Mathematical foundations for Machine learning

Name : B. Haripriya

ID : 2025AA05640

Question

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

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

$$J(w_0, w_1) = \frac{1}{2} \sum_{i=1}^{5} (\hat{y}_i - y_i)^2$$

$$(a \pm b)^2 = a^2 \pm 2ab + b^2$$

$$\Rightarrow \left((a+b)-c\right)^2 = (a+b)^2 - 2(a+b)c + c^2$$

$$\Rightarrow J(w_0, w_1) = \frac{1}{2}\sum_{i=1}^{5}(\hat{y}_i - y_i)^2$$

$$= \frac{1}{2}\sum_{i=1}^{5}(w_0 + w_1 x_i - y_i)^2 \hookrightarrow (1)$$

$$\Rightarrow \text{Now, } (w_0 + w_1 x_i - y_i)^2 = (w_0 + w_1 x_i)^2 - 2(w_0 + w_1 x_i)y_i + y_i^2 \hookrightarrow (2)$$

$\Rightarrow$ from (1) & (2)

$$J(w_0, w_1) = \frac{1}{2}\sum_{i=1}^{5}\Big(\underbrace{w_0^2}_{a} + \underbrace{2w_0 w_1 x_i}_{b} + \underbrace{w_1^2 x_i^2}_{c} - \underbrace{2w_0 y_i}_{d} - \underbrace{2w_1 x_i y_i}_{e} + \underbrace{y_i^2}_{f}\Big) \hookrightarrow (3)$$

$$\Rightarrow a = \sum_{i=1}^{5} w_0^2 = 5w_0^2$$

$$b = \sum_{i=1}^{5} 2 w_0 w_1 x_i = 2 w_0 w_1 \sum_{i=1}^{5} x_i$$

$$= 2 w_0 w_1 (1 + 2 + 4 + 6 + 8)$$

$$= 42 w_0 w_1$$

$$c = \sum_{i=1}^{5} w_1^2 x_i^2 = w_1^2 \sum_{i=1}^{5} x_i^2$$

$$= w_1^2 (1^2 + 2^2 + 4^2 + 6^2 + 8^2) = 121 w_1^2$$

$$d = -2 w_0 \sum_{i=1}^{5} y_i = -2 w_0 (3 + 5 + 9 + 13 + 17)$$

$$= -94 w_0$$

$$e = -2 w_1 \sum_{i=1}^{5} x_i y_i = -2 w_1 (1 \times 3 + 2 \times 5 + 4 \times 9$$

$$+ 6 \times 13 + 8 \times 17) = -526 w_1$$

$$f = \sum_{i=1}^{5} y_i^2 = (3^2 + 5^2 + 9^2 + 13^2 + 17^2)$$

$$= 573$$

from (3) :-

$$J(w_0, w_1) = \frac{1}{2}\left(5w_0^2 + 42w_0w_1 + 121w_1^2 - 94w_0 - 526w_1 + 573\right)$$

$$J(w_0, w_1) = 2.5w_0^2 + 21w_0w_1 + 60.5w_1^2 - 47w_0 - 263w_1 + 286.5 \rightarrow (4)$$

$$\frac{\partial J}{\partial w_0} = 5w_0 + 21w_1 - 47 \longrightarrow (5)$$

$$\frac{\partial J}{\partial w_1} = 21w_0 + 121w_1 - 263 \longrightarrow (6)$$

$$\frac{\partial^2 J}{\partial w_0^2} = 5 ; \frac{\partial^2 J}{\partial w_1^2} = 121, \frac{\partial^2 J}{\partial w_0 \partial w_1} = 21,$$

$$\frac{\partial^2 J}{\partial w_1 \partial w_0} = 21$$

a) $$H = \begin{bmatrix} \dfrac{\partial^2 J}{\partial w_0^2} & \dfrac{\partial^2 J}{\partial w_0 \partial w_1} \\ \dfrac{\partial^2 J}{\partial w_1 \partial w_0} & \dfrac{\partial^2 J}{\partial w_1^2} \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 21 \\ 21 & 121 \end{bmatrix}$$

First leading principal minor $= 5 > 0$

$\det(H) = 5 \times 121 - (21) \times (21)$

$\cancel{= 146}$

$= 164 > 0$

$\therefore$ H is +ve definite

$\Rightarrow$ A strictly convex function requires H to be +ve definite

$\Rightarrow \therefore J(w_0, w_1)$ is strictly convex.

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

(from 5 & 6) $\longrightarrow$ 7

(c) (i) Constant learning rate $= \eta$

$= 0.05,$

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

from 7

$$\nabla J' = \begin{bmatrix} 5 \times 2 + 21 \times 2 - 47 \\ 21 \times 2 + 121 \times 2 - 263 \end{bmatrix} = \begin{bmatrix} 5 \\ 21 \end{bmatrix}$$

$\longrightarrow$ 8

(i) $t=1$, Iteration 1

$$w_0^{(0)} - \eta \frac{\partial J}{\partial w_0} = 2 - 0.05 \times 5 = \underline{1.75}$$

$$w_1^{(1)} = w_1^{(0)} - \eta \frac{\partial J}{\partial w_1} = 2 - 0.05(21)$$

$$= \underline{0.95}$$

After iteration 1:-

$$w_0^{(1)} = 1.75, \quad w_1^{(1)} = 0.95$$

(ii) $t=2$ Iteration 2:-

$$\frac{\partial J}{\partial w_0} = 5w_0 + 21w_1 - 47$$

$$= 5(1.75) + 21(0.95) - 47$$

$$= \boxed{-18.3}$$

$$\frac{\partial J}{\partial w_1} = 21w_0 + 121w_1 - 263$$

$$= 21(1.75) + 121(0.95) - 263$$

$$= \boxed{-111.3}$$

$\therefore \nabla J = \begin{bmatrix} -18.3 \\ -111.3 \end{bmatrix}$

$$w_0^{(2)} = w_0^{(1)} - \eta \frac{\partial J}{\partial w_0}$$

$$w_1^{(2)} = w_1^{(1)} - \eta \frac{\partial J}{\partial w_1}$$

$$w_0^{(2)} = 1.75 - 0.05(-18.3) = \boxed{2.665}$$

(c)

$$w_1^{(2)} = 0.95 - 0.05(-111.3) = \boxed{6.515}$$

$\Rightarrow$ After Iteration 2,

$$\underline{w_0^{(2)} = 2.665} \quad \& \quad \underline{w_1^{(2)} = 6.515}$$

2. $\eta_t = \eta_0 e^{-kt}$, $\eta_0 = 0.1$, $k = 0.4$,

$t = 1, 2, 3, \ldots$

$w_0^{(0)} = 2, \; w_1^{(0)} = 2$

(i) Iteration 1 :-

$t=1$:-

$$\eta_1 = \eta_0 e^{-0.4 \times 1} = 0.067$$

$$\nabla J = \begin{bmatrix} 5 \\ 21 \end{bmatrix} \quad \text{[from ⑧]}.$$

(c) 2. (i) $w_0^{(1)} = w_0^{(0)} - \eta_1 \frac{\partial J}{\partial w_0}$

$= 2 - 0.067 \times 5$

$= 1.665$

$w_1^{(1)} = w_1^{(0)} - \eta \frac{\partial J}{\partial w_1}$

$= 2 - 0.067 \times 21$

$= 0.593$

$\therefore w_0^{(1)} = 1.665, \; w_1^{(1)} = 0.593$ after iteration 1

(ii) Iteration 2, $t=2$:-

$$\eta_2 = \eta_0 e^{-0.4\times 2} = 0.1 e^{-0.8}$$

$$\boxed{\eta_2 = 0.0449}$$

$$\Rightarrow \frac{\partial J}{\partial w_0} = 5w_0 + 21w_1 - 47$$

$$= 5(1.665) + 21(0.593) - 47$$

$$= \underline{-26.222}$$

$$\Rightarrow \frac{\partial J}{\partial w_1} = 21w_0 + 121w_1 - 263$$

$$= 21(1.665) + 121(0.593) - 263$$

$$= \underline{-156.282}$$

$$\Rightarrow w_0^{(2)} = w_0^{(1)} - \eta \frac{\partial J}{\partial w_0}$$

$$= 1.665 - 0.0449(-26.222)$$

$$= \underline{2.8424}$$

$\Rightarrow w$

(c)

$$\Rightarrow w_1^{(2)} = w_1^{(1)} - \eta_2 \frac{\partial J}{\partial w_1}$$

$$= 0.593 - 0.0449\,(-156.282)$$

$$= \underline{7.6101}$$

$\therefore$ Iteration 2, $\underline{w_0^{(2)} = 2.8424}$

& $\underline{w_1^{(2)} = 7.6101}$

(c) $w_0^{(1)} = 1.75$, $w_0^{(1)} = 0.05$

Constant ($\eta = 0.05$)

$w_0^{(2)} = 2.665$, $w_1^{(2)} = 6.515$

Decaying η [starts with $\eta = 0.1$]

$w_0^{(1)} = 1.665$, $w_1^{(1)} = 0.593$

$w_0^{(2)} = 2.8424$, $w_1^{(2)} = 7.6101$

From input table :-

$y = 1 + 2x$ is optimal solution

~~$q_1 = 0.067$~~

~~$\lambda_2 = 0.0449$~~

$$H = \begin{bmatrix} 5 & 21 \\ 21 & 121 \end{bmatrix}$$

$\det(A - \lambda I) = 0$

$\therefore$ ~~5~~ $(5-\lambda)(121-\lambda) - 21 \times 21 = 0$

$\therefore 605 - 5\lambda - 121\lambda + \lambda^2 - 441 = 0$

$\therefore \lambda^2 - 126\lambda + 164 = 0$

$a = 1, \; b = -126, \; c = 164$

$$\Rightarrow \lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$\therefore \lambda = \frac{-(-126) \pm \sqrt{(126)^2 - 4(1)(164)}}{2 \times 1}$$

$$\lambda = \frac{126 \pm \sqrt{15220}}{2} = 124.6847, \; 1.3153$$

$$\boxed{\lambda_{max} = 124.6847}$$

$$\frac{2}{\lambda_{max}} = 0.016, \text{ for stability } \eta < \lambda_{max},$$

i.e, $\eta < 0.016$

$\Rightarrow$ In (.1 & (.2 $\eta = 0.05$

& $\eta_1 = 0.067$ & $\eta_2 = 0.0749$) $\nless 0.016$

$\Rightarrow$ while η improves convergence, initial η was aggressive causing overshooting & instability.

$\therefore$ Past c $\because$ decay η didn't improve stability & convergence behaviour compared to constant η.

$\Rightarrow$ It infact made first iteration update larger then constant η therefore overshoot risk

$\Rightarrow$ Also there is a larger ~~gar~~ gradient leading to instability.

$\Rightarrow$ Both runs produced large jump in parameters

(e) (i) from (4)

$$J(w_0, w_1) = 2.5 w_0^2 + 21 w_0 w_1 + 60.5 w_1^2 - 47 w_0 - 263 w_1 + 286.5$$

$w_0^{(0)} = 2$, $w_1^{(0)} = 2$ from 2.(i) $\nabla J = \begin{bmatrix} 5 \\ 21 \end{bmatrix}$

$$d = -\nabla J = \begin{bmatrix} -5 \\ -21 \end{bmatrix}$$

$\phi(\eta) = J(w + \eta d)$; $w_0(\eta) = w_0^{(0)} - \eta \frac{\partial J}{\partial w_0}$

$$= 2 - \eta(5) = 2 - 5\eta$$

$\Rightarrow w_1(\eta) = w_1^{(0)} - \eta \frac{\partial J}{\partial w_1} = 2 - \eta(21)$

$$= 2 - 21\eta$$

Binary Search

Let current interval be $[a, b]$

& $m = \frac{a+b}{2}$

$\Rightarrow$ $\phi(m) < \phi(m+\epsilon) \Rightarrow$ fn. is $\uparrow$ at 'm' (slope > 0)

min. = smaller ?

$\Rightarrow$ $\phi(m) > \phi(m+\varepsilon) \Rightarrow$ fn. is $\downarrow$ at 'm' (slope < 0)

min = larger ?

$\Rightarrow$ Iteration 1 :-

$[0, 1] \Rightarrow m = \frac{0+1}{2} = 0.5$

$\epsilon = 10^{-3} = 0.001$

$0.5 + \epsilon = 0.501$

$\Rightarrow \phi(0.5) = J(2-5\eta, 2-21\eta)$

$= J(2-5(0.5), 2-21(0.5))$

$= J(-0.5, -8.5)$

$= 2.5(0.5)^2 + 21(-0.5)(-8.5)$

$+ 60.5(-8.5)^2 - 47(-0.5)$

$- 263(-8.5) + 286.5$

$\therefore \underline{J(-0.5, -8.5) = \boxed{7006.5}}$

$\Rightarrow \phi(0.501) = J(2-5\times 0.501, 2-21\times 0.501)$

$= J(-0.505, -8.521)$

$= 2.5(-0.505)^2 + 21(-0.505)(-8.521)$

$+ 60.5(-8.541)^2 - 47(-0.505)$

$-263(-8.521) + 286.5$

$\therefore \underline{J(-0.505, -8.521)} = \boxed{7035.0109}$

Iteration 2 :-

$[0, 0.5] \Rightarrow m = \frac{0+0.5}{2} = 0.25$

$m + \epsilon = 0.25 + 0.001 = 0.251$

$\phi(0.25) = J(2 - 5 \times 0.25, 2 - 21 \times 0.25)$

$= J(0.75, -3.25)$

$= 2.5(0.75)^2 + 21(0.75)(-3.25)$
$+ 60.5(3.25)^2 - 47(0.75)$
$- 263(-3.25) + 286.5$

$\phi(0.25) = 1695.25 \longrightarrow (9)$

$\phi(0.251) = \phi(0.251) = J(2.5 \times 0.251,$
$2 - 21 \times 0.251)$

$= J(0.745, -2.371)$

$= \boxed{1709.2869}$

ϕ (m

f

(ii)

ϕ (

ϕ (

=

$\phi(m) < \phi(m+\epsilon)$, again,

Reduced interval $= [0, 0.25]$

fn is $\uparrow$ at M_2, min lies in left.

(ii) $[a, b] = [0, 1]$, $M_1 = 1/4 = 0.25$,

$M_2 = 3/4 = 0.75$

$\phi(0.25) = J(2 - 5 \times 0.25, 2 - 21 \times 0.25)$

$= J(0.75, -3.25) = \underline{1695.25}$

$\hookrightarrow$ From (9)

$\phi(0.75) = J(2 - 5 \times 0.75, 2 - 21 \times 0.75)$

$= J(-1.75, -13.75)$

$= 2.5(-1.75)^2 + 21(-1.75)(-13.75) + 60.5(-13.75)^2 - 47(-1.75) - 263(-13.75) + 286.5 = \underline{15936.25}$

$\therefore \; \phi(m_1) < \phi(m_2)$

$\therefore$ min. in $[a, m_2]$ i.e,

$$[0, 0.75]$$

(e) (ii) L = Interval width $= b - a$

$$= 0.75 - 0$$

$$= 0.75$$

→ golden ratio $\Rightarrow \tau = \frac{1+\sqrt{5}}{2} = 0.618$

$$m_1 = a + (1 - \tau) \times L = 0 + (1 - 0.618) \times 0.75$$

$$= \underline{0.2865}$$

$$m_2 = 0 + \tau L = 0 + 0.618 \times 0.75$$

$$= \underline{0.4635}$$

∴ one valid pair

(0.2865, 0.4635) using golden ratio.

→ Another option using specific ratio logic :-

$m_1 = 0.25 \times 0.75 = 0.1875$ ($\frac{1}{4}$th of way)

$m_2 = 0.75 \times 0.75 = 0.5625$

($\frac{3}{4}$th of way)