

Objectives

- Math Review
 - Backpropagation
-

1 Math Review for Backpropagation

1.1 Function Types and Domains

- $f : \mathbb{R} \rightarrow \mathbb{R}$: Maps a number to a number (single variable).
- $f : \mathbb{R}^n \rightarrow \mathbb{R}$: Multivariable function.
- $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$: Multivariable vector-valued function.

1.2 Understanding the Derivative from First Principles

Definition of Derivative: Let $y = f(x)$.

$$y' = \frac{dy}{dx} = f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

Exercise 1: Find $f'(x)$ given the function $f(x) = x^2$

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} \\ &= \lim_{h \rightarrow 0} \frac{x^2 + 2xh + h^2 - x^2}{h} \\ &= \lim_{h \rightarrow 0} \frac{2xh + h^2}{h} \\ &= \lim_{h \rightarrow 0} \frac{h(2x + h)}{h} \\ &= \lim_{h \rightarrow 0} (2x + h) \\ &= 2x \end{aligned}$$

Exercise 2: Given $g(x) = C$, where C is a constant, find $g'(x)$.

$$\begin{aligned} g'(x) &= \lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{C - C}{h} \\ &= \lim_{h \rightarrow 0} \frac{0}{h} \\ &= 0 \end{aligned}$$

1.3 The Power Rule via Binomial Expansion

$$f(x) = x^2, f'(x) = nx^{n-1}$$

Exercise 3: Use the binomial expansion to find the derivative of $f(x) = x^n$.

The binomial coefficient is defined as:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad \text{where } 0! = 1.$$

For example:

$$\binom{n}{0} = \frac{n!}{0!(n-0)!} = \frac{n!}{n!} = 1.$$

These coefficients appear in Pascal's triangle, which lists the coefficients of the binomial expansion for increasing n . For $n = 0$ to $n = 4$:

$$\begin{array}{cccccc} & & & & & 1 & (n=0) \\ & & & & 1 & & 1 & (n=1) \\ & & 1 & & 2 & & 1 & (n=2) \\ & 1 & & 3 & & 3 & & 1 & (n=3) \\ 1 & & 4 & & 6 & & 4 & & 1 & (n=4) \end{array}$$

Using the binomial expansion, $(x + h)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} h^k$, we compute the derivative:

$$\begin{aligned}
 f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{(x+h)^n - x^n}{h} \\
 &= \lim_{h \rightarrow 0} \frac{\sum_{k=0}^n \binom{n}{k} x^{n-k} h^k - x^n}{h} \\
 &= \lim_{h \rightarrow 0} \frac{\binom{n}{0} x^n + \binom{n}{1} x^{n-1} h + \binom{n}{2} x^{n-2} h^2 + \cdots + \binom{n}{n} h^n - x^n}{h} \\
 &= \lim_{h \rightarrow 0} \frac{\binom{n}{1} x^{n-1} h + \binom{n}{2} x^{n-2} h^2 + \cdots + \binom{n}{n} h^n}{h} \\
 &= \lim_{h \rightarrow 0} \left(\binom{n}{1} x^{n-1} + \binom{n}{2} x^{n-2} h + \cdots + \binom{n}{n} h^{n-1} \right) \\
 &= \binom{n}{1} x^{n-1} \\
 &= nx^{n-1}.
 \end{aligned}$$

1.4 Product, Quotient, and Chain Rules

Let's say we have a product of two functions, then the derivative is given by the **Product rule**:

$$(f(x) \cdot g(x))' = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

Similarly, if we have a quotient of two functions, the **Quotient rule** gives:

$$\left(\frac{f(x)}{g(x)} \right)' = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$$

Final important rule is the **Chain Rule** which gives:

$$f(g(x)) = f'(g(x)) \cdot g'(x)$$

$$f(g(h(x))) = f'(g(h(x))) \cdot g'(h(x)) \cdot h'(x)$$

Exercise 4: Find $f'(x)$ given $f(x) = (3x^2 - 5x + 6)^3$

$$\begin{aligned}
 f(x) &= x^3, \quad f'(x) = 3x^2 \\
 g(x) &= 3x^2 - 5x + 6, \quad g'(x) = 6x - 5 \\
 &= 3(3x^2 - 5x + 6)^2 \cdot (6x - 5)
 \end{aligned}$$

Exercise 5: Find $f'(x)$ given the function $f(x) = \sqrt{\sin(x^2)}$

First, decompose it into a composition of three functions:

$$\begin{aligned}y &= \sqrt{v} = \sqrt{\sin(x^2)} \\v &= \sin(u) \\u &= x^2\end{aligned}$$

Now compute the derivative using the chain rule:

$$\begin{aligned}\frac{dy}{dv} &= \frac{1}{2}v^{-1/2} = \frac{1}{2}(\sin(x^2))^{-1/2}, \\ \frac{dv}{du} &= \cos(u) = \cos(x^2), \\ \frac{du}{dx} &= 2x.\end{aligned}$$

Applying the chain rule:

$$\frac{dy}{dx} = \frac{dy}{dv} \cdot \frac{dv}{du} \cdot \frac{du}{dx} = \frac{1}{2}(\sin(x^2))^{-1/2} \cdot \cos(x^2) \cdot 2x = \frac{x \cos(x^2)}{\sqrt{\sin(x^2)}}.$$

1.5 Multivariable Derivatives

For a function of multiple variables, such as:

$$f(x, y) = x^3 + 3y^2,$$

we can take partial derivatives with respect to each variable:

$$\frac{\partial f}{\partial x} = 3x^2, \quad \frac{\partial f}{\partial y} = 6y.$$

These partial derivatives together form the **gradient** of f , which is a vector that points in the direction of the greatest rate of increase of the function. The gradient is denoted by ∇f and defined as:

$$\nabla f(x, y) = \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle = \langle 3x^2, 6y \rangle.$$

So, the gradient of a scalar function $f(x, y)$ is a **vector-valued function** that contains all its first partial derivatives.

In multivariable calculus, when working with products of functions, the product rule still applies. For example, if

$$h(x, y) = f(x, y) \cdot g(x, y),$$

then the partial derivative with respect to x is:

$$\frac{\partial h}{\partial x} = \frac{\partial f}{\partial x} \cdot g(x, y) + f(x, y) \cdot \frac{\partial g}{\partial x}.$$

Exercise 6: Given $C(x, y) = xy^2 + \ln(\sin x)$, find ∇C .

$$\nabla C = \begin{bmatrix} \frac{\partial C}{\partial x} \\ \frac{\partial C}{\partial y} \end{bmatrix} = \begin{bmatrix} y^2 + \cot x \\ 2xy \end{bmatrix}$$

2 Backpropagation

These mathematical exercises are crucial for developing a solid understanding of backpropagation. The gradients of the loss function with respect to each weight are computed using partial derivatives. By iteratively applying the chain rule, backpropagation efficiently updates the model parameters to minimize the error.