ML Assignment 4

Bohan YANG
Student ID: 2330016056

November 24, 2025

Question 1

1) Derive the equations

1. **Hidden Layer Linear Transformation:**

$$Z_1 = W_1 x + b_1$$

2. **Hidden Layer Activation:**

$$H = f(Z_1) = \mathrm{ReLU}(Z_1) + \sin(Z_1)$$

3. **Output Layer Linear Transformation:**

$$Z_2 = W_2 H + b_2$$

4. **Output Layer Activation (Softmax):**

$$\hat{y} = \sigma(Z_2) = \frac{e^{Z_2}}{\sum_{j=1}^{k} e^{Z_{2,j}}}$$

2) Calculate outputs explicitly

Given:

$$x = \begin{pmatrix} 1 \\ -1 \\ 0.5 \\ 2 \end{pmatrix}, \quad W_1 = \begin{pmatrix} 0.1 & -0.2 & 0.3 & 0.4 \\ 0.5 & -0.3 & 0.1 & -0.2 \\ 0.4 & 0.2 & -0.5 & 0.3 \end{pmatrix}, \quad b_1 = \begin{pmatrix} 0.1 \\ -0.1 \\ 0.05 \end{pmatrix}$$

$$W_2 = \begin{pmatrix} -0.3 & 0.2 & 0.1 \\ 0.4 & -0.5 & 0.3 \end{pmatrix}, \quad b_2 = \begin{pmatrix} 0.05 \\ -0.05 \end{pmatrix}$$

Step 1: Calculate Z_1

$$Z_1 = \begin{pmatrix} 0.1 & -0.2 & 0.3 & 0.4 \\ 0.5 & -0.3 & 0.1 & -0.2 \\ 0.4 & 0.2 & -0.5 & 0.3 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0.5 \\ 2 \end{pmatrix} + \begin{pmatrix} 0.1 \\ -0.1 \\ 0.05 \end{pmatrix}$$

$$Z_{1,1} = (0.1)(1) + (-0.2)(-1) + (0.3)(0.5) + (0.4)(2) + 0.1 = 0.1 + 0.2 + 0.15 + 0.8 + 0.1 = 1.35$$
$$Z_{1,2} = (0.5)(1) + (-0.3)(-1) + (0.1)(0.5) + (-0.2)(2) - 0.1 = 0.5 + 0.3 + 0.05 - 0.4 - 0.1 = 0.35$$
$$Z_{1,3} = (0.4)(1) + (0.2)(-1) + (-0.5)(0.5) + (0.3)(2) + 0.05 = 0.4 - 0.2 - 0.25 + 0.6 + 0.05 = 0.60$$

$$\mathbf{Z_1} = \begin{pmatrix} 1.3500 \\ 0.3500 \\ 0.6000 \end{pmatrix}$$

Step 2: Calculate H Since all elements of $Z_1 > 0$, $\mathrm{ReLU}(z) = z$. Thus $H = Z_1 + \sin(Z_1)$.

$$H_1 = 1.35 + \sin(1.35) \approx 1.35 + 0.9757 = 2.3257$$
$$H_2 = 0.35 + \sin(0.35) \approx 0.35 + 0.3429 = 0.6929$$
$$H_3 = 0.60 + \sin(0.60) \approx 0.60 + 0.5646 = 1.1646$$

$$\mathbf{H} = \begin{pmatrix} 2.3257 \\ 0.6929 \\ 1.1646 \end{pmatrix}$$

Step 3: Calculate Z_2

$$Z_2 = \begin{pmatrix} -0.3 & 0.2 & 0.1 \\ 0.4 & -0.5 & 0.3 \end{pmatrix} \begin{pmatrix} 2.3257 \\ 0.6929 \\ 1.1646 \end{pmatrix} + \begin{pmatrix} 0.05 \\ -0.05 \end{pmatrix}$$

$$\begin{aligned} Z_{2,1} &= -0.3(2.3257) + 0.2(0.6929) + 0.1(1.1646) + 0.05 \\ &= -0.6977 + 0.1386 + 0.1165 + 0.05 = -0.3926 \\ Z_{2,2} &= 0.4(2.3257) - 0.5(0.6929) + 0.3(1.1646) - 0.05 \\ &= 0.9303 - 0.3465 + 0.3494 - 0.05 = 0.8832 \end{aligned}$$

$$\mathbf{Z_2} = \begin{pmatrix} -0.3926 \\ 0.8832 \end{pmatrix}$$

Step 4: Calculate $\hat{y}$ First, calculate the denominator $\sum e^{Z_{2,j}}$:

$$\text{Sum} = e^{-0.3926} + e^{0.8832} \approx 0.6753 + 2.4186 = 3.0939$$
$$\hat{y}_1 = \frac{0.6753}{3.0939} \approx 0.2183, \quad \hat{y}_2 = \frac{2.4186}{3.0939} \approx 0.7817$$
$$\hat{\mathbf{y}} = \begin{pmatrix} 0.2183 \\ 0.7817 \end{pmatrix}$$

Question 2

Derivation

The loss function is Cross-Entropy $L = -\sum y_i \log(\hat{y}_i)$.

1. Error at Output Layer (δ_2): For Softmax combined with Cross-Entropy loss, the gradient with respect to the pre-activation output Z_2 is:

$$\delta_2 = \frac{\partial L}{\partial Z_2} = \hat{y} - y$$

$$\frac{\partial L}{\partial W_2} = \delta_2 H^T, \quad \frac{\partial L}{\partial b_2} = \delta_2$$

2. Error at Hidden Layer (δ_1): We backpropagate δ_2 to the hidden layer.

$$\frac{\partial L}{\partial H} = W_2^T \delta_2$$

Next, we account for the activation function $f(z) = \text{ReLU}(z) + \sin(z)$. The derivative $f'(z)$ is:

$$f'(z) = \begin{cases} 1 + \cos(z) & \text{if } z > 0 \\ \cos(z) & \text{if } z \leq 0 \end{cases}$$

The error term δ_1 (gradient w.r.t Z_1) is computed using the Hadamard product ($\odot$):

$$\delta_1 = \frac{\partial L}{\partial Z_1} = (W_2^T \delta_2) \odot f'(Z_1)$$

$$\frac{\partial L}{\partial W_1} = \delta_1 x^T, \quad \frac{\partial L}{\partial b_1} = \delta_1$$

Calculation

Given target $y = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

1. Calculate δ_2 and Gradients for Layer 2:

$$\delta_2 = \begin{pmatrix} 0.2183 \\ 0.7817 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -0.7817 \\ 0.7817 \end{pmatrix}$$

$$\frac{\partial L}{\partial b_2} = \begin{pmatrix} -0.7817 \\ 0.7817 \end{pmatrix}$$

$$\frac{\partial L}{\partial W_2} = \delta_2 H^T = \begin{pmatrix} -0.7817 \\ 0.7817 \end{pmatrix} \begin{pmatrix} 2.3257 & 0.6929 & 1.1646 \end{pmatrix}$$

$$\frac{\partial L}{\partial W_2} \approx \begin{pmatrix} -1.8180 & -0.5416 & -0.9104 \\ 1.8180 & 0.5416 & 0.9104 \end{pmatrix}$$

2. Calculate δ_1 and Gradients for Layer 1: First, compute the backpropagated error term $W_2^T \delta_2$:

$$W_2^T \delta_2 = \begin{pmatrix} -0.3 & 0.4 \\ 0.2 & -0.5 \\ 0.1 & 0.3 \end{pmatrix} \begin{pmatrix} -0.7817 \\ 0.7817 \end{pmatrix}$$

Row 1: $(-0.3)(-0.7817) + 0.4(0.7817) = 0.2345 + 0.3127 = 0.5472$
Row 2: $(0.2)(-0.7817) - 0.5(0.7817) = -0.1563 - 0.3909 = -0.5472$
Row 3: $(0.1)(-0.7817) + 0.3(0.7817) = -0.0782 + 0.2345 = 0.1563$

$$W_2^T \delta_2 = \begin{pmatrix} 0.5472 \\ -0.5472 \\ 0.1563 \end{pmatrix}$$

Next, calculate $f'(Z_1)$. Since all components of Z_1 are positive $(1.35, 0.35, 0.60)$, $f'(z) = 1 + \cos(z)$.

$$f'(1.35) = 1 + \cos(1.35) \approx 1 + 0.2190 = 1.2190$$
$$f'(0.35) = 1 + \cos(0.35) \approx 1 + 0.9394 = 1.9394$$
$$f'(0.60) = 1 + \cos(0.60) \approx 1 + 0.8253 = 1.8253$$

$$\delta_1 = \begin{pmatrix} 0.5472 \\ -0.5472 \\ 0.1563 \end{pmatrix} \odot \begin{pmatrix} 1.2190 \\ 1.9394 \\ 1.8253 \end{pmatrix} = \begin{pmatrix} 0.6670 \\ -1.0612 \\ 0.2853 \end{pmatrix}$$

Now, compute gradients:

$$\frac{\partial L}{\partial b_1} = \begin{pmatrix} 0.6670 \\ -1.0612 \\ 0.2853 \end{pmatrix}$$

$$\frac{\partial L}{\partial W_1} = \delta_1 x^T = \begin{pmatrix} 0.6670 \\ -1.0612 \\ 0.2853 \end{pmatrix} \begin{pmatrix} 1 & -1 & 0.5 & 2 \end{pmatrix}$$

$$\frac{\partial L}{\partial W_1} \approx \begin{pmatrix} 0.6670 & -0.6670 & 0.3335 & 1.3340 \\ -1.0612 & 1.0612 & -0.5306 & -2.1224 \\ 0.2853 & -0.2853 & 0.1427 & 0.5706 \end{pmatrix}$$

Question 3

$\alpha = 0.001$.

$\theta_{new} = \theta_{old} - \alpha \frac{\partial L}{\partial \theta}$.

1. (W_2, b_2):

$$b_2^{new} = \begin{pmatrix} 0.05 \\ -0.05 \end{pmatrix} - 0.001 \begin{pmatrix} -0.7817 \\ 0.7817 \end{pmatrix} = \begin{pmatrix} 0.0508 \\ -0.0508 \end{pmatrix}$$

$$W_2^{new} = \begin{pmatrix} -0.3 & 0.2 & 0.1 \\ 0.4 & -0.5 & 0.3 \end{pmatrix} - 0.001 \begin{pmatrix} -1.8180 & -0.5416 & -0.9104 \\ 1.8180 & 0.5416 & 0.9104 \end{pmatrix}$$

$$W_2^{new} = \begin{pmatrix} -0.2982 & 0.2005 & 0.1009 \\ 0.3982 & -0.5005 & 0.2991 \end{pmatrix}$$

2. (W_1, b_1):

$$b_1^{new} = \begin{pmatrix} 0.1 \\ -0.1 \\ 0.05 \end{pmatrix} - 0.001 \begin{pmatrix} 0.6670 \\ -1.0612 \\ 0.2853 \end{pmatrix} = \begin{pmatrix} 0.0993 \\ -0.0989 \\ 0.0497 \end{pmatrix}$$

$$W_1^{new} = W_1 - 0.001 \frac{\partial L}{\partial W_1}$$

Performing the element-wise subtraction (multiplying gradients by 0.001 first):

$$\begin{aligned}
\text{Row 1: } & [0.1, -0.2, 0.3, 0.4] - [0.0007, -0.0007, 0.0003, 0.0013] \\
& = [0.0993, -0.1993, 0.2997, 0.3987] \\
\text{Row 2: } & [0.5, -0.3, 0.1, -0.2] - [-0.0011, 0.0011, -0.0005, -0.0021] \\
& = [0.5011, -0.3011, 0.1005, -0.1979] \\
\text{Row 3: } & [0.4, 0.2, -0.5, 0.3] - [0.0003, -0.0003, 0.0001, 0.0006] \\
& = [0.3997, 0.2003, -0.5001, 0.2994]
\end{aligned}$$

$$W_1^{new} = \begin{pmatrix} 0.0993 & -0.1993 & 0.2997 & 0.3987 \\ 0.5011 & -0.3011 & 0.1005 & -0.1979 \\ 0.3997 & 0.2003 & -0.5001 & 0.2994 \end{pmatrix}$$