Chain Rule

Mathematical Foundation

Chain Rule Formula

$$\frac{d}{dx} f(g(x)) = f'(g(x)) \cdot g'(x)$$

$$\text{Input } x \xrightarrow{g'(x)} \text{Inner } g(x) \xrightarrow{f'(g)} \text{Output } f(g)$$

The derivative flows backward:

$$\partial f/\partial x = \partial f/\partial g \times \partial g/\partial x$$

Step-by-Step Example

Compute: $d/dx \sin(x^2)$

1. $f(g) = \sin(g), \; g(x) = x^2$
2. $f'(g) = \cos(g) = \cos(x^2)$
3. $g'(x) = 2x$
4. $d/dx = \cos(x^2) \cdot 2x$

Applications in Neural Networks

Multi-layer Composition

Neural networks chain multiple functions through layers

$$y = f_3(f_2(f_1(x)))$$
$$\partial y/\partial x = \partial f_3/\partial f_2 \cdot \partial f_2/\partial f_1 \cdot \partial f_1/\partial x$$

Backpropagation

Gradients propagate backward through the network

$$\partial L/\partial w^{(l)} = \partial L/\partial z^{(l)} \cdot \partial z^{(l)}/\partial w^{(l)}$$
$$\text{where } z^{(l)} = w^{(l)}h^{(l-1)} + b^{(l)}$$

💡 **Key Insight**

The chain rule is the mathematical foundation of backpropagation. It allows us to compute gradients efficiently by decomposing complex derivatives into simpler parts, multiplying local gradients as we traverse backward through the computational graph.