

By-Hand Training of a 1-Hidden-Layer ReLU Network to Fit $y = x^2 + 2x$

Worksheet + Fully Worked Answer Key (rounded to 2 decimals; full arithmetic shown)

Learning rate $\alpha = 0.05$; Training inputs $x \in \{-2, -1, 0, 1, 2\}$

Problem Setup (shared)

We fit a tiny neural network with a single hidden layer (2 units) and ReLU activation to the target $y = x^2 + 2x$ using full-batch gradient descent.

Network:

$$\text{Hidden (2 units):} \quad z_1 = W_1x + b_1, \quad h = \text{ReLU}(z_1)$$

$$\text{Output (scalar):} \quad \hat{y} = W_2h + b_2$$

Parameters: $W_1 = \begin{bmatrix} w_{11} \\ w_{21} \end{bmatrix}$, $b_1 = \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix}$, $W_2 = [w_{2,1} \quad w_{2,2}]$, $b_2 \in \mathbb{R}$.

Loss: $L = \frac{1}{2} \sum_s (y_s - \hat{y}_s)^2$. We take $\text{ReLU}'(0) = 0$.

Part A. Student Worksheet

This section provides tables where you will explicitly compute values, rounding to **2 decimals**.

A1. Parameter Initialization

Suggested start:

$$W_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad b_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad W_2 = [1 \quad -1], \quad b_2 = 0.$$

Write these in the table below (fill in new values each iteration after updates):

Iteration	W_1	b_1	W_2, b_2
0	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$[1, -1], 0$
1			
2			
3			
4			
5			
6			

A2. Forward Pass Table

For each $x \in \{-2, -1, 0, 1, 2\}$, fill in values:

x	$y = x^2 + 2x$	$z_1 = W_1x + b_1$	$h = \text{ReLU}(z_1)$	$\hat{y} = W_2h + b_2$	$e = \hat{y} - y$
-2					
-1					
0					
1					
2					

A3. Gradient Computations

For each sample, compute step by step:

- $\partial W_2|_s = e h^\top$ (show each multiplication)
- $\partial b_2|_s = e$
- eW_2 , then $\delta^{(h)} = (eW_2) \odot 1_{z_1 > 0}$
- $\partial W_1|_s = \delta^{(h)}x$, $\partial b_1|_s = \delta^{(h)}$

Sum across samples to get batch gradients.

A4. Update

Show $\theta \leftarrow \theta - 0.05\nabla L$, substitution, and rounded result.

Part B. Answer Key (fully detailed; all numbers rounded to 2 decimals)

We use the suggested starting values. Every multiplication/addition shown below is rounded to 2 decimals *at the time it is shown*, and these rounded values are propagated to the subsequent displayed steps.

Iteration 0: forward, gradients, update to Iteration 1

Forward pass (parameters at start of Iteration 0). Loss $L = 22.00$.

x	y	z_1	h	\hat{y}	e
-2	0.00	[-2.00, 2.00]	[0.00, 2.00]	-2.00	-2.00
-1	-1.00	[-1.00, 1.00]	[0.00, 1.00]	-1.00	0.00
0	0.00	[0.00, 0.00]	[0.00, 0.00]	0.00	0.00
1	3.00	[1.00, -1.00]	[1.00, 0.00]	1.00	-2.00
2	8.00	[2.00, -2.00]	[2.00, 0.00]	2.00	-6.00

Per-sample gradients (no abbreviation; 2-dec rounding at each product):

- $x = -2$: $e = -2.00$, $h = [0.00, 2.00]$, $1_{z>0} = [0, 1]$.

$$\begin{aligned}\partial W_2|_s &= e h = [-2.00 \cdot 0.00, -2.00 \cdot 2.00] = [0.00, -4.00], \\ \partial b_2|_s &= e = -2.00, \\ eW_2 &= -2.00[1.00, -1.00] = [-2.00, 2.00], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [-2.00 \cdot 0, 2.00 \cdot 1] = [0.00, 2.00], \\ \partial W_1|_s &= \delta^{(h)} x = [0.00 \cdot -2, 2.00 \cdot -2] = [0.00, -4.00], \\ \partial b_1|_s &= \delta^{(h)} = [0.00, 2.00].\end{aligned}$$

- $x = -1$: $e = 0.00$, $h = [0.00, 1.00]$, $1_{z>0} = [0, 1]$.

$$\begin{aligned}\partial W_2|_s &= e h = [0.00 \cdot 0.00, 0.00 \cdot 1.00] = [0.00, 0.00], \\ \partial b_2|_s &= e = 0.00, \\ eW_2 &= 0.00[1.00, -1.00] = [0.00, 0.00], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [0.00 \cdot 0, 0.00 \cdot 1] = [0.00, 0.00], \\ \partial W_1|_s &= \delta^{(h)} x = [0.00 \cdot -1, 0.00 \cdot -1] = [0.00, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [0.00, 0.00].\end{aligned}$$

- $x = 0$: $e = 0.00$, $h = [0.00, 0.00]$, $1_{z>0} = [0, 0]$.

$$\begin{aligned}\partial W_2|_s &= e h = [0.00 \cdot 0.00, 0.00 \cdot 0.00] = [0.00, 0.00], \\ \partial b_2|_s &= e = 0.00, \\ eW_2 &= 0.00[1.00, -1.00] = [0.00, 0.00], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [0.00 \cdot 0, 0.00 \cdot 0] = [0.00, 0.00], \\ \partial W_1|_s &= \delta^{(h)} x = [0.00 \cdot 0, 0.00 \cdot 0] = [0.00, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [0.00, 0.00].\end{aligned}$$

- $x = 1$: $e = -2.00$, $h = [1.00, 0.00]$, $1_{z>0} = [1, 0]$.

$$\begin{aligned}\partial W_2|_s &= e h = [-2.00 \cdot 1.00, -2.00 \cdot 0.00] = [-2.00, 0.00], \\ \partial b_2|_s &= e = -2.00, \\ eW_2 &= -2.00[1.00, -1.00] = [-2.00, 2.00], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [-2.00 \cdot 1, 2.00 \cdot 0] = [-2.00, 0.00], \\ \partial W_1|_s &= \delta^{(h)} x = [-2.00 \cdot 1, 0.00 \cdot 1] = [-2.00, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [-2.00, 0.00].\end{aligned}$$

- $x = 2$: $e = -6.00$, $h = [2.00, 0.00]$, $1_{z>0} = [1, 0]$.

$$\partial W_2|_s = e h = [-6.00 \cdot 2.00, -6.00 \cdot 0.00] = [-12.00, 0.00],$$

$$\partial b_2|_s = e = -6.00,$$

$$e W_2 = -6.00[1.00, -1.00] = [-6.00, 6.00],$$

$$\delta^{(h)} = e W_2 \odot 1_{z>0} = [-6.00 \cdot 1, 6.00 \cdot 0] = [-6.00, 0.00],$$

$$\partial W_1|_s = \delta^{(h)} x = [-6.00 \cdot 2, 0.00 \cdot 2] = [-12.00, 0.00],$$

$$\partial b_1|_s = \delta^{(h)} = [-6.00, 0.00].$$

Batch sums:

$$\partial W_2 = [-14.00, -4.00], \quad \partial b_2 = -10.00, \quad \partial W_1 = [-14.00, -4.00], \quad \partial b_1 = [-8.00, 2.00].$$

Update ($\alpha = 0.05$):

$$W_2^+ = [1.00, -1.00] - 0.05[-14.00, -4.00] = [1.00, -1.00] - [-0.70, -0.20] = [1.70, -0.80],$$

$$b_2^+ = 0.00 - 0.05 \cdot (-10.00) = 0.00 - -0.50 = 0.50,$$

$$W_1^+ = \begin{bmatrix} 1.00 \\ -1.00 \end{bmatrix} - 0.05 \begin{bmatrix} -14.00 \\ -4.00 \end{bmatrix} = \begin{bmatrix} 1.00 \\ -1.00 \end{bmatrix} - \begin{bmatrix} -0.70 \\ -0.20 \end{bmatrix} = \begin{bmatrix} 1.70 \\ -0.80 \end{bmatrix},$$

$$b_1^+ = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} - 0.05 \begin{bmatrix} -8.00 \\ 2.00 \end{bmatrix} = \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix} - \begin{bmatrix} -0.40 \\ 0.10 \end{bmatrix} = \begin{bmatrix} 0.40 \\ -0.10 \end{bmatrix}.$$

Iteration 1: forward, gradients, update to Iteration 2

Forward pass (parameters at start of Iteration 1). Loss $L = 2.50$.

x	y	z_1	h	\hat{y}	e
-2	0.00	[-3.00, 1.50]	[0.00, 1.50]	-0.70	-0.70
-1	-1.00	[-1.30, 0.70]	[0.00, 0.70]	-0.06	0.94
0	0.00	[0.40, -0.10]	[0.40, 0.00]	1.18	1.18
1	3.00	[2.10, -0.90]	[2.10, 0.00]	4.07	1.07
2	8.00	[3.80, -1.70]	[3.80, 0.00]	6.96	-1.04

Per-sample gradients (no abbreviation; 2-dec rounding at each product):

- $x = -2$: $e = -0.70$, $h = [0.00, 1.50]$, $1_{z>0} = [0, 1]$.

$$\partial W_2|_s = e h = [-0.70 \cdot 0.00, -0.70 \cdot 1.50] = [0.00, -1.05],$$

$$\partial b_2|_s = e = -0.70,$$

$$e W_2 = -0.70[1.70, -0.80] = [-1.19, 0.56],$$

$$\delta^{(h)} = e W_2 \odot 1_{z>0} = [-1.19 \cdot 0, 0.56 \cdot 1] = [0.00, 0.56],$$

$$\partial W_1|_s = \delta^{(h)} x = [0.00 \cdot -2, 0.56 \cdot -2] = [0.00, -1.12],$$

$$\partial b_1|_s = \delta^{(h)} = [0.00, 0.56].$$

- $x = -1$: $e = 0.94$, $h = [0.00, 0.70]$, $1_{z>0} = [0, 1]$.

$$\begin{aligned}\partial W_2|_s &= e h = [0.94 \cdot 0.00, 0.94 \cdot 0.70] = [0.00, 0.66], \\ \partial b_2|_s &= e = 0.94, \\ eW_2 &= 0.94[1.70, -0.80] = [1.60, -0.75], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [1.60 \cdot 0, -0.75 \cdot 1] = [0.00, -0.75], \\ \partial W_1|_s &= \delta^{(h)}x = [0.00 \cdot -1, -0.75 \cdot -1] = [0.00, 0.75], \\ \partial b_1|_s &= \delta^{(h)} = [0.00, -0.75].\end{aligned}$$

- $x = 0$: $e = 1.18$, $h = [0.40, 0.00]$, $1_{z>0} = [1, 0]$.

$$\begin{aligned}\partial W_2|_s &= e h = [1.18 \cdot 0.40, 1.18 \cdot 0.00] = [0.47, 0.00], \\ \partial b_2|_s &= e = 1.18, \\ eW_2 &= 1.18[1.70, -0.80] = [2.01, -0.94], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [2.01 \cdot 1, -0.94 \cdot 0] = [2.01, 0.00], \\ \partial W_1|_s &= \delta^{(h)}x = [2.01 \cdot 0, 0.00 \cdot 0] = [0.00, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [2.01, 0.00].\end{aligned}$$

- $x = 1$: $e = 1.07$, $h = [2.10, 0.00]$, $1_{z>0} = [1, 0]$.

$$\begin{aligned}\partial W_2|_s &= e h = [1.07 \cdot 2.10, 1.07 \cdot 0.00] = [2.25, 0.00], \\ \partial b_2|_s &= e = 1.07, \\ eW_2 &= 1.07[1.70, -0.80] = [1.82, -0.86], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [1.82 \cdot 1, -0.86 \cdot 0] = [1.82, 0.00], \\ \partial W_1|_s &= \delta^{(h)}x = [1.82 \cdot 1, 0.00 \cdot 1] = [1.82, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [1.82, 0.00].\end{aligned}$$

- $x = 2$: $e = -1.04$, $h = [3.80, 0.00]$, $1_{z>0} = [1, 0]$.

$$\begin{aligned}\partial W_2|_s &= e h = [-1.04 \cdot 3.80, -1.04 \cdot 0.00] = [-3.95, 0.00], \\ \partial b_2|_s &= e = -1.04, \\ eW_2 &= -1.04[1.70, -0.80] = [-1.77, 0.83], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [-1.77 \cdot 1, 0.83 \cdot 0] = [-1.77, 0.00], \\ \partial W_1|_s &= \delta^{(h)}x = [-1.77 \cdot 2, 0.00 \cdot 2] = [-3.54, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [-1.77, 0.00].\end{aligned}$$

Batch sums:

$$\partial W_2 = [-1.23, -0.39], \quad \partial b_2 = 1.45, \quad \partial W_1 = [-1.72, -0.37], \quad \partial b_1 = [2.06, -0.19].$$

Update ($\alpha = 0.05$):

$$W_2^+ = [1.70, -0.80] - 0.05 [-1.23, -0.39] = [1.70, -0.80] - [-0.06, -0.02] = [1.76, -0.78],$$

$$b_2^+ = 0.50 - 0.05 \cdot (1.45) = 0.50 - 0.07 = 0.43,$$

$$W_1^+ = \begin{bmatrix} 1.70 \\ -0.80 \end{bmatrix} - 0.05 \begin{bmatrix} -1.72 \\ -0.37 \end{bmatrix} = \begin{bmatrix} 1.70 \\ -0.80 \end{bmatrix} - \begin{bmatrix} -0.09 \\ -0.02 \end{bmatrix} = \begin{bmatrix} 1.79 \\ -0.78 \end{bmatrix},$$

$$b_1^+ = \begin{bmatrix} 0.40 \\ -0.10 \end{bmatrix} - 0.05 \begin{bmatrix} 2.06 \\ -0.19 \end{bmatrix} = \begin{bmatrix} 0.40 \\ -0.10 \end{bmatrix} - \begin{bmatrix} 0.10 \\ -0.01 \end{bmatrix} = \begin{bmatrix} 0.30 \\ -0.09 \end{bmatrix}.$$

Iteration 2: forward, gradients, update to Iteration 3

Forward pass (parameters at start of Iteration 2). Loss $L = 2.01$.

x	y	z_1	h	\hat{y}	e
-2	0.00	[-3.28, 1.47]	[0.00, 1.47]	-0.72	-0.72
-1	-1.00	[-1.49, 0.69]	[0.00, 0.69]	-0.11	0.89
0	0.00	[0.30, -0.09]	[0.30, 0.00]	0.96	0.96
1	3.00	[2.09, -0.87]	[2.09, 0.00]	4.11	1.11
2	8.00	[3.88, -1.65]	[3.88, 0.00]	7.26	-0.74

Per-sample gradients (no abbreviation; 2-dec rounding at each product):

- $x = -2$: $e = -0.72$, $h = [0.00, 1.47]$, $1_{z>0} = [0, 1]$.

$$\partial W_2|_s = e h = [-0.72 \cdot 0.00, -0.72 \cdot 1.47] = [0.00, -1.06],$$

$$\partial b_2|_s = e = -0.72,$$

$$e W_2 = -0.72 [1.76, -0.78] = [-1.27, 0.56],$$

$$\delta^{(h)} = e W_2 \odot 1_{z>0} = [-1.27 \cdot 0, 0.56 \cdot 1] = [0.00, 0.56],$$

$$\partial W_1|_s = \delta^{(h)} x = [0.00 \cdot -2, 0.56 \cdot -2] = [0.00, -1.12],$$

$$\partial b_1|_s = \delta^{(h)} = [0.00, 0.56].$$

- $x = -1$: $e = 0.89$, $h = [0.00, 0.69]$, $1_{z>0} = [0, 1]$.

$$\partial W_2|_s = e h = [0.89 \cdot 0.00, 0.89 \cdot 0.69] = [0.00, 0.61],$$

$$\partial b_2|_s = e = 0.89,$$

$$e W_2 = 0.89 [1.76, -0.78] = [1.57, -0.69],$$

$$\delta^{(h)} = e W_2 \odot 1_{z>0} = [1.57 \cdot 0, -0.69 \cdot 1] = [0.00, -0.69],$$

$$\partial W_1|_s = \delta^{(h)} x = [0.00 \cdot -1, -0.69 \cdot -1] = [0.00, 0.69],$$

$$\partial b_1|_s = \delta^{(h)} = [0.00, -0.69].$$

- $x = 0$: $e = 0.96$, $h = [0.30, 0.00]$, $1_{z>0} = [1, 0]$.

$$\begin{aligned}\partial W_2|_s &= e h = [0.96 \cdot 0.30, 0.96 \cdot 0.00] = [0.29, 0.00], \\ \partial b_2|_s &= e = 0.96, \\ eW_2 &= 0.96[1.76, -0.78] = [1.69, -0.75], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [1.69 \cdot 1, -0.75 \cdot 0] = [1.69, 0.00], \\ \partial W_1|_s &= \delta^{(h)} x = [1.69 \cdot 0, 0.00 \cdot 0] = [0.00, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [1.69, 0.00].\end{aligned}$$

- $x = 1$: $e = 1.11$, $h = [2.09, 0.00]$, $1_{z>0} = [1, 0]$.

$$\begin{aligned}\partial W_2|_s &= e h = [1.11 \cdot 2.09, 1.11 \cdot 0.00] = [2.32, 0.00], \\ \partial b_2|_s &= e = 1.11, \\ eW_2 &= 1.11[1.76, -0.78] = [1.95, -0.87], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [1.95 \cdot 1, -0.87 \cdot 0] = [1.95, 0.00], \\ \partial W_1|_s &= \delta^{(h)} x = [1.95 \cdot 1, 0.00 \cdot 1] = [1.95, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [1.95, 0.00].\end{aligned}$$

- $x = 2$: $e = -0.74$, $h = [3.88, 0.00]$, $1_{z>0} = [1, 0]$.

$$\begin{aligned}\partial W_2|_s &= e h = [-0.74 \cdot 3.88, -0.74 \cdot 0.00] = [-2.87, 0.00], \\ \partial b_2|_s &= e = -0.74, \\ eW_2 &= -0.74[1.76, -0.78] = [-1.30, 0.58], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [-1.30 \cdot 1, 0.58 \cdot 0] = [-1.30, 0.00], \\ \partial W_1|_s &= \delta^{(h)} x = [-1.30 \cdot 2, 0.00 \cdot 2] = [-2.60, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [-1.30, 0.00].\end{aligned}$$

Batch sums:

$$\partial W_2 = [-0.26, -0.45], \quad \partial b_2 = 1.50, \quad \partial W_1 = [-0.65, -0.43], \quad \partial b_1 = [2.34, -0.13].$$

Update ($\alpha = 0.05$):

$$\begin{aligned}W_2^+ &= [1.76, -0.78] - 0.05 [-0.26, -0.45] = [1.76, -0.78] - [-0.01, -0.02] = [1.77, -0.76], \\ b_2^+ &= 0.43 - 0.05 \cdot (1.50) = 0.43 - 0.08 = 0.35, \\ W_1^+ &= \begin{bmatrix} 1.79 \\ -0.78 \end{bmatrix} - 0.05 \begin{bmatrix} -0.65 \\ -0.43 \end{bmatrix} = \begin{bmatrix} 1.79 \\ -0.78 \end{bmatrix} - \begin{bmatrix} -0.03 \\ -0.02 \end{bmatrix} = \begin{bmatrix} 1.82 \\ -0.76 \end{bmatrix}, \\ b_1^+ &= \begin{bmatrix} 0.30 \\ -0.09 \end{bmatrix} - 0.05 \begin{bmatrix} 2.34 \\ -0.13 \end{bmatrix} = \begin{bmatrix} 0.30 \\ -0.09 \end{bmatrix} - \begin{bmatrix} 0.12 \\ -0.01 \end{bmatrix} = \begin{bmatrix} 0.18 \\ -0.08 \end{bmatrix}.\end{aligned}$$

Iteration 3: forward, gradients, update to Iteration 4

Forward pass (parameters at start of Iteration 3). Loss $L = 1.63$.

x	y	z_1	h	\hat{y}	e
-2	0.00	[-3.46, 1.44]	[0.00, 1.44]	-0.74	-0.74
-1	-1.00	[-1.64, 0.68]	[0.00, 0.68]	-0.17	0.83
0	0.00	[0.18, -0.08]	[0.18, 0.00]	0.67	0.67
1	3.00	[2.00, -0.84]	[2.00, 0.00]	3.89	0.89
2	8.00	[3.82, -1.60]	[3.82, 0.00]	7.11	-0.89

Per-sample gradients (no abbreviation; 2-dec rounding at each product):

- $x = -2$: $e = -0.74$, $h = [0.00, 1.44]$, $1_{z>0} = [0, 1]$.

$$\partial W_2|_s = e h = [-0.74 \cdot 0.00, -0.74 \cdot 1.44] = [0.00, -1.07],$$

$$\partial b_2|_s = e = -0.74,$$

$$e W_2 = -0.74[1.77, -0.76] = [-1.31, 0.56],$$

$$\delta^{(h)} = e W_2 \odot 1_{z>0} = [-1.31 \cdot 0, 0.56 \cdot 1] = [0.00, 0.56],$$

$$\partial W_1|_s = \delta^{(h)} x = [0.00 \cdot -2, 0.56 \cdot -2] = [0.00, -1.12],$$

$$\partial b_1|_s = \delta^{(h)} = [0.00, 0.56].$$

- $x = -1$: $e = 0.83$, $h = [0.00, 0.68]$, $1_{z>0} = [0, 1]$.

$$\partial W_2|_s = e h = [0.83 \cdot 0.00, 0.83 \cdot 0.68] = [0.00, 0.56],$$

$$\partial b_2|_s = e = 0.83,$$

$$e W_2 = 0.83[1.77, -0.76] = [1.47, -0.63],$$

$$\delta^{(h)} = e W_2 \odot 1_{z>0} = [1.47 \cdot 0, -0.63 \cdot 1] = [0.00, -0.63],$$

$$\partial W_1|_s = \delta^{(h)} x = [0.00 \cdot -1, -0.63 \cdot -1] = [0.00, 0.63],$$

$$\partial b_1|_s = \delta^{(h)} = [0.00, -0.63].$$

- $x = 0$: $e = 0.67$, $h = [0.18, 0.00]$, $1_{z>0} = [1, 0]$.

$$\partial W_2|_s = e h = [0.67 \cdot 0.18, 0.67 \cdot 0.00] = [0.12, 0.00],$$

$$\partial b_2|_s = e = 0.67,$$

$$e W_2 = 0.67[1.77, -0.76] = [1.19, -0.51],$$

$$\delta^{(h)} = e W_2 \odot 1_{z>0} = [1.19 \cdot 1, -0.51 \cdot 0] = [1.19, 0.00],$$

$$\partial W_1|_s = \delta^{(h)} x = [1.19 \cdot 0, 0.00 \cdot 0] = [0.00, 0.00],$$

$$\partial b_1|_s = \delta^{(h)} = [1.19, 0.00].$$

- $x = 1$: $e = 0.89$, $h = [2.00, 0.00]$, $1_{z>0} = [1, 0]$.

$$\partial W_2|_s = e h = [0.89 \cdot 2.00, 0.89 \cdot 0.00] = [1.78, 0.00],$$

$$\partial b_2|_s = e = 0.89,$$

$$e W_2 = 0.89[1.77, -0.76] = [1.58, -0.68],$$

$$\delta^{(h)} = e W_2 \odot 1_{z>0} = [1.58 \cdot 1, -0.68 \cdot 0] = [1.58, 0.00],$$

$$\partial W_1|_s = \delta^{(h)} x = [1.58 \cdot 1, 0.00 \cdot 1] = [1.58, 0.00],$$

$$\partial b_1|_s = \delta^{(h)} = [1.58, 0.00].$$

- $x = 2$: $e = -0.89$, $h = [3.82, 0.00]$, $1_{z>0} = [1, 0]$.

$$\partial W_2|_s = e h = [-0.89 \cdot 3.82, -0.89 \cdot 0.00] = [-3.40, 0.00],$$

$$\partial b_2|_s = e = -0.89,$$

$$eW_2 = -0.89[1.77, -0.76] = [-1.58, 0.68],$$

$$\delta^{(h)} = eW_2 \odot 1_{z>0} = [-1.58 \cdot 1, 0.68 \cdot 0] = [-1.58, 0.00],$$

$$\partial W_1|_s = \delta^{(h)} x = [-1.58 \cdot 2, 0.00 \cdot 2] = [-3.16, 0.00],$$

$$\partial b_1|_s = \delta^{(h)} = [-1.58, 0.00].$$

Batch sums:

$$\partial W_2 = [-1.50, -0.51], \quad \partial b_2 = 0.76, \quad \partial W_1 = [-1.58, -0.49], \quad \partial b_1 = [1.19, -0.07].$$

Update ($\alpha = 0.05$):

$$W_2^+ = [1.77, -0.76] - 0.05[-1.50, -0.51] = [1.77, -0.76] - [-0.07, -0.03] = [1.84, -0.73],$$

$$b_2^+ = 0.35 - 0.05 \cdot (0.76) = 0.35 - 0.04 = 0.31,$$

$$W_1^+ = \begin{bmatrix} 1.82 \\ -0.76 \end{bmatrix} - 0.05 \begin{bmatrix} -1.58 \\ -0.49 \end{bmatrix} = \begin{bmatrix} 1.82 \\ -0.76 \end{bmatrix} - \begin{bmatrix} -0.08 \\ -0.02 \end{bmatrix} = \begin{bmatrix} 1.90 \\ -0.74 \end{bmatrix},$$

$$b_1^+ = \begin{bmatrix} 0.18 \\ -0.08 \end{bmatrix} - 0.05 \begin{bmatrix} 1.19 \\ -0.07 \end{bmatrix} = \begin{bmatrix} 0.18 \\ -0.08 \end{bmatrix} - \begin{bmatrix} 0.06 \\ -0.00 \end{bmatrix} = \begin{bmatrix} 0.12 \\ -0.08 \end{bmatrix}.$$

Iteration 4: forward, gradients, update to Iteration 5

Forward pass (parameters at start of Iteration 4). Loss $L = 1.38$.

x	y	z_1	h	\hat{y}	e
-2	0.00	[-3.68, 1.40]	[0.00, 1.40]	-0.71	-0.71
-1	-1.00	[-1.78, 0.66]	[0.00, 0.66]	-0.17	0.83
0	0.00	[0.12, -0.08]	[0.12, 0.00]	0.53	0.53
1	3.00	[2.02, -0.82]	[2.02, 0.00]	4.03	1.03
2	8.00	[3.92, -1.56]	[3.92, 0.00]	7.52	-0.48

Per-sample gradients (no abbreviation; 2-dec rounding at each product):

- $x = -2$: $e = -0.71$, $h = [0.00, 1.40]$, $1_{z>0} = [0, 1]$.

$$\partial W_2|_s = e h = [-0.71 \cdot 0.00, -0.71 \cdot 1.40] = [0.00, -0.99],$$

$$\partial b_2|_s = e = -0.71,$$

$$eW_2 = -0.71[1.84, -0.73] = [-1.31, 0.52],$$

$$\delta^{(h)} = eW_2 \odot 1_{z>0} = [-1.31 \cdot 0, 0.52 \cdot 1] = [0.00, 0.52],$$

$$\partial W_1|_s = \delta^{(h)} x = [0.00 \cdot -2, 0.52 \cdot -2] = [0.00, -1.04],$$

$$\partial b_1|_s = \delta^{(h)} = [0.00, 0.52].$$

- $x = -1$: $e = 0.83$, $h = [0.00, 0.66]$, $1_{z>0} = [0, 1]$.

$$\begin{aligned}\partial W_2|_s &= e h = [0.83 \cdot 0.00, 0.83 \cdot 0.66] = [0.00, 0.55], \\ \partial b_2|_s &= e = 0.83, \\ eW_2 &= 0.83[1.84, -0.73] = [1.53, -0.61], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [1.53 \cdot 0, -0.61 \cdot 1] = [0.00, -0.61], \\ \partial W_1|_s &= \delta^{(h)}x = [0.00 \cdot -1, -0.61 \cdot -1] = [0.00, 0.61], \\ \partial b_1|_s &= \delta^{(h)} = [0.00, -0.61].\end{aligned}$$

- $x = 0$: $e = 0.53$, $h = [0.12, 0.00]$, $1_{z>0} = [1, 0]$.

$$\begin{aligned}\partial W_2|_s &= e h = [0.53 \cdot 0.12, 0.53 \cdot 0.00] = [0.06, 0.00], \\ \partial b_2|_s &= e = 0.53, \\ eW_2 &= 0.53[1.84, -0.73] = [0.98, -0.39], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [0.98 \cdot 1, -0.39 \cdot 0] = [0.98, 0.00], \\ \partial W_1|_s &= \delta^{(h)}x = [0.98 \cdot 0, 0.00 \cdot 0] = [0.00, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [0.98, 0.00].\end{aligned}$$

- $x = 1$: $e = 1.03$, $h = [2.02, 0.00]$, $1_{z>0} = [1, 0]$.

$$\begin{aligned}\partial W_2|_s &= e h = [1.03 \cdot 2.02, 1.03 \cdot 0.00] = [2.08, 0.00], \\ \partial b_2|_s &= e = 1.03, \\ eW_2 &= 1.03[1.84, -0.73] = [1.90, -0.75], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [1.90 \cdot 1, -0.75 \cdot 0] = [1.90, 0.00], \\ \partial W_1|_s &= \delta^{(h)}x = [1.90 \cdot 1, 0.00 \cdot 1] = [1.90, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [1.90, 0.00].\end{aligned}$$

- $x = 2$: $e = -0.48$, $h = [3.92, 0.00]$, $1_{z>0} = [1, 0]$.

$$\begin{aligned}\partial W_2|_s &= e h = [-0.48 \cdot 3.92, -0.48 \cdot 0.00] = [-1.88, 0.00], \\ \partial b_2|_s &= e = -0.48, \\ eW_2 &= -0.48[1.84, -0.73] = [-0.88, 0.35], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [-0.88 \cdot 1, 0.35 \cdot 0] = [-0.88, 0.00], \\ \partial W_1|_s &= \delta^{(h)}x = [-0.88 \cdot 2, 0.00 \cdot 2] = [-1.76, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [-0.88, 0.00].\end{aligned}$$

Batch sums:

$$\partial W_2 = [0.26, -0.44], \quad \partial b_2 = 1.20, \quad \partial W_1 = [0.14, -0.43], \quad \partial b_1 = [2.00, -0.09].$$

Update ($\alpha = 0.05$):

$$\begin{aligned}
W_2^+ &= [1.84, -0.73] - 0.05 [0.26, -0.44] = [1.84, -0.73] - [0.01, -0.02] = [1.83, -0.71], \\
b_2^+ &= 0.31 - 0.05 \cdot (1.20) = 0.31 - 0.06 = 0.25, \\
W_1^+ &= \begin{bmatrix} 1.90 \\ -0.74 \end{bmatrix} - 0.05 \begin{bmatrix} 0.14 \\ -0.43 \end{bmatrix} = \begin{bmatrix} 1.90 \\ -0.74 \end{bmatrix} - \begin{bmatrix} 0.01 \\ -0.02 \end{bmatrix} = \begin{bmatrix} 1.89 \\ -0.72 \end{bmatrix}, \\
b_1^+ &= \begin{bmatrix} 0.12 \\ -0.08 \end{bmatrix} - 0.05 \begin{bmatrix} 2.00 \\ -0.09 \end{bmatrix} = \begin{bmatrix} 0.12 \\ -0.08 \end{bmatrix} - \begin{bmatrix} 0.10 \\ -0.00 \end{bmatrix} = \begin{bmatrix} 0.02 \\ -0.08 \end{bmatrix}.
\end{aligned}$$

Iteration 5: forward, gradients, update to Iteration 6

Forward pass (parameters at start of Iteration 5). Loss $L = 1.22$.

x	y	z_1	h	\hat{y}	e
-2	0.00	[-3.76, 1.36]	[0.00, 1.36]	-0.72	-0.72
-1	-1.00	[-1.87, 0.64]	[0.00, 0.64]	-0.20	0.80
0	0.00	[0.02, -0.08]	[0.02, 0.00]	0.29	0.29
1	3.00	[1.91, -0.80]	[1.91, 0.00]	3.75	0.75
2	8.00	[3.80, -1.52]	[3.80, 0.00]	7.20	-0.80

Per-sample gradients (no abbreviation; 2-dec rounding at each product):

- $x = -2$: $e = -0.72$, $h = [0.00, 1.36]$, $1_{z>0} = [0, 1]$.

$$\begin{aligned}
\partial W_2|_s &= e h = [-0.72 \cdot 0.00, -0.72 \cdot 1.36] = [0.00, -0.98], \\
\partial b_2|_s &= e = -0.72, \\
e W_2 &= -0.72 [1.83, -0.71] = [-1.32, 0.51], \\
\delta^{(h)} &= e W_2 \odot 1_{z>0} = [-1.32 \cdot 0, 0.51 \cdot 1] = [0.00, 0.51], \\
\partial W_1|_s &= \delta^{(h)} x = [0.00 \cdot -2, 0.51 \cdot -2] = [0.00, -1.02], \\
\partial b_1|_s &= \delta^{(h)} = [0.00, 0.51].
\end{aligned}$$

- $x = -1$: $e = 0.80$, $h = [0.00, 0.64]$, $1_{z>0} = [0, 1]$.

$$\begin{aligned}
\partial W_2|_s &= e h = [0.80 \cdot 0.00, 0.80 \cdot 0.64] = [0.00, 0.51], \\
\partial b_2|_s &= e = 0.80, \\
e W_2 &= 0.80 [1.83, -0.71] = [1.46, -0.57], \\
\delta^{(h)} &= e W_2 \odot 1_{z>0} = [1.46 \cdot 0, -0.57 \cdot 1] = [0.00, -0.57], \\
\partial W_1|_s &= \delta^{(h)} x = [0.00 \cdot -1, -0.57 \cdot -1] = [0.00, 0.57], \\
\partial b_1|_s &= \delta^{(h)} = [0.00, -0.57].
\end{aligned}$$

- $x = 0$: $e = 0.29$, $h = [0.02, 0.00]$, $1_{z>0} = [1, 0]$.

$$\begin{aligned}\partial W_2|_s &= e h = [0.29 \cdot 0.02, 0.29 \cdot 0.00] = [0.01, 0.00], \\ \partial b_2|_s &= e = 0.29, \\ eW_2 &= 0.29[1.83, -0.71] = [0.53, -0.21], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [0.53 \cdot 1, -0.21 \cdot 0] = [0.53, 0.00], \\ \partial W_1|_s &= \delta^{(h)}x = [0.53 \cdot 0, 0.00 \cdot 0] = [0.00, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [0.53, 0.00].\end{aligned}$$

- $x = 1$: $e = 0.75$, $h = [1.91, 0.00]$, $1_{z>0} = [1, 0]$.

$$\begin{aligned}\partial W_2|_s &= e h = [0.75 \cdot 1.91, 0.75 \cdot 0.00] = [1.43, 0.00], \\ \partial b_2|_s &= e = 0.75, \\ eW_2 &= 0.75[1.83, -0.71] = [1.37, -0.53], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [1.37 \cdot 1, -0.53 \cdot 0] = [1.37, 0.00], \\ \partial W_1|_s &= \delta^{(h)}x = [1.37 \cdot 1, 0.00 \cdot 1] = [1.37, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [1.37, 0.00].\end{aligned}$$

- $x = 2$: $e = -0.80$, $h = [3.80, 0.00]$, $1_{z>0} = [1, 0]$.

$$\begin{aligned}\partial W_2|_s &= e h = [-0.80 \cdot 3.80, -0.80 \cdot 0.00] = [-3.04, 0.00], \\ \partial b_2|_s &= e = -0.80, \\ eW_2 &= -0.80[1.83, -0.71] = [-1.46, 0.57], \\ \delta^{(h)} &= eW_2 \odot 1_{z>0} = [-1.46 \cdot 1, 0.57 \cdot 0] = [-1.46, 0.00], \\ \partial W_1|_s &= \delta^{(h)}x = [-1.46 \cdot 2, 0.00 \cdot 2] = [-2.92, 0.00], \\ \partial b_1|_s &= \delta^{(h)} = [-1.46, 0.00].\end{aligned}$$

Batch sums:

$$\partial W_2 = [-1.60, -0.47], \quad \partial b_2 = 0.32, \quad \partial W_1 = [-1.55, -0.45], \quad \partial b_1 = [0.44, -0.06].$$

Update ($\alpha = 0.05$):

$$\begin{aligned}W_2^+ &= [1.83, -0.71] - 0.05 [-1.60, -0.47] = [1.83, -0.71] - [-0.08, -0.02] = [1.91, -0.69], \\ b_2^+ &= 0.25 - 0.05 \cdot (0.32) = 0.25 - 0.02 = 0.23, \\ W_1^+ &= \begin{bmatrix} 1.89 \\ -0.72 \end{bmatrix} - 0.05 \begin{bmatrix} -1.55 \\ -0.45 \end{bmatrix} = \begin{bmatrix} 1.89 \\ -0.72 \end{bmatrix} - \begin{bmatrix} -0.08 \\ -0.02 \end{bmatrix} = \begin{bmatrix} 1.97 \\ -0.70 \end{bmatrix}, \\ b_1^+ &= \begin{bmatrix} 0.02 \\ -0.08 \end{bmatrix} - 0.05 \begin{bmatrix} 0.44 \\ -0.06 \end{bmatrix} = \begin{bmatrix} 0.02 \\ -0.08 \end{bmatrix} - \begin{bmatrix} 0.02 \\ -0.00 \end{bmatrix} = \begin{bmatrix} 0.00 \\ -0.08 \end{bmatrix}.\end{aligned}$$

Iteration 6: forward only (to show convergence)

Forward pass (parameters at start of Iteration 6). Loss $L = 1.10$.

x	y	z_1	h	\hat{y}	e
-2	0.00	[-3.94, 1.32]	[0.00, 1.32]	-0.68	-0.68
-1	-1.00	[-1.97, 0.62]	[0.00, 0.62]	-0.20	0.80
0	0.00	[0.00, -0.08]	[0.00, 0.00]	0.23	0.23
1	3.00	[1.97, -0.78]	[1.97, 0.00]	3.99	0.99
2	8.00	[3.94, -1.48]	[3.94, 0.00]	7.76	-0.24

Per-sample gradients (no abbreviation; 2-dec rounding at each product):

- $x = -2$: $e = -0.68$, $h = [0.00, 1.32]$, $1_{z>0} = [0, 1]$.

$$\partial W_2|_s = e h = [-0.68 \cdot 0.00, -0.68 \cdot 1.32] = [0.00, -0.90],$$

$$\partial b_2|_s = e = -0.68,$$

$$eW_2 = -0.68[1.91, -0.69] = [-1.30, 0.47],$$

$$\delta^{(h)} = eW_2 \odot 1_{z>0} = [-1.30 \cdot 0, 0.47 \cdot 1] = [0.00, 0.47],$$

$$\partial W_1|_s = \delta^{(h)} x = [0.00 \cdot -2, 0.47 \cdot -2] = [0.00, -0.94],$$

$$\partial b_1|_s = \delta^{(h)} = [0.00, 0.47].$$

- $x = -1$: $e = 0.80$, $h = [0.00, 0.62]$, $1_{z>0} = [0, 1]$.

$$\partial W_2|_s = e h = [0.80 \cdot 0.00, 0.80 \cdot 0.62] = [0.00, 0.50],$$

$$\partial b_2|_s = e = 0.80,$$

$$eW_2 = 0.80[1.91, -0.69] = [1.53, -0.55],$$

$$\delta^{(h)} = eW_2 \odot 1_{z>0} = [1.53 \cdot 0, -0.55 \cdot 1] = [0.00, -0.55],$$

$$\partial W_1|_s = \delta^{(h)} x = [0.00 \cdot -1, -0.55 \cdot -1] = [0.00, 0.55],$$

$$\partial b_1|_s = \delta^{(h)} = [0.00, -0.55].$$

- $x = 0$: $e = 0.23$, $h = [0.00, 0.00]$, $1_{z>0} = [0, 0]$.

$$\partial W_2|_s = e h = [0.23 \cdot 0.00, 0.23 \cdot 0.00] = [0.00, 0.00],$$

$$\partial b_2|_s = e = 0.23,$$

$$eW_2 = 0.23[1.91, -0.69] = [0.44, -0.16],$$

$$\delta^{(h)} = eW_2 \odot 1_{z>0} = [0.44 \cdot 0, -0.16 \cdot 0] = [0.00, 0.00],$$

$$\partial W_1|_s = \delta^{(h)} x = [0.00 \cdot 0, 0.00 \cdot 0] = [0.00, 0.00],$$

$$\partial b_1|_s = \delta^{(h)} = [0.00, 0.00].$$

- $x = 1$: $e = 0.99$, $h = [1.97, 0.00]$, $1_{z>0} = [1, 0]$.

$$\partial W_2|_s = e h = [0.99 \cdot 1.97, 0.99 \cdot 0.00] = [1.95, 0.00],$$

$$\partial b_2|_s = e = 0.99,$$

$$eW_2 = 0.99[1.91, -0.69] = [1.89, -0.68],$$

$$\delta^{(h)} = eW_2 \odot 1_{z>0} = [1.89 \cdot 1, -0.68 \cdot 0] = [1.89, 0.00],$$

$$\partial W_1|_s = \delta^{(h)} x = [1.89 \cdot 1, 0.00 \cdot 1] = [1.89, 0.00],$$

$$\partial b_1|_s = \delta^{(h)} = [1.89, 0.00].$$

- $x = 2$: $e = -0.24$, $h = [3.94, 0.00]$, $1_{z>0} = [1, 0]$.

$$\partial W_2|_s = e h = [-0.24 \cdot 3.94, -0.24 \cdot 0.00] = [-0.95, 0.00],$$

$$\partial b_2|_s = e = -0.24,$$

$$eW_2 = -0.24[1.91, -0.69] = [-0.46, 0.17],$$

$$\delta^{(h)} = eW_2 \odot 1_{z>0} = [-0.46 \cdot 1, 0.17 \cdot 0] = [-0.46, 0.00],$$

$$\partial W_1|_s = \delta^{(h)} x = [-0.46 \cdot 2, 0.00 \cdot 2] = [-0.92, 0.00],$$

$$\partial b_1|_s = \delta^{(h)} = [-0.46, 0.00].$$