we get

$$\phi(\eta) = 82\eta^2 - 26\eta + 2.5$$

Clearly, $\phi(\eta)$ is a convex quadratic in η.

1. Binary Search

Initial interval: $[0, 1]$

$$\text{midpoint} = \frac{0+1}{2} = 0.5$$

$$\phi(0.5) = 82 \times 0.25 - 26 \times 0.5 + 2.5$$

$$= 10$$

$$\phi(0.5 + 0.001) = 82 \times 0.501^2 - 26 \times 0.501 + 2.5$$

$$\approx 10.005$$

Shubhang Agarwal
2021AB05184

Since $\phi(m) < \phi(m + 10^{-3})$ $\therefore$ minimum lies in left interval

Interval 2 : $[0, 0.5]$

midpoint $= \frac{0 + 0.5}{2} = 0.25$

$\phi(0.25) = 82 \times 0.25^2 - 26 \times 0.25 + 2.5$

$= 1.125$

$\phi(0.25 + 0.001) = 82 \times 0.251^2 - 26 \times 0.251 + 2.5$

≈ 1.13

Since $\phi(0.25) < \phi(0.251)$ $\therefore$ minimum lies in interval left of 0.25

Interval 3 : $[0, 0.25]$

$\therefore$ Binary search gives interval $[0, 0.25]$ after 2 iteration

2. Golden section search

Interval 0 : $[0, 1]$

Given interior points: $M_1 = \frac{1}{4}$ $M_2 = \frac{3}{4}$

$\phi(0.25) = 1.125$

$\phi(0.75) = 29.625$

Shubhangi Agarwal
2015cs35126

$\phi(M_1) < \phi(M_2)$

$\therefore$ The minimum lies in interval $[0, 0.75]$

for the second iteration,

interval : $[0, 0.75]$

valid choice for $M_1 = 0.25$ $M_2 = 0.5$

(one of which is already calculated)