

Derivatives of Common Activation Functions

Questions

1. Sigmoid Function

The sigmoid (logistic) function is defined as:

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

- (a) Find $\frac{d}{dx}\sigma(x)$ using the chain rule.
- (b) Show that this derivative can be written in the form

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)).$$

2. Hyperbolic Tangent (Tanh)

The hyperbolic tangent function is defined as:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

- (a) Find $\frac{d}{dx}\tanh(x)$ using the quotient rule.
- (b) Show that this derivative can be written as

$$\frac{d}{dx}\tanh(x) = 1 - \tanh^2(x).$$

3. ReLU (Rectified Linear Unit)

The ReLU function is defined as:

$$\text{ReLU}(x) = \begin{cases} 0, & x < 0, \\ x, & x \geq 0. \end{cases}$$

- (a) Find $\frac{d}{dx}\text{ReLU}(x)$ for $x < 0$ and $x > 0$.
- (b) Discuss what happens to the derivative at $x = 0$ and what is commonly used in deep learning practice.

Step-by-Step Solutions

1. Sigmoid Function

We are given

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

(a) Compute $\frac{d}{dx}\sigma(x)$

First, rewrite the function in a form suitable for the chain rule:

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

Let

$$u(x) = 1 + e^{-x} \Rightarrow \sigma(x) = u(x)^{-1}.$$

Now apply the chain rule:

$$\frac{d}{dx}\sigma(x) = \frac{d}{dx}(u^{-1}) = -u^{-2} \cdot \frac{du}{dx}.$$

Compute $\frac{du}{dx}$:

$$u(x) = 1 + e^{-x} \Rightarrow \frac{du}{dx} = 0 + \frac{d}{dx}(e^{-x}) = -e^{-x}.$$

Therefore,

$$\begin{aligned} \frac{d}{dx}\sigma(x) &= -u^{-2} \cdot (-e^{-x}) \\ &= \frac{e^{-x}}{(1 + e^{-x})^2}. \end{aligned}$$

So,

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

(b) Show that $\sigma'(x) = \sigma(x)(1 - \sigma(x))$

Recall:

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

Then

$$\begin{aligned} 1 - \sigma(x) &= 1 - \frac{1}{1 + e^{-x}} \\ &= \frac{1 + e^{-x}}{1 + e^{-x}} - \frac{1}{1 + e^{-x}} \\ &= \frac{e^{-x}}{1 + e^{-x}}. \end{aligned}$$

Now compute $\sigma(x)(1 - \sigma(x))$:

$$\begin{aligned}\sigma(x)(1 - \sigma(x)) &= \frac{1}{1 + e^{-x}} \cdot \frac{e^{-x}}{1 + e^{-x}} \\ &= \frac{e^{-x}}{(1 + e^{-x})^2}.\end{aligned}$$

But this is exactly the expression we found for $\sigma'(x)$:

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

Therefore,

$$\boxed{\sigma'(x) = \sigma(x)(1 - \sigma(x))}.$$

2. Hyperbolic Tangent (Tanh)

We are given

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

Let

$$N(x) = e^x - e^{-x}, \quad D(x) = e^x + e^{-x},$$

so that

$$\tanh(x) = \frac{N(x)}{D(x)}.$$

(a) Compute $\frac{d}{dx} \tanh(x)$ using the quotient rule

The quotient rule states:

$$\frac{d}{dx} \left(\frac{N}{D} \right) = \frac{N'D - ND'}{D^2}.$$

First compute $N'(x)$ and $D'(x)$:

$$N(x) = e^x - e^{-x} \Rightarrow N'(x) = e^x + e^{-x},$$

$$D(x) = e^x + e^{-x} \Rightarrow D'(x) = e^x - e^{-x}.$$

Now apply the quotient rule:

$$\begin{aligned}\frac{d}{dx} \tanh(x) &= \frac{N'(x)D(x) - N(x)D'(x)}{D(x)^2} \\ &= \frac{(e^x + e^{-x})(e^x + e^{-x}) - (e^x - e^{-x})(e^x - e^{-x})}{(e^x + e^{-x})^2}.\end{aligned}$$

Simplify the numerator term-by-term:

First product:

$$(e^x + e^{-x})^2 = e^{2x} + 2 + e^{-2x}.$$

Second product:

$$(e^x - e^{-x})^2 = e^{2x} - 2 + e^{-2x}.$$

Now subtract:

$$\begin{aligned}(e^x + e^{-x})^2 - (e^x - e^{-x})^2 &= (e^{2x} + 2 + e^{-2x}) - (e^{2x} - 2 + e^{-2x}) \\ &= e^{2x} + 2 + e^{-2x} - e^{2x} + 2 - e^{-2x} \\ &= 4.\end{aligned}$$

Thus,

$$\frac{d}{dx} \tanh(x) = \frac{4}{(e^x + e^{-x})^2}.$$

So we have

$$\tanh'(x) = \frac{4}{(e^x + e^{-x})^2}.$$

(b) Show that $\tanh'(x) = 1 - \tanh^2(x)$

We know:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

Then

$$\tanh^2(x) = \left(\frac{e^x - e^{-x}}{e^x + e^{-x}} \right)^2 = \frac{(e^x - e^{-x})^2}{(e^x + e^{-x})^2} = \frac{e^{2x} - 2 + e^{-2x}}{(e^x + e^{-x})^2}.$$

Now compute:

$$\begin{aligned}1 - \tanh^2(x) &= 1 - \frac{e^{2x} - 2 + e^{-2x}}{(e^x + e^{-x})^2} \\ &= \frac{(e^x + e^{-x})^2}{(e^x + e^{-x})^2} - \frac{e^{2x} - 2 + e^{-2x}}{(e^x + e^{-x})^2} \\ &= \frac{(e^x + e^{-x})^2 - (e^{2x} - 2 + e^{-2x})}{(e^x + e^{-x})^2}.\end{aligned}$$

Using the expansions again:

$$(e^x + e^{-x})^2 = e^{2x} + 2 + e^{-2x},$$

so the numerator becomes

$$\begin{aligned}(e^x + e^{-x})^2 - (e^{2x} - 2 + e^{-2x}) &= (e^{2x} + 2 + e^{-2x}) - (e^{2x} - 2 + e^{-2x}) \\ &= 4.\end{aligned}$$

Therefore,

$$1 - \tanh^2(x) = \frac{4}{(e^x + e^{-x})^2}.$$

But we already found:

$$\tanh'(x) = \frac{4}{(e^x + e^{-x})^2}.$$

Hence,

$$\boxed{\tanh'(x) = 1 - \tanh^2(x)}.$$

3. ReLU (Rectified Linear Unit)

The ReLU function is defined piecewise:

$$\text{ReLU}(x) = \begin{cases} 0, & x < 0, \\ x, & x \geq 0. \end{cases}$$

(a) Compute $\frac{d}{dx}\text{ReLU}(x)$ for $x < 0$ and $x > 0$

For $x < 0$:

$$\text{ReLU}(x) = 0 \Rightarrow \frac{d}{dx}\text{ReLU}(x) = 0.$$

For $x > 0$:

$$\text{ReLU}(x) = x \Rightarrow \frac{d}{dx}\text{ReLU}(x) = 1.$$

So, for $x \neq 0$,

$$\text{ReLU}'(x) = \begin{cases} 0, & x < 0, \\ 1, & x > 0. \end{cases}$$

(b) What happens at $x = 0$?

At $x = 0$, the function changes from 0 (for $x < 0$) to x (for $x > 0$). The left-hand derivative at $x = 0$ is

$$\lim_{h \rightarrow 0^-} \frac{\text{ReLU}(0 + h) - \text{ReLU}(0)}{h} = \lim_{h \rightarrow 0^-} \frac{0 - 0}{h} = 0,$$

and the right-hand derivative is

$$\lim_{h \rightarrow 0^+} \frac{\text{ReLU}(0 + h) - \text{ReLU}(0)}{h} = \lim_{h \rightarrow 0^+} \frac{h - 0}{h} = 1.$$

Since the left-hand and right-hand derivatives are not equal, the function is *not differentiable* at $x = 0$ in the strict mathematical sense.

However, in deep learning practice, it is common to define a *subgradient* at $x = 0$, and many implementations choose

$$\text{ReLU}'(0) = 0 \quad \text{or sometimes } 1,$$

for computational convenience.

We can summarize the practical derivative used in neural networks as:

$\text{ReLU}'(x) = \begin{cases} 0, & x < 0, \\ (\text{any value between } 0 \text{ and } 1), & x = 0 \text{ (subgradient),} \\ 1, & x > 0. \end{cases}$
--