Name - Sagar Powar
BITS ID - 2025AA05421

MFML - Assignment 2

Q. a) Given loss function $J(w_0, w_1) = \frac{1}{2}\sum_{i=1}^{5} (\hat{y}_i - y_i)^2$

substitute each datapoint value, we get.

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

explicit squared-error loss function.

The Hessian matrix H contains the second-order partial derivatives.

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

① First order derivatives

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

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

② second order derivatives

$$\frac{\partial^2 J}{\partial w_0^2} = \sum_{i=1}^{5} 1 = 5$$

$$\frac{\partial^2 J}{\partial w_1^2} = \sum_{i=1}^{5} x_i^2 = 1^2 + 2^2 + 4^2 + 6^2 + 8^2 = 121$$

$$\frac{\partial^2 J}{\partial w_0 \partial w_1} = \sum_{i=1}^{5} x_i = 1 + 2 + 4 + 6 + 8 = 21$$

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

Name - Sagar Powar
BITS ID - 2025AA05421

MFML - Assignment - 2

a) Convexity proof

- A function is convex if its Hessian is positive-semi definite
- for a 2x2 matrix, check below conditions

① $H_{11} > 0$
② $\det(H) \geq 0$

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

→ $H_{11} = 5 > 0$ — (I)
→ $\det(H) = (5 \times 121) - (21)^2 = 605 - 441 = 164 > 0$ — (II)

As both (I) & (II) are satisfied, the squared-loss fun.

$\boxed{J(w_0, w_1) \text{ is a convex-function}}$

b) Gradient of function $\nabla J = \begin{bmatrix} \frac{\partial J}{\partial w_0} \\ \frac{\partial J}{\partial w_1} \end{bmatrix}$

①
$$\frac{\partial J}{\partial w_0} = \sum_{i=1}^{5} (w_0 + w_1 x_i - y_i)$$
$$= (w_0 + w_1 - 3) + (w_0 + 2w_1 - 5) + (w_0 + 4w_1 - 9) + (w_0 + 6w_1 - 13) + (w_0 + 8w_1 - 17)$$

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

②
$$\frac{\partial J}{\partial w_1} = \sum_{i=1}^{5} (w_0 + w_1 x_i - y_i) x_i$$
$$= (w_0 + w_1 - 3)(1) + (w_0 + 2w_1 - 5)2 + (w_0 + 4w_1 - 9)4 + (w_0 + 6w_1 - 13)6 + (w_0 + 8w_1 - 17)8$$

$$\frac{\partial J}{\partial w_1} = 21w_0 + 121w_1 - \cancel{259}\ 263$$

$$\nabla J = \begin{bmatrix} 5w_0 + 21w_1 - 47 \\ 21w_0 + 121w_1 - \cancel{259}\ 263 \end{bmatrix}$$

Name - Sagar powar
BITS ID - 2025AA05421

MFML - Assignment 2

Q. c) Given, initial parameters. $w_0^{(0)} = 2$ $w_1^{(0)} = 2$

Gradient Descent update Rule,

$$\boxed{w^{(t+1)} = w^{(t)} - \eta \nabla J(w^{(t)})}$$

$\eta = 0.05$ (learning rate)

Iteration 1

i) Find gradient at initial point, $w_0 = 2$, $w_1 = 2$

$$\frac{\partial J}{\partial w_0} = 5w_0 + 21w_1 - 47 = 5(2) + 21(2) - 47 = 5$$

$$\frac{\partial J}{\partial w_1} = 21w_0 + 121w_1 - 263 = 21(2) + 121(2) - 263 = 21$$

$$\nabla J(2,2) = \begin{bmatrix} 5 \\ \cancel{25} \end{bmatrix} = \begin{bmatrix} 5 \\ 21 \end{bmatrix}$$

ii) update parameters,

$$w_0^{(1)} = w^{(0)} - \eta \nabla J(w_0^{(0)}) = 2 - (0.05)(5) = \boxed{1.75}$$

$$w_1^{(1)} = 2 - (0.05)(21) = 2 - 1.05 = \boxed{0.95}$$

Iteration 2

i) Find gradients at new point, $w_0 = 1.75$ $w_1 = 0.95$

$$\frac{\partial J}{\partial w_0} = 5(1.75) + 21(0.95) - 47 = 8.75 + 19.95 - 47 = \underline{-18.3}$$

$$\frac{\partial J}{\partial w_1} = 21(1.75) + 121(0.95) - 263 = 36.75 + 114.95 - 263 = \underline{-111.3}$$

$$\nabla J(1.75, 0.95) = \begin{bmatrix} -18.3 \\ -111.3 \end{bmatrix}$$

ii) update parameters

$$w_0^{(2)} = 1.75 - 0.05(-18.3) = 1.75 + 0.915 = \boxed{2.665}$$

$$w_1^{(2)} = 0.95 - 0.05(-111.3) = 0.95 + 5.565 = \boxed{6.515}$$

Name - sagar powar
BITS ID - 2025AA05421

MFML - Assignment - 2

Q c) 2) Exponentially decaying learning rate

$\eta_t = \eta_0 e^{-kt}$ $\quad \cancel{\eta_0 = 0.1}$ $\eta_0 = 0.1$, $k = 0.4$

$\eta_1 = 0.1 e^{-0.4} \approx 0.1(0.6703) = \boxed{0.06703}$

$\eta_2 = 0.1 e^{-0.8} \approx 0.1(0.4493) = \boxed{0.04493}$

1) Iteration - I

i) Gradient at $(2,2)$. $\nabla J(2,2) = \begin{bmatrix} 5 \\ 21 \end{bmatrix}$

ii) Updates,

$w_0^{(1)} = 2 - 0.06703(5) = 2 - 0.33515 = \boxed{1.66485}$

$w_1^{(1)} = 2 - 0.06703(21) = 2 - 1.40763 = \boxed{0.59237}$

Iteration II

i) Find gradients at 2nd point

$\frac{\partial J}{\partial w_0} = 5(1.66485) + 21(0.59237) - 47$
$= 8.32425 + 12.43977 - 47 = \underline{-26.23598}$

$\frac{\partial J}{\partial w_1} = 21(1.66485) + 121(0.59237) - 263$
$= 34.96185 + 71.67677 - 263 = \underline{-156.36138}$

ii) Update parameters

$w_0^{(2)} = 1.66485 - 0.04493(-26.23598) \approx \boxed{2.84374}$

$w_1^{(2)} = 0.59237 - 0.04493(-156.36138) \approx \boxed{7.61837}$

Name - Sagar Powar | MFML - Assigment 2 | 5
BITS ID - 2025AA05421

d) Comparing two methods from (c):

(I) Constant learning rate ($\eta = 0.05$) - The parameters moved from (2,2) to [2.665, 6.515].
- Gradient increased significantly, indicates fixed step size might be too large
- This leads to oscillation or overshooting global minimum.

(II) Exponential Decay -
- This method starts with higher learning rate (0.067) to make large initial progress & then reduced it.

Conclusion
1. Exponential decay improves stability and
2. convergence compared to a constant learning rate.

e) Step-size selection

Given, $d = -\nabla J(w)$ & 1-D function $\phi_\eta = J(w + \eta d)$

Goal minimize ϕ_η over [0, 1]

I) Binary search method

1. initial interval [0, 1], midpoint = 0.5
- Find loss at midpoint ($\eta = 0.5$)
- $w_0 = 2 - 5(0.5) = \cancel{0.5} - 0.5$
- $w_1 = 2 - 21(0.5) = \cancel{-08.6} - 8.5$

- Find loss at m,

$$J = \frac{1}{2}\left[(-0.5 + 1(8.5) - 3)^2 + (-0.5 + 2(-8.5) - 5)^2 + (-0.5 + 4(-8.5) - 9)^2 + (-0.5 + 6(-8.5) - 13)^2 + (-0.5 + 8(-8) - 17)^2\right]$$

e) $J = \frac{1}{2}[(-12)^2 + (-22.5)^2 + (43.5)^2 + (-64.5)^2 + (-85.5)^2]$

$= \frac{1}{2}[144 + 506.25 + 1892.25 + 4160.25 + 7310.25]$

$= \frac{1}{2}[14013] = \boxed{7006.5}$ — ①

② Calculate loss at perturbed point ($\eta = 0.501$).

$w_0 = 2 - 5(0.501) = -0.505$

$w_1 = 2 - 21(0.501) = -8.521$

$J = \frac{1}{2}[(-0.505 + 1(-8.521) - 3)^2 + (-0.505 + 2(-8.521) - 5)^2 + (-0.505 + 4(-8.521) - 9)^2 + (-0.505 + 6(-8.521) - 13)^2 + (-0.505 + 8(-8.521) - 17)^2]$

$= \frac{1}{2}[144.62 + 508.37 + 1900 + 4177.16 + 7339.86]$

$= \frac{1}{2}(14070.01) = \boxed{7035.0}$ — ②

As $\boxed{7006.5 < 7035.0}$, New interval $= [0, 0.5]$

① Iteration - 2 : midpoint $= 0.25$ & interval $[0, 0.5]$

weights, $w_0 = 2 - 5(0.25) = 0.75$

$w_1 = 2 - 21(0.25) = -3.25$

find loss,

$J = \frac{1}{2}[(0.75 + (-3.25) - 3)^2 + (0.75 + 2(-3.25) - 5)^2 + (0.75 + 4(-3.25) - 9)^2 + (0.75 + 6(-3.25) - 13)^2 + (0.75 + 8(-3.25) - 17)^2]$

$= \frac{1}{2}[30.25 + 115.56 + 451.56 + 1008.06 + 1785.06]$ — ③

$= \frac{1}{2}(3390.5) = \boxed{1695.25}$

$\eta = 0.251 = 0.25 + 0.001$

find loss at perturbed point

$w_0 = 2 - 5(0.251) = 0.745$

$w_1 = 2 - 21(0.251) = -3.271$

Name - sagar powar
Bits ID - 2023AA05421

Find loss,

$$J = \frac{1}{2}\Big[(0.745+(-3.271)-3)^2 + (0.745+(-3.271)2-5)^2 + (0.745+4(-3.271)-9)^2 + (0.745+6(-3.271)-13)^2 + (0.745+8(-3.271)-17)^2\Big]$$

$$J = \frac{1}{2}[30.54+116.58+455.35+1016.39+1299.71]$$

$$= \frac{1}{2}(3418.57) = \boxed{1709.28} \text{ — (4)}$$

As, per (3) & (4) ~~1695~~ $\boxed{1695.25 < 1709.28}$, we are far from right

New interval $[0, 0.25]$

e) II) $\boxed{\text{Golden-Section Search}}$

- Find two points m_1 (at 25% distance) & m_2 (at 75%) distance
  Keep section containing lower loss.

$\boxed{\text{Iteration-I}}$ interval $[0, 1]$ $m_1 = 0.25$ $m_2 = 0.75$

(A) Find loss at m_1 $(\eta = 0.25)$

$\boxed{J(m_1) = 1695.25}$ (from (3)) — (5)

(B) Find loss at m_2 $(\eta = 0.75)$

$w_0 = 2 - 5(0.75) = -1.75$

$w_1 = 2 - 21(0.75) = -13.75$

$$J = \frac{1}{2}\Big[((-1.75)+(-13.75)-3)^2 + (-1.75+2(-13.75)-5)^2 + (-1.75+4(-13.75)-9)^2 + (-1.75+6(-13.75)-13)^2 + (-1.75+8(-13.75)-17)^2\Big]$$

$$= \frac{1}{2}[342.25+1173.06+4323.06+9457.56+16576.56]$$

$$= \frac{1}{2}(31872.5) = \boxed{15936.25} \text{ — (6)}$$

Decision $J(m_1) < J(m_2)$, the minimum is in left section.

New interval $[0, 0.75]$

Name - sagar Powar
Bits ID - 2025AA05421

MFML-2

e **Iteration: 2**

$m_1 = 0 + 0.25(0.75) = 0.1875$

$m_2 = 0 + 0.75(0.75) = 0.5625$

(A) Evaluate m_1 ($\eta = 0.1875$)

$w_0 = 2 - 5(0.1875) = 1.0625$

$w_1 = 2 - 21(0.1875) = -1.9375$

$$J = \frac{1}{2}\left[(1.06 - 1.94 - 3)^2 + (1.06 - 3.88 - 5)^2 + (1.06 - 7.75 - 9)^2 + (1.06 - 11.62 - 13)^2 + (1.06 - 15.5 - 17)^2\right]$$

$$= \frac{1}{2}\left[15.01 + 61.03 + 246.1 + 551.19 + 988.31\right] =$$

$$J = \frac{1}{2}(1865.64) = \boxed{932.82} \text{ — } (7)$$

(B) Evaluate m_2 ($\eta = 0.5625$)

$w_0 = 2 - 5(0.5625) = -0.8125$

$w_1 = 2 - 21(0.5625) = -9.8125$

Find loss,

$$J(m_2) = \frac{1}{2}\left[(-0.81 - \overset{9.81}{\cancel{9.81}} - 3)^2 + (-0.81 - 19.62 - 5)^2 + (-0.81 - 39.25 - 9)^2 + (-0.81 - 58.88 - 13)^2 + (-0.81 - 78.5 - 17)^2\right]$$

$$J = \frac{1}{2}\left[185.64 + 647.06 + 2407.13 + 5283.47 + 9276.10\right]$$

$$= \frac{1}{2}\left[17799.4\right] = \boxed{8899.7} \text{ — } (8)$$

Decision: $J(m_1) < J(m_2)$. The minimum is in left section

New interval $= [0, 0.5625]$