Partial Derivatives and Chain Rule in Artificial Intelligence

Aymen Negadi

1 Derivation and Partial Derivatives

1.1 Definition of Derivation

In mathematics, a **derivative** measures how a function changes when its input changes.

Let $f(x)$ be a real-valued function. The derivative of f at a point x is defined as:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

1.2 Definition of Partial Derivatives

Let $f(x, y)$ be a function of two variables. The partial derivatives are defined as:

$$\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h}$$

$$\frac{\partial f}{\partial y} = \lim_{h \to 0} \frac{f(x, y+h) - f(x, y)}{h}$$

1.3 Why Derivatives Are Important in Artificial Intelligence

Derivatives are essential in Artificial Intelligence because learning algorithms optimize a loss function using gradient-based methods such as Gradient Descent.

- Updating model parameters
- Minimizing error functions
- Training neural networks

1.4 Example of Derivative Calculation

$$f(x) = 3x^2 + 2x + 1$$

$$f'(x) = 6x + 2$$

1.5 Example of Partial Derivative Calculation

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

$$\frac{\partial f}{\partial x} = 2x + 3y$$

$$\frac{\partial f}{\partial y} = 3x + 2y$$

1.6 Second Example of Partial Derivatives

Consider the following function:

$$w = e^x \cos y$$

1.6.1 Partial Derivative with Respect to x

To compute the partial derivative of w with respect to x, we consider y as a constant.

$$\frac{\partial w}{\partial x} = \frac{\partial}{\partial x} \left(e^x \cos y\right)$$

Since $\cos y$ is constant with respect to x, we obtain:

$$\frac{\partial w}{\partial x} = e^x \cos y$$

1.6.2 Partial Derivative with Respect to y

Now, we compute the partial derivative of w with respect to y, considering x as constant.

$$\frac{\partial w}{\partial y} = e^x \frac{\partial}{\partial y}(\cos y)$$

Knowing that $\frac{d}{dy}(\cos y) = -\sin y$, we obtain:

$$\frac{\partial w}{\partial y} = -e^x \sin y$$

Importance in Artificial Intelligence This example is particularly important in Artificial Intelligence because functions involving exponential and trigonometric terms such as e^x, $\sin(x)$, and $\cos(x)$ frequently appear in AI models.

These functions are commonly used in:

- **Activation functions**, where non-linearities are required to model complex patterns
- **Loss functions**, which measure the error between predictions and true values
- **Backpropagation**, where derivatives are computed to update model parameters efficiently

Therefore, mastering partial derivatives of such functions is essential for understanding and implementing learning algorithms in Artificial Intelligence.

2 Chain Rule

2.1 Definition of the Chain Rule

The **Chain Rule** is a fundamental rule of differentiation used when a variable depends on another variable, which itself depends on a third one. In other words, it applies to **composite functions** (functions of functions).

Let $y = f(u)$ and $u = g(x)$. Then y is a function of x through u, and the derivative of y with respect to x is given by:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

This rule allows us to decompose a complex derivative into simpler derivatives.

2.2 Why the Chain Rule Is Important in Machine Learning and Artificial Intelligence

The Chain Rule is essential in Machine Learning and Artificial Intelligence because most models are built as a sequence of nested functions.

In particular, it is the mathematical foundation of the **Backpropagation algorithm**, which is used to train neural networks.

The Chain Rule allows:

- Computing gradients of loss functions with respect to model parameters
- Propagating error signals backward through neural network layers
- Updating weights efficiently using optimization algorithms such as Gradient Descent

Without the Chain Rule, learning in deep neural networks would not be possible.

2.3 First Numerical Example

Consider the function:

$$y = (3x + 1)^2$$

Let:

$$u = 3x + 1 \quad \text{and} \quad y = u^2$$

We compute:

$$\frac{dy}{du} = 2u \quad \text{and} \quad \frac{du}{dx} = 3$$

Using the Chain Rule:

$$\frac{dy}{dx} = 2(3x + 1) \cdot 3 = 6(3x + 1)$$

At $x = 1$:

$$\frac{dy}{dx}(1) = 6(4) = 24$$

2.4 Second Numerical Example

Consider the function:

$$w = e^{x^2}$$

Let:

$$u = x^2 \quad \text{and} \quad w = e^u$$

We compute:

$$\frac{dw}{du} = e^u \quad \text{and} \quad \frac{du}{dx} = 2x$$

Applying the Chain Rule:

$$\frac{dw}{dx} = e^{x^2} \cdot 2x$$

At $x = 1$:

$$\frac{dw}{dx}(1) = 2e$$

2.5 Third Numerical Example

Consider the function:

$$y = \sin(\cos(\tan x))$$

This function is a composition of several functions. We define the intermediate variables as follows:

$$u = \tan x$$

$$v = \cos u$$

$$y = \sin v$$

We compute each derivative step by step:

$$\frac{dy}{dv} = \cos v$$

$$\frac{dv}{du} = -\sin u$$

$$\frac{du}{dx} = \sec^2 x$$

Using the Chain Rule, we obtain:

$$\frac{dy}{dx} = \frac{dy}{dv} \cdot \frac{dv}{du} \cdot \frac{du}{dx}$$

$$\frac{dy}{dx} = \cos(\cos(\tan x)) \cdot (-\sin(\tan x)) \cdot \sec^2 x$$

Therefore, the final result is:

$$y' = -\cos(\cos(\tan x)) \sin(\tan x) \sec^2 x$$