Solution 1 (a)

Step 1

Let w be the loss function on vectors $w \in \mathbb{R}^4$:

$$L(w) = w_1^2 + 2w_2^2 + w_3^2 - 2w_3w_4 + w_4^2 + 2w_1 - 4w_2 + 4$$

Step 2

Calculate $\frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \frac{\partial L}{\partial w_3}, \frac{\partial L}{\partial w_4}$

$$\begin{aligned}
\frac{\partial L}{\partial w_1} &= \left(w_1^2 + 2w_2^2 + w_3^2 - 2w_3w_4 + w_4^2 + 2w_1 - 4w_2 + 4\right)\frac{\partial}{\partial w_1} = & 2w_1 + 2 \\
\frac{\partial L}{\partial w_2} &= \left(w_1^2 + 2w_2^2 + w_3^2 - 2w_3w_4 + w_4^2 + 2w_1 - 4w_2 + 4\right)\frac{\partial}{\partial w_2} = & 4w_2 - 4 \\
\frac{\partial L}{\partial w_3} &= \left(w_1^2 + 2w_2^2 + w_3^2 - 2w_3w_4 + w_4^2 + 2w_1 - 4w_2 + 4\right)\frac{\partial}{\partial w_3} = & 2w_3 - 2w_4 \\
\frac{\partial L}{\partial w_4} &= \left(w_1^2 + 2w_2^2 + w_3^2 - 2w_3w_4 + w_4^2 + 2w_1 - 4w_2 + 4\right)\frac{\partial}{\partial w_4} = & -2w_3 + 2w_4
\end{aligned}$$

$\therefore$ The partial derivatives with respect to w_1, w_2, w_3, w_4, are:

$$\frac{\partial L}{\partial w_1} = 2w_1 + 2 \qquad \frac{\partial L}{\partial w_2} = 4w_2 - 4 \qquad \frac{\partial L}{\partial w_3} = 2w_3 - 2w_4 \qquad \frac{\partial L}{\partial w_4} = -2w_3 + 2w_4$$

Solution 1 (b)

Step 1

The gradient of $L(w)$ is the vector of partial derivatives:

$$\nabla L(w) = \begin{pmatrix} \frac{\partial L}{\partial w_1} \\ \vdots \\ \frac{\partial L}{\partial w_n} \end{pmatrix}$$

Step 2

Subsitute partial derivatives from **Solution 1 (a)**.

$$\frac{\partial L}{\partial w_1} = 2w_1 + 2 \qquad \frac{\partial L}{\partial w_2} = 4w_2 - 4 \qquad \frac{\partial L}{\partial w_3} = 2w_3 - 2w_4 \qquad \frac{\partial L}{\partial w_4} = -2w_3 + 2w_4$$

$$\nabla L(w) = \begin{pmatrix} \frac{\partial L}{\partial w_1} \\ \frac{\partial L}{\partial w_2} \\ \frac{\partial L}{\partial w_3} \\ \frac{\partial L}{\partial w_4} \end{pmatrix} = \begin{pmatrix} 2w_1 + 2 \\ 4w_2 - 4 \\ 2w_3 - 2w_4 \\ -2w_3 + 2w_4 \end{pmatrix}$$

$\therefore$ The gradient of $L(w)$ is:

$$\nabla L(w) = \begin{pmatrix} 2w_1 + 2 \\ 4w_2 - 4 \\ 2w_3 - 2w_4 \\ -2w_3 + 2w_4 \end{pmatrix}$$

Solution 1 (c)

Step 1

The Gradient Descent Update is defined as:

$$w_{t+1} = w_t - \eta \nabla L(w_t)$$

Step 2

Calculate Gradient Descent Update:

Let $t = 0$, $\eta = 0.1$, $w_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, and $\nabla L(w) = \begin{pmatrix} 2w_1 + 2 \\ 4w_2 - 4 \\ 2w_3 - 2w_4 \\ -2w_3 + 2w_4 \end{pmatrix}$

$$w_1 = w_0 - \eta \nabla L(w_0) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} - 0.1 \cdot \begin{pmatrix} 2 \cdot 0 + 2 \\ 4 \cdot 0 - 4 \\ 2 \cdot 0 - 2 \cdot 0 \\ -2 \cdot 0 + 2 \cdot 0 \end{pmatrix} = 0.1 \cdot \begin{pmatrix} 2 \\ -4 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.2 \\ -0.4 \\ 0 \\ 0 \end{pmatrix}$$

$\therefore$ The next estimate is:

$$w_1 = \begin{pmatrix} 0.2 \\ -0.4 \\ 0 \\ 0 \end{pmatrix}$$

Solution 1 (d)

Step 1

Set the gradient to zero to find the minimizer:

$$\begin{aligned}
2w_1 + 2 = 0 &\implies w_1^* = -1 \\
4w_2 - 4 = 0 &\implies w_2^* = 1 \\
2w_3 - 2w_4 = 0 &\implies w_3^* = w_4^* \\
-2w_3 + 2w_4 = 0 &\implies w_3^* = w_4^*
\end{aligned}$$

Let $w_3^* = w_4^* = \alpha$.

Step 2

Plug into $L(w)$:

$$\begin{aligned}
L(w^*) &= (w_1^*)^2 + 2(w_2^*)^2 + (w_3^*)^2 - 2w_3^* w_4^* + (w_4^*)^2 + 2w_1^* - 4w_2^* + 4 \\
&= (-1)^2 + 2(1)^2 + \alpha^2 - 2\alpha^2 + \alpha^2 + 2(-1) - 4(1) + 4 \\
&= 1 + 2 + \alpha^2 - 2\alpha^2 + \alpha^2 - 2 - 4 + 4 \\
&= 3 + (0) - 2 - 4 + 4 \\
&= 1
\end{aligned}$$

$\therefore$ The minimum value is $L(w^*) = 1$.

Solution 1 (e)

Step 1

From part (d), the minimizer is:

$$w^* = (-1,\, 1,\, \alpha,\, \alpha)$$

for any $\alpha \in \mathbb{R}$.

$\therefore$ The loss function is strictly convex in w_1 and w_2, but only jointly convex in w_3 and w_4 along the direction $w_3 = w_4$. The Hessian has a zero eigenvalue corresponding to the direction $(0, 0, 1, -1)$, so the minimum is not unique. There is a **line** of minimizers parameterized by α.
