Exercise 2: Derivatives of Activation Function and Loss

14 May 2024

Differentiation Rules Used

1. **Power Rule:**
$$\frac{d}{dx}[x^n] = nx^{n-1}$$
2. **Chain Rule:**
$$\frac{d}{dx}[f(g(x))] = f'(g(x)) \cdot g'(x)$$
3. **Sum/Difference Rule:**
$$\frac{d}{dx}[f(x) \pm g(x)] = f'(x) \pm g'(x)$$
4. **Constant Multiple Rule:**
$$\frac{d}{dx}[c \cdot f(x)] = c \cdot f'(x)$$
5. **Derivative of the Exponential Function:**
$$\frac{d}{dz}[e^z] = e^z$$

Step 2: Backpropagation Phase

Compute output error

1. **Error (loss) derivative with respect to output:**
$$\frac{\partial E}{\partial o_1} = 2 \cdot (o_1 - y)$$
2. **Derivative of the sigmoid function at the output:**
$$\frac{\partial o_1}{\partial z_o} = o_1 \cdot (1 - o_1)$$
3. **Error (loss) with respect to z_o:**
$$\delta_o = \frac{\partial E}{\partial o_1} \cdot \frac{\partial o_1}{\partial z_o}$$

Derivations

1. Error (loss) derivative with respect to output

Given the MSE loss:

$$E = \frac{1}{2}(o - y)^2$$

We need to find the derivative of E with respect to the output o:

$$\frac{\partial E}{\partial o} = \frac{\partial}{\partial o}\left(\frac{1}{2}(o - y)^2\right)$$

Using the chain rule and power rule:

$$\frac{\partial E}{\partial o} = \frac{1}{2} \cdot 2 \cdot (o - y) = (o - y)$$

Including the factor of 2:

$$\frac{\partial E}{\partial o_1} = 2 \cdot (o_1 - y)$$

1. **Error (loss) derivative with respect to output:**

$$\frac{\partial E}{\partial o_1} = 2 \cdot (o_1 - y)$$

2. **Derivative of the sigmoid function at the output:**

$$\frac{\partial o_1}{\partial z_o} = o_1 \cdot (1 - o_1)$$

3. **Error (loss) with respect to z_o:**

$$\delta_o = \frac{\partial E}{\partial o_1} \cdot \frac{\partial o_1}{\partial z_o}$$

Derivations

1. Error (loss) derivative with respect to output

Given the MSE loss:

$$E = \frac{1}{2}(o - y)^2$$

We need to find the derivative of E with respect to the output o:

$$\frac{\partial E}{\partial o} = \frac{\partial}{\partial o}\left(\frac{1}{2}(o - y)^2\right)$$

Using the chain rule and power rule:

$$\frac{\partial E}{\partial o} = \frac{1}{2} \cdot 2 \cdot (o - y) = (o - y)$$

Including the factor of 2:

$$\frac{\partial E}{\partial o_1} = 2 \cdot (o_1 - y)$$

2. Derivative of the sigmoid function at the output

The sigmoid function is defined as:

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

To find the derivative of the sigmoid function, we use the chain rule:

$$\sigma'(z) = \frac{d}{dz}\left(\frac{1}{1+e^{-z}}\right)$$

Rewriting the sigmoid function:

$$\sigma(z) = (1+e^{-z})^{-1}$$

Using the chain rule and power rule:

$$\sigma'(z) = -1 \cdot (1+e^{-z})^{-2} \cdot \frac{d}{dz}(1+e^{-z})$$

Differentiate the inner function:

$$\frac{d}{dz}(1+e^{-z}) = 0 + (-e^{-z}) \cdot (-1) = e^{-z}$$

Substitute back:

$$\sigma'(z) = \frac{e^{-z}}{(1+e^{-z})^2}$$

Simplify using $\sigma(z) = \frac{1}{1+e^{-z}}$:

$$e^{-z} = \frac{1-\sigma(z)}{\sigma(z)}$$

Substitute:

$$\sigma'(z) = \frac{\frac{1-\sigma(z)}{\sigma(z)}}{(1+e^{-z})^2} = \frac{1-\sigma(z)}{\sigma(z)(1+e^{-z})^2}$$

Using $1+e^{-z} = \frac{1}{\sigma(z)}$:

$$\sigma'(z) = \frac{1-\sigma(z)}{\sigma(z)\left(\frac{1}{\sigma(z)}\right)^2} = \frac{1-\sigma(z)}{\sigma(z)} \cdot \sigma(z)^2 = (1-\sigma(z))\sigma(z)$$

Thus:

$$\sigma'(z) = \sigma(z)(1-\sigma(z))$$

3. Error (loss) with respect to z_o

To find the error with respect to z_o, we combine the derivatives using the chain rule:

$$\delta_o = \frac{\partial E}{\partial o_1} \cdot \frac{\partial o_1}{\partial z_o}$$

Substitute the expressions derived in steps 1 and 2:

$$\delta_o = 2 \cdot (o_1 - y) \cdot o_1 \cdot (1 - o_1)$$

Weight and Bias Updates from Hidden to Output Layer

Weight Update Derivation

The weight update rule is given by:

$$w_{oi} \leftarrow w_{oi} - \eta \cdot \delta_o \cdot h_i$$

1. **Gradient of the Loss with Respect to Weights:**

$$z_o = \sum_i w_{oi} \cdot h_i + b_o$$

2. **Gradient of z_o with Respect to the Weights w_{oi}:**

$$\frac{\partial z_o}{\partial w_{oi}} = h_i$$

3. **Gradient of the Loss with Respect to the Weights w_{oi}:**

$$\frac{\partial E}{\partial w_{oi}} = \frac{\partial E}{\partial z_o} \cdot \frac{\partial z_o}{\partial w_{oi}} = \delta_o \cdot h_i$$

4. **Update Rule for Weights:**

$$w_{oi} \leftarrow w_{oi} - \eta \cdot \frac{\partial E}{\partial w_{oi}} = w_{oi} - \eta \cdot \delta_o \cdot h_i$$

Bias Update Derivation

The bias update rule is given by:

$$b_o \leftarrow b_o - \eta \cdot \delta_o$$

1. **Gradient of z_o with Respect to the Bias b_o:**

$$\frac{\partial z_o}{\partial b_o} = 1$$

2. **Gradient of the Loss with Respect to the Bias b_o:**

$$\frac{\partial E}{\partial b_o} = \frac{\partial E}{\partial z_o} \cdot \frac{\partial z_o}{\partial b_o} = \delta_o$$

3. **Update Rule for Bias:**

$$b_o \leftarrow b_o - \eta \cdot \frac{\partial E}{\partial b_o} = b_o - \eta \cdot \delta_o$$