Activation Functions in Neural Networks

1. Sigmoid Activation Function

The Sigmoid activation function maps input values into the range (0, 1). It is defined as:

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

Step-by-Step Calculation

For an input $z = 0.5$:

$$f(0.5) = \frac{1}{1 + e^{-0.5}} \approx \frac{1}{1 + 0.6065} \approx 0.6225$$

Derivative of Sigmoid

The derivative of the Sigmoid function is:

$$\frac{\partial f(z)}{\partial z} = f(z) \cdot (1 - f(z))$$

Substituting $f(z) = 0.6225$:

$$\frac{\partial f(z)}{\partial z} \approx 0.6225 \cdot (1 - 0.6225) \approx 0.2350$$

2. ReLU (Rectified Linear Unit) Activation Function

The ReLU activation function outputs the input directly if it is positive, otherwise, it outputs zero:

$$f(z) = \max(0, z)$$

Step-by-Step Calculation

For an input $z = 0.5$:

$$f(0.5) = \max(0, 0.5) = 0.5$$

Derivative of ReLU

The derivative of the ReLU function is:

$$\frac{\partial f(z)}{\partial z} = \begin{cases} 1, \text{if } z > 0 \\ 0, \text{otherwise} \end{cases}$$

For $z = 0.5$:

$$\frac{\partial f(0.5)}{\partial z} = 1$$

3. Tanh Activation Function

The Tanh activation function outputs values between -1 and 1. It is defined as:

$$f(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

Step-by-Step Calculation

For an input $z = 0.5$:

$$f(0.5) = \frac{e^{0.5} - e^{-0.5}}{e^{0.5} + e^{-0.5}} \approx \frac{1.6487 - 0.6065}{1.6487 + 0.6065} \approx 0.4621$$

Derivative of Tanh

The derivative of the Tanh function is:

$$\frac{\partial f(z)}{\partial z} = 1 - f(z)^2$$

Substituting $f(0.5) = 0.4621$:

$$\frac{\partial f(0.5)}{\partial z} = 1 - 0.4621^2 \approx 1 - 0.2135 \approx 0.7865$$

4. Softmax Activation Function

The Softmax activation function is commonly used for multi-class classification. It converts logits into probabilities. It is defined as:

$$f_i(z) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

Step-by-Step Calculation

For a set of inputs $z_1 = 1, z_2 = 2, z_3 = 0$:

$$f_1 = \frac{e^1}{e^1 + e^2 + e^0} = \frac{2.718}{2.718 + 7.389 + 1} \approx 0.2368$$
$$f_2 = \frac{e^2}{e^1 + e^2 + e^0} = \frac{7.389}{2.718 + 7.389 + 1} \approx 0.6439$$
$$f_3 = \frac{e^0}{e^1 + e^2 + e^0} = \frac{1}{2.718 + 7.389 + 1} \approx 0.1193$$

Derivative of Softmax

The derivative of the Softmax function is more complex, as it involves partial derivatives. It is given by:

$$\frac{\partial f_i(z)}{\partial z_k} = f_i(z) \cdot (\delta_{ik} - f_k(z))$$

Here, δ_{ik} is the Kronecker delta, which is 1 if $i = k$ and 0 otherwise.

5. Leaky ReLU Activation Function

The Leaky ReLU activation function is similar to ReLU but allows a small, non-zero gradient when the input is negative:

$$f(z) = \begin{cases} z, & \text{if } z > 0 \\ 0.01 \cdot z, & \text{otherwise} \end{cases}$$

Step-by-Step Calculation

For an input $z = -0.5$:

$$f(-0.5) = 0.01 \cdot -0.5 = -0.005$$

Derivative of Leaky ReLU

The derivative of the Leaky ReLU function is:

$$\frac{\partial f(z)}{\partial z} = \begin{cases} 1, & \text{if } z > 0 \\ 0.01, & \text{otherwise} \end{cases}$$

For $z = -0.5$:

$$\frac{\partial f(-0.5)}{\partial z} = 0.01$$

6. Swish Activation Function

The Swish activation function is defined as:

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

Step-by-Step Calculation

For an input $z = 0.5$:

$$f(0.5) = 0.5 \cdot \frac{1}{1 + e^{-0.5}} \approx 0.5 \cdot 0.6225 \approx 0.3113$$

Derivative of Swish

The derivative of the Swish function is:

$$\frac{\partial f(z)}{\partial z} = f(z) + \sigma(z) \cdot (1 - f(z))$$

Where $\sigma(z) = \frac{1}{1 + e^{-z}}$ is the Sigmoid function. Substituting $z = 0.5$:

$$\sigma(0.5) = 0.6225, \quad f(0.5) = 0.3113$$
$$\frac{\partial f(0.5)}{\partial z} \approx 0.3113 + 0.6225 \cdot (1 - 0.3113) \approx 0.3113 + 0.4284 \approx 0.7397$$