

Exercise 3

Saturday, 15. April 2023 15:32

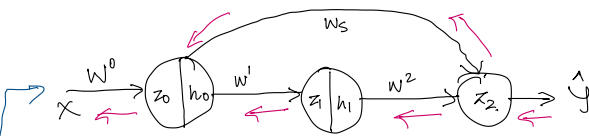
Exercise 3

1. A neural network with 3 layers (input, hidden, output) is shown below. The input layer has 2 nodes, the hidden layer has 2 nodes, and the output layer has 1 node. The weights are given in the table below.

From \ To	Node 1	Node 2
Input Node 1	0.1	0.2
Input Node 2	0.3	0.4
Hidden Node 1	0.5	0.6
Hidden Node 2	0.7	0.8
Output Node 1	0.9	1.0

2. The activation function for the hidden layer is the sigmoid function, and for the output layer is the linear function. The bias for the hidden layer is 0.5, and for the output layer is 0.1.

3. The input vector is  $x = [1, 2]^T$ . Calculate the output of the network.



**Forward pass:**

Unit 0

$$z^0 = w^0 x + b^0$$
$$h^0 = g^0(z^0)$$

Unit 1

$$z^1 = w^1 h^0 + b^1$$
$$h^1 = g^1(z^1)$$

Unit 2

$$z^2 = w^2 h^1 + w^3 h^0$$
$$y = z^2$$

Loss

$$L(y, \hat{y}) = |y - \hat{y}|$$

**Backward pass**

1)  $\frac{\partial L}{\partial z_2} = |y - z_2| = -\frac{(y - z_2)}{|y - z_2|} = L'$

2)  $\frac{\partial L}{\partial w_5} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_5} = L' \cdot h_1$

3)  $\frac{\partial L}{\partial w_4} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_4} = L' \cdot h_0$

4)  $\frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_3} = L' \cdot h_0$

5)  $\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_2} = L' \cdot h_1$

6)  $\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1} = L' \cdot h_0$

7)  $\frac{\partial L}{\partial w_0} = \frac{\partial L}{\partial z_0} \cdot \frac{\partial z_0}{\partial w_0} = L' \cdot x$

8)  $\frac{\partial L}{\partial b_0} = \frac{\partial L}{\partial z_0} \cdot \frac{\partial z_0}{\partial b_0} = L'$

9)  $\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial z_1} \cdot \frac{\partial z_1}{\partial b_1} = L'$

10)  $\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial b_2} = L'$

**Backpropagation:**

ReLU:  $\max(0, x)$

ReLU':  $\max(0, \frac{x}{|x|})$

Linear:  $w^0 x + b^0$

Linear':  $x^T$

Linear:  $w$

Linear':  $1$

CE loss with preapplied softmax.

CE loss =  $-\sum [a \log \text{softmax}(s) + (1-a) \log (1 - \text{softmax}(s))]$

Softmax(y) =  $\frac{\exp(y)}{\sum \exp(y)}$

Softmax' =  $\frac{1}{V} - \frac{y_1^i - y_2^i}{V^2} = \frac{\exp(y_1) \cdot \exp(y_2) - \exp(y_2)^2}{[\sum \exp(y)]^2}$

**Final**

$$\frac{\partial L}{\partial w_0} = (L' \cdot h_1 + L' \cdot h_0) \cdot \text{ReLU}'(z_0)$$
$$\frac{\partial L}{\partial w_1} = L' \cdot h_1 \cdot 1 \cdot \text{ReLU}'(x_1)$$
$$\frac{\partial L}{\partial w_2} = -\frac{(y - z_2)}{|y - z_2|} \cdot h_1$$
$$\frac{\partial L}{\partial w_3} = -\frac{(y - z_2)}{|y - z_2|} \cdot h_0$$

Due to skip connection, the gradients of both add up!

$$\frac{\partial L}{\partial w_0} = L' \cdot h_1 + L' \cdot h_0$$

$$\begin{aligned}
 (CE_{loss})' &= \sum_{j=1}^J \exp(x_j) R \left| \right. \\
 &= a \frac{1}{\text{softmax}(y)} \cdot \text{softmax}(y)' + (1-a) \frac{1}{1-\text{softmax}(y)} \cdot \text{softmax}(y)' \\
 &= \left[ \frac{a}{\text{softmax}(y)} - \frac{(1-a)}{1-\text{softmax}(y)} \right] \text{softmax}(y)' \\
 &= \left[ \frac{a(1-\text{softmax}) + (1-a)\text{softmax}}{\text{softmax}(1-\text{softmax})} \right] \text{softmax}'(y) \\
 &= \left[ \frac{a - a\text{softmax} + a\text{softmax} - \text{softmax}}{\text{softmax}(1-\text{softmax})} \right] \text{softmax}'(y) \\
 &= \left[ \frac{a - \text{softmax}}{\text{softmax}(1-\text{softmax})} \right] \text{softmax}'(y) \\
 &= \sum \left[ a_j - \frac{\exp(x_j)}{\sum \exp(x)} \right] \cdot \left[ \frac{\exp(x_k) [\sum \exp(x) - \exp(x_k)]}{[\sum \exp(x)]^2} \right]
 \end{aligned}$$

$$= \frac{\exp(x_k) [\sum \exp(x) - \exp(x_k)]}{[\sum \exp(x)]^2}$$