Differentiation

What is Differentiation?

- Differentiation is the process of finding the **derivative** of a function.
- The **derivative** tells us **how fast something is changing**.

Example: Speed of a Car

Imagine you're driving a car, and the **speedometer** shows your speed at every moment.

- If at **time = 5 seconds**, your speed is **30 km/h**,
- And at **time = 6 seconds**, your speed is **35 km/h**,
- Then your **rate of change of speed** is **(35 - 30) / (6 - 5) = 5 km/h per second**.

This rate of change is what **differentiation** calculates.

-

Key Idea: Calculus (specifically differentiation) is used in ML to optimize models.

- **Derivative**: Measures the **instantaneous rate of change** of a function.
  - *Example*: Speed is the derivative of distance with respect to time.

Why "Instantaneous"?

- It's the rate of change at an **exact point** (not an average over time).
- *Analogy*: **Your car's speedometer shows your speed *right now*, not your average speed for the trip.**

$$\lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

Relation to Slope:

💡 **The derivative is the slope of the tangent line to the function's curve at a point.**

- ***Example: For $y = x^2$, the slope at x=2 is 4.***

Example:

- If a hill has a steep **upward** slope, the derivative is **positive**.
- If the hill is going **downward**, the derivative is **negative**.
- If the hill is **flat**, the derivative is **zero**.

💡 **So, differentiation helps us understand whether a function is increasing, decreasing, or flat at a particular point.**

Derivative of a Constant

- **Rule**: The derivative of any constant (e.g., y=5) is **0**.
- **Why?** Constants don't change, so their rate of change is zero.

◦ *Example*: If $y = 7$, $\frac{dy}{dx} = 0$.

Maxima and Minima

- **Maxima**: The highest point of a curve (like the top of a hill).
- **Minima**: The lowest point of a curve (like the bottom of a valley).

At these points, the
derivative is zero because the slope is flat.

Mathematically,

$$f'(x) = 0$$

means a function has reached a maximum or minimum.

To confirm whether it's a maximum or minimum, we use the **second derivative**:

- If **f''(x) > 0**, it's a **minimum** (valley).
- If **f''(x)< 0**, it's a **maximum** (peak).

Common Derivatives Cheatsheet

- **Power Rule**:
  - *Example*: $\frac{d}{dx}(x^3) = 3x^2$.
- **Trigonometric Functions**:
  - $\frac{d}{dx}(\sin x) = \cos x$.
  - $\frac{d}{dx}(\cos x) = -\sin x$.
- **Exponentials/Logarithms**:
  - $\frac{d}{dx}(e^x) = e^x$.
  - $\frac{d}{dx}(\ln x) = \frac{1}{x}$.

Power Rule

$$\frac{d}{dx}(x^n) = nx^{n-1}$$

Example:

$$\frac{d}{dx}(x^3) = 3x^2$$

- Bring the exponent down in front.
- Subtract 1 from the exponent.

Sum Rule

- The derivative of a sum is just the sum of the derivatives.

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

$$\frac{d}{dx}(x^3+x^2) = 3x^2+2x$$

Product Rule

$$\frac{d}{dx}[f(x)g(x)] = f'(x)g(x) + f(x)g'(x)$$

Example:

If $f(x) = x^2$ and $g(x) = \sin x$ then:

$$\frac{d}{dx}(x^2 \sin x) = 2x \overset{\text{as is}}{\sin x} + \overset{\text{as is}}{x^2} \cos x$$

Quotient Rule

For division of two functions:

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

$$\frac{d}{dx}\left(\frac{x}{x+1}\right)=\frac{(1)(\cancel{x}+1)-\cancel{x}(1)}{(x+1)^2}=\frac{1}{(x+1)^2}$$

Chain Rule

Used when differentiating a function inside another function.

$$h(x)=f(g(x))$$

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

Example:

If f(x)= $\sin(x^2)$ then:

$$\frac{d}{dx}(\sin x^2) = \cos(x^2) \cdot 2x$$

Partial Differentiation

- When a function has **multiple variables**, we differentiate **one variable at a time** while keeping others constant.

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

$$\frac{\partial f}{\partial x} = 2x, \qquad \frac{\partial f}{\partial y} = 3y^2$$

- This is useful in **machine learning** and **physics**, where functions often depend on multiple variables.

Example: Temperature in a Room

Imagine the temperature **T** inside a room depends on:

- x = Distance from the heater
- y = Distance from the window

The temperature function might be:

$$T(x, y) = 5x^2 + 3y + 10$$

- Here, x and y both affect T.

Now, what if we only want to know how T changes with x while ignoring y? →That's where partial differentiation comes in!

Example: Let $f(x, y) = 3x^2y + 2xy^3$.

- **Partial derivative with respect to x:**

$$\frac{\partial f}{\partial x} = 6xy + 2y^3 \quad \text{(since } y \text{ is constant).}$$

- **Partial derivative with respect to y:**

$$\frac{\partial f}{\partial y} = 3x^2 + 6xy^2 \quad \text{(since } x \text{ is constant).}$$

Higher-Order Derivatives

- **First derivative** f′(x) → tells us slope.
- **Second derivative** f″(x) → tells us if the function is concave up or down.
- **Third derivative** f‴(x) → used in physics (jerk in motion).

Example:

$$f(x) = x^4, \quad f'(x) = 4x^3, \quad f''(x) = 12x^2, \quad f'''(x) = 24x$$

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

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

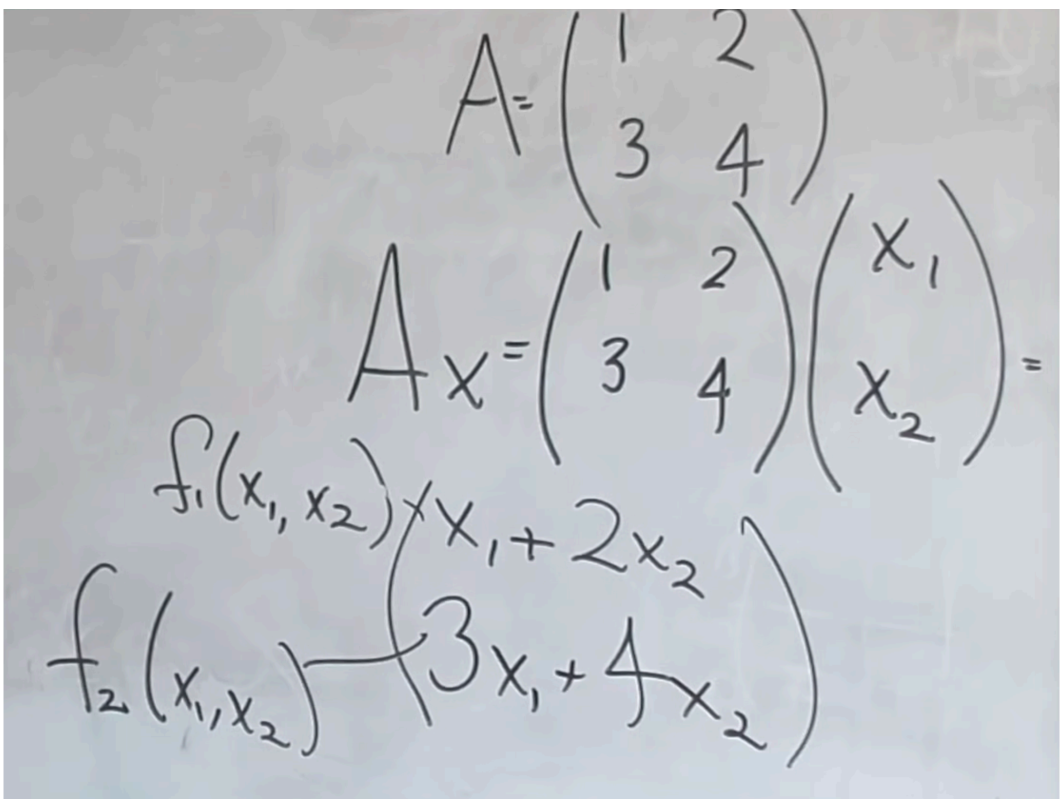
$$\frac{\partial^2 f}{\partial x^2} = 6x$$

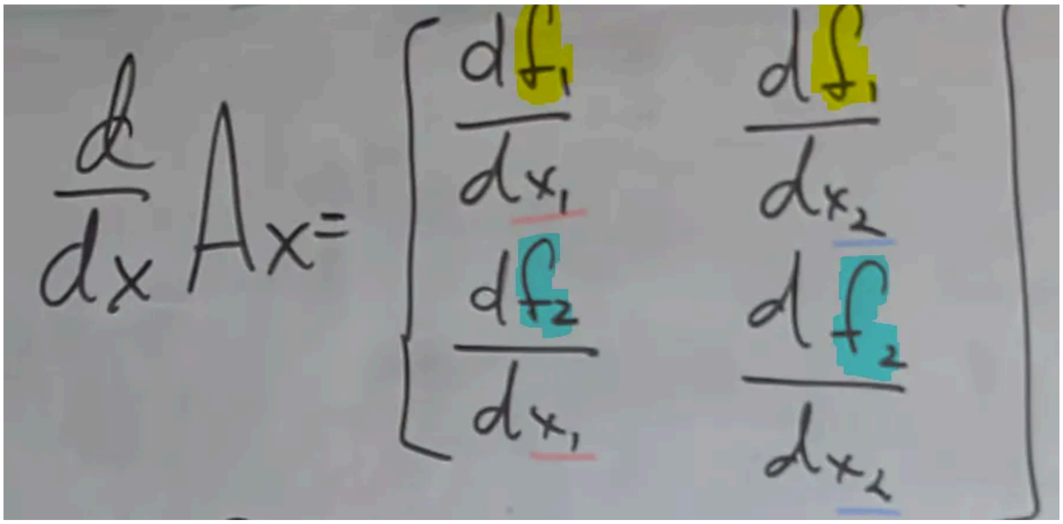
Matrix Differentiation

If $\mathbf{X}$ is a matrix and $\mathbf{w}$ is a vector, differentiation is done using special rules:

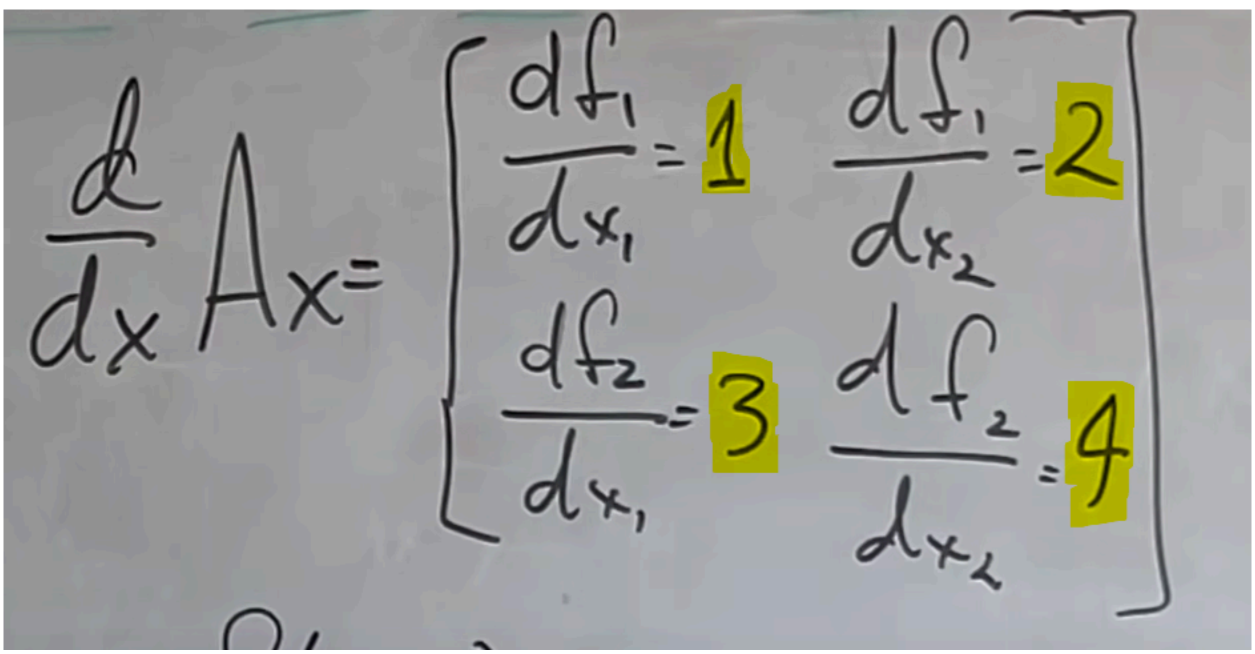
$$\frac{d}{d\mathbf{w}}(\mathbf{X}\mathbf{w}) = \mathbf{X}$$

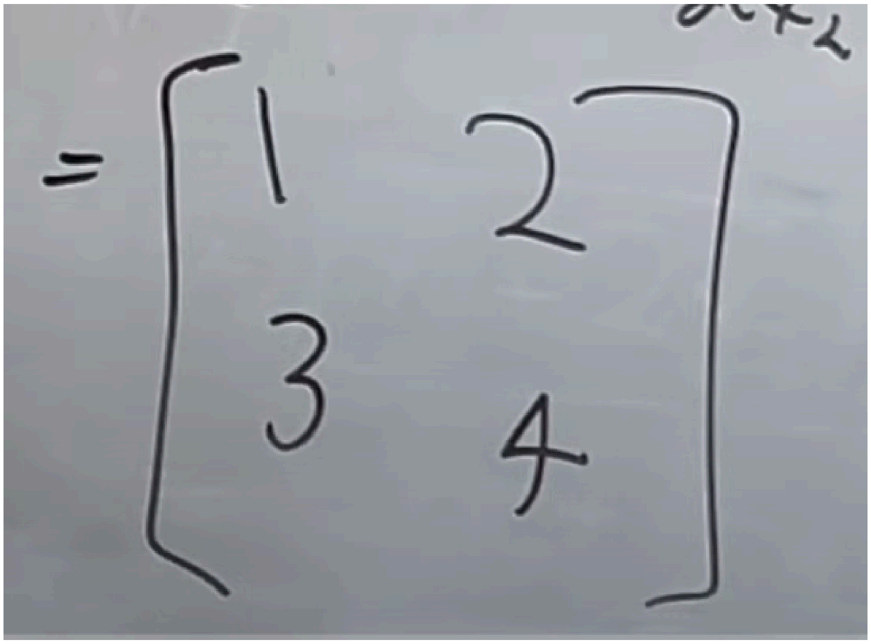
- Fundamental tool in machine learning, optimization, and deep learning



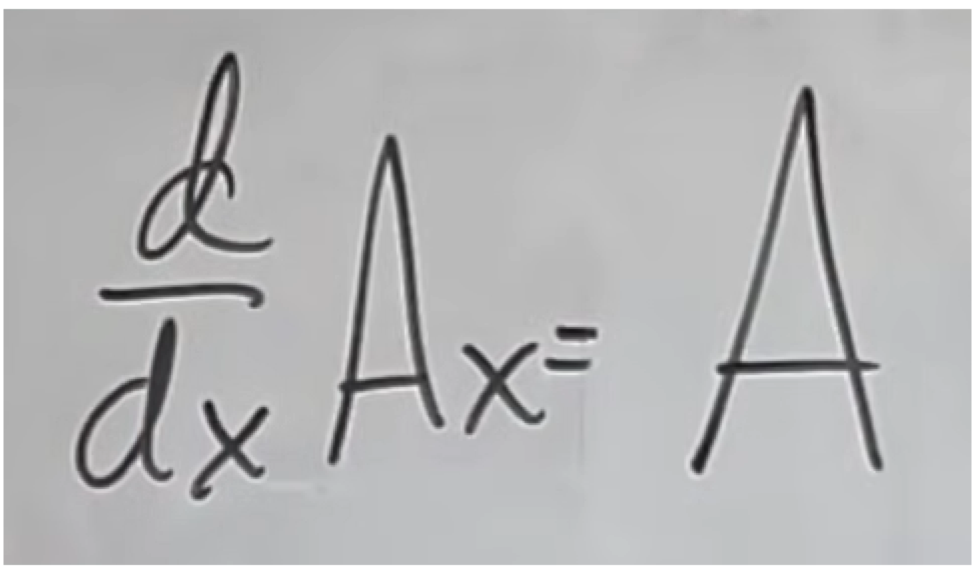


- We know f_1& f_2
- Find out the derivatives of them wrt x_1 & x_2





- 👆This is the original matrix itself.



If $\mathbf{X}$ is a matrix and $\mathbf{w}$ is a vector, differentiation is done using special rules:

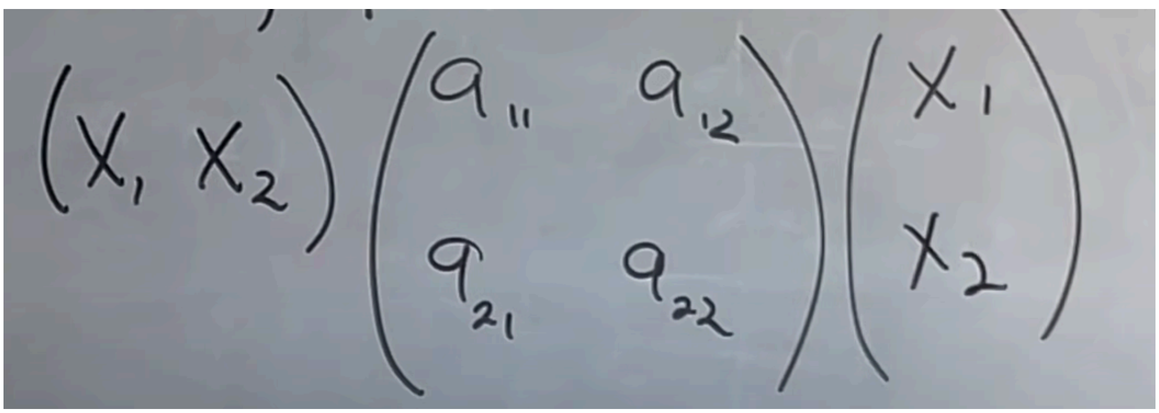
$$\frac{d}{d\mathbf{w}}(\mathbf{X}\mathbf{w}) = \mathbf{X}$$

This is heavily used in
gradient descent optimization.

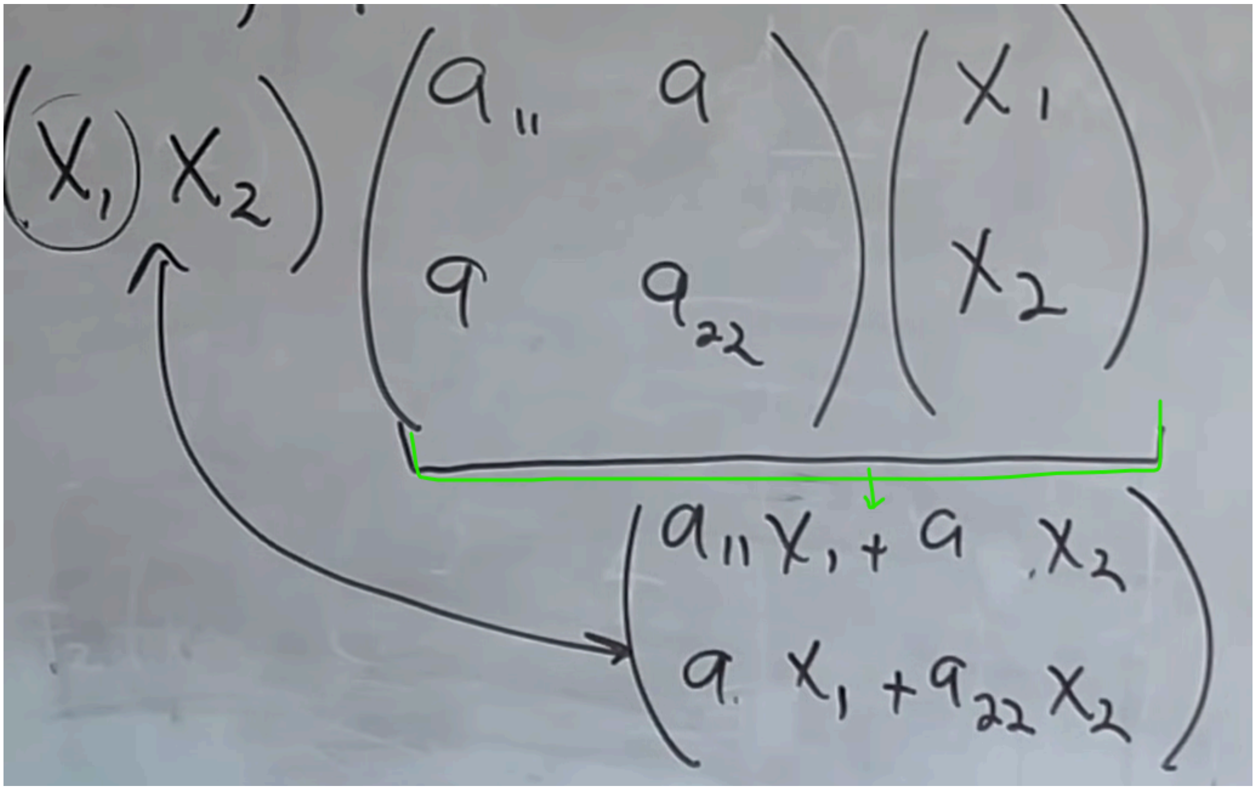
Little tougher example→

Derivative of:

$$y = X^T \underbrace{A X}$$



- First let's multiply A & X



We called a_{12}& a_{21} as→ a

- Now multiply this with (x_1, x_2)

$$a_{11}x_1^2 + ax_1x_2 + ax_1x_2 + a_{22}x_2^2$$

$$a_{11}x_1^2 + \underline{2ax_1x_2} + a_{22}x_2^2$$

- We'll call this👆 some function $f(x_1, x_2)$

$$= f(x_1, x_2)$$

$$\frac{d}{dx} x^T A x \qquad \begin{pmatrix} \frac{df}{dx_1} \\ \frac{df}{dx_2} \end{pmatrix}$$

$$\begin{pmatrix} 2a_{11}x_1 + 2a x_2 \\ 2a x_1 + 2a_{22} x_2 \end{pmatrix}$$

$$2\begin{pmatrix} a_1 x_1 + a x_2 \\ a x_1 + a_{22} x_2 \end{pmatrix}$$

$$2\begin{pmatrix} a_{11} & a \\ a & a_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= 2Ax$$

Therefore,

$$\frac{d}{dx} x^T A x = 2Ax$$

$$\boxed{2x^T A}$$