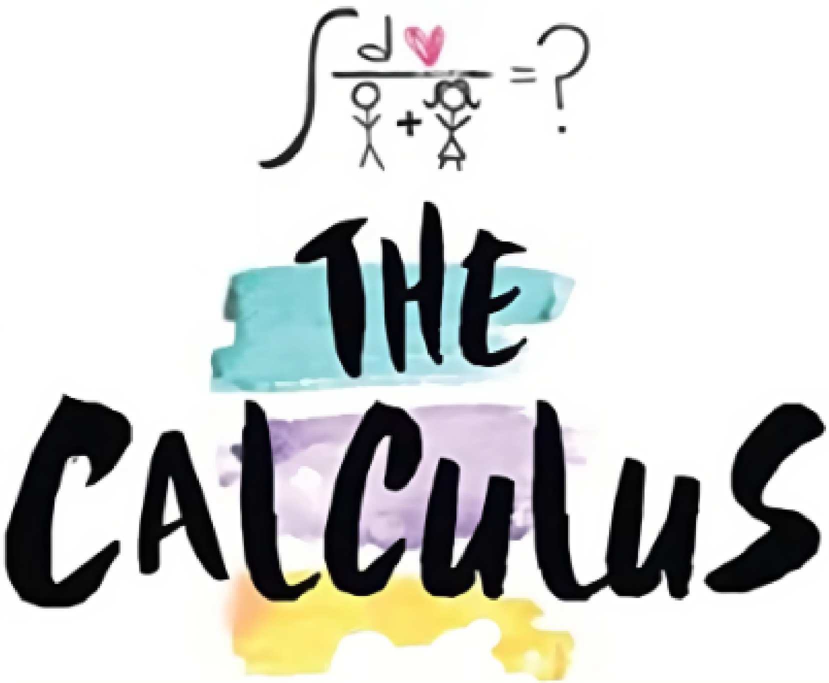
What is Differential calculus ?

Differential calculus is a branch of mathematics that studies the rates at which quantities change. It is one of the two traditional divisions of calculus, the other being integral calculus—the study of the area beneath a curve.

Differentiation is the process of finding the derivative of a function. The derivative of a function represents the **instantaneous rate of change of the function** with respect to its variable, typically denoted as 'x'. For example, the derivative of the function f(x) = x^2 is 2x, which means that the function f(x) increases by 2x for every unit increase in x.



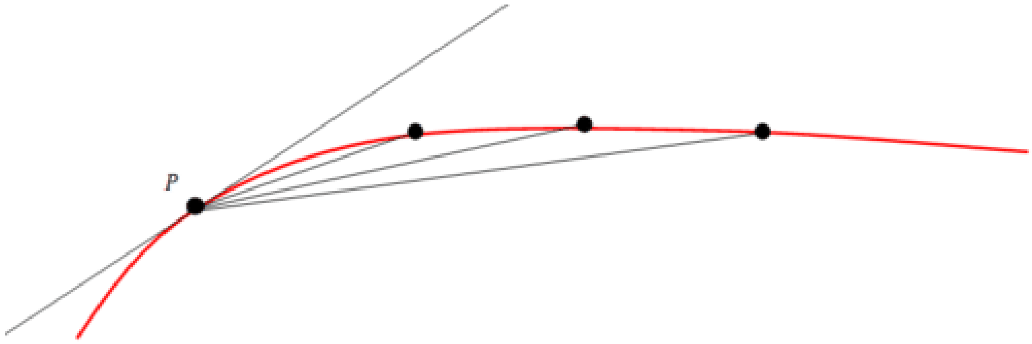
Here are some of the basic concepts of differential calculus:

- **Derivative:** The derivative of a function at a point is a measure of how much the function changes as its input changes at that point.
- **Differential equation:** A differential equation is an equation that contains derivatives. Differential equations are used to model a wide variety of phenomena, such as the motion of objects, the spread of diseases, and the growth of populations.
- **Integration:** Integration is the opposite of differentiation. It is the process of finding the function whose derivative is a given function.

Differential calculus is a powerful tool for analyzing and understanding the behavior of quantities that change over time or space. It is an essential tool for scientists, engineers, and other professionals who need to understand how things work.

Why instantaneous?

The term "instantaneous" in the context of differential calculus refers to the concept of measuring a quantity at an extremely small and specific point in time or space. It is used to describe the rate of change of a function at an exact moment or a specific point on a curve.



In calculus, we often deal with functions that represent continuously changing quantities, such as position, velocity, or temperature. To understand how these quantities are changing at any given point, we use the concept of the derivative. The derivative gives us the rate of change of a function with respect to its independent variable (usually denoted as x), and it can be interpreted as the slope of a tangent line to the graph of the function at a particular point.

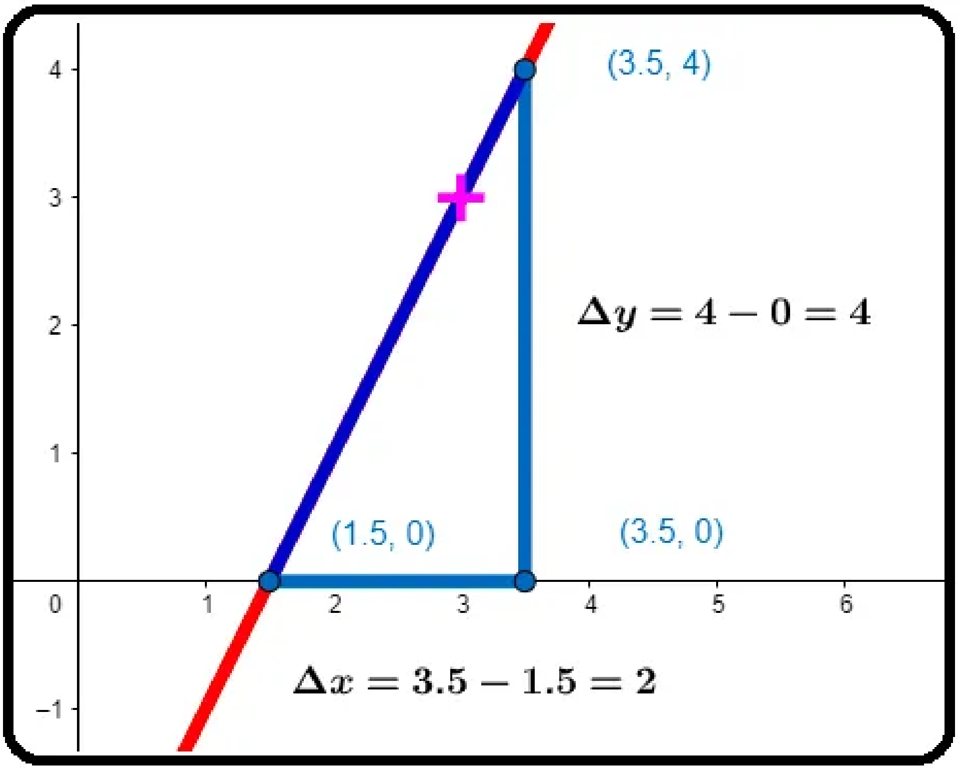
The term "instantaneous" is used because the derivative provides us with the rate of change at an infinitesimally small interval around a specific point. It captures the behavior of the function at that precise moment or location without considering the overall behavior of the function over larger intervals.

By studying instantaneous rates of change, we gain insight into how a function behaves locally and can make predictions about its behavior near a specific point. This concept of instantaneous rates of change is fundamental to differential calculus and is crucial for analyzing the behavior of functions in various contexts.

Relation with slope?

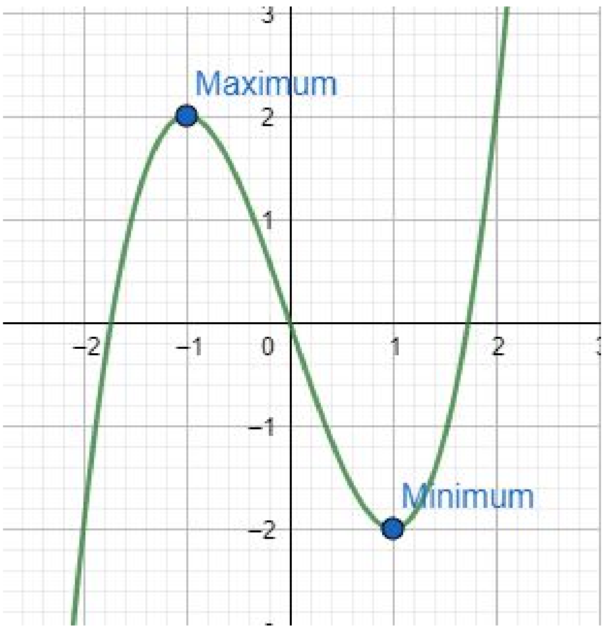
Differential calculus is closely related to the concept of slope. In fact, the derivative of a function at a specific point represents the slope of the tangent line to the graph of the function at that point.

Consider a function f(x) that represents a curve on a coordinate plane. The derivative of f(x) with respect to x, denoted as f'(x) or dy/dx, gives us the instantaneous rate of change of the function at any point x. This rate of change is equivalent to the slope of the tangent line to the graph of the function at that point.



Maxima and Minima

- **Maxima**: A maximum is the highest point on the graph of a function. It is a point where the function is increasing and then starts to decrease. The derivative of the function at a maximum is equal to 0.
- **Minima** :A minimum is the lowest point on the graph of a function. It is a point where the function is decreasing and then starts to increase. The derivative of the function at a minimum is equal to 0.



the key points about finding maxima and minima in calculus:

1. Identify the domain of the function.
2. Find critical points by setting the derivative equal to zero and solving for x.

3. Use the first derivative test to determine if each critical point is a local maximum or minimum.
4. Check the endpoints and boundaries of the function's domain for potential extrema.
5. Compare the values at critical points, endpoints, and boundaries to find the global maximum and minimum.

How to calculate derivative?

To calculate the derivative of a function, you can follow these general steps:

1. **Identify the function**: Let's say you have a function denoted as f(x).
2. **Express the derivative**: The derivative of f(x) with respect to x can be denoted as f'(x), dy/dx, or df/dx. These notations represent the rate of change of the function f(x) with respect to x.
3. **Apply differentiation rules**: Use various rules and techniques to differentiate the function f(x) with respect to x. Some commonly used rules include:
   - Power rule: If f(x) = **x^n, where n is a constant, then f'(x) = nx^(n-1).**

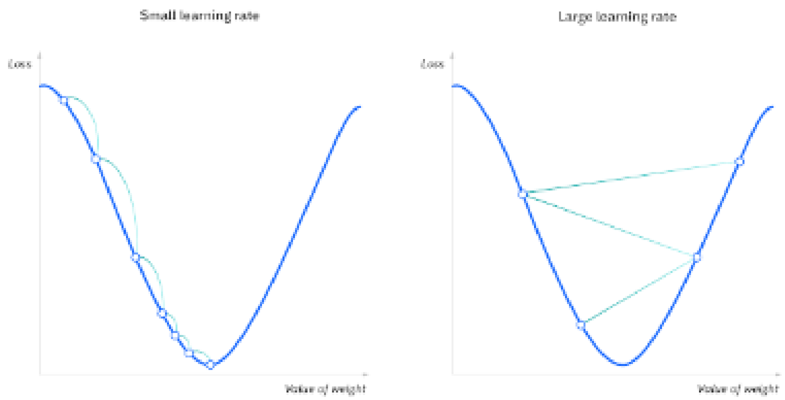
- Product rule: **If f(x) = u(x)v(x), then f'(x) = u'(x)v(x) + u(x)v'(x)**.
- Quotient rule: **If f(x) = u(x) / v(x), then f'(x) = [u'(x)v(x) - u(x)v'(x)] / [v(x)]^2**.
- Chain rule: **If f(x) = g(h(x)), then f'(x) = g'(h(x)) * h'(x), where g'(x) and h'(x) represent the derivatives of g(x) and h(x) respectively.**

4. **Simplify the derivative**: Apply algebraic simplification techniques to simplify the expression obtained after differentiation, if possible.
5. **Evaluate the derivative**: If you need to find the derivative at a specific point x=a, substitute the value of a into the derivative expression to calculate the numerical value of the derivative at that point.

how derivatives are used in machine learning?

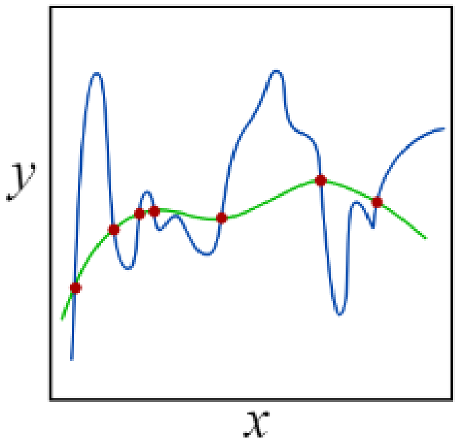
- **Gradient descent:**

Gradient descent is an optimization algorithm that uses derivatives to find the minimum of a loss function. The loss function is a measure of how well the model fits the training data. The gradient of the loss function is a vector that points in the direction of the steepest descent. Gradient descent works by iteratively moving in the direction of the gradient until it reaches a minimum.



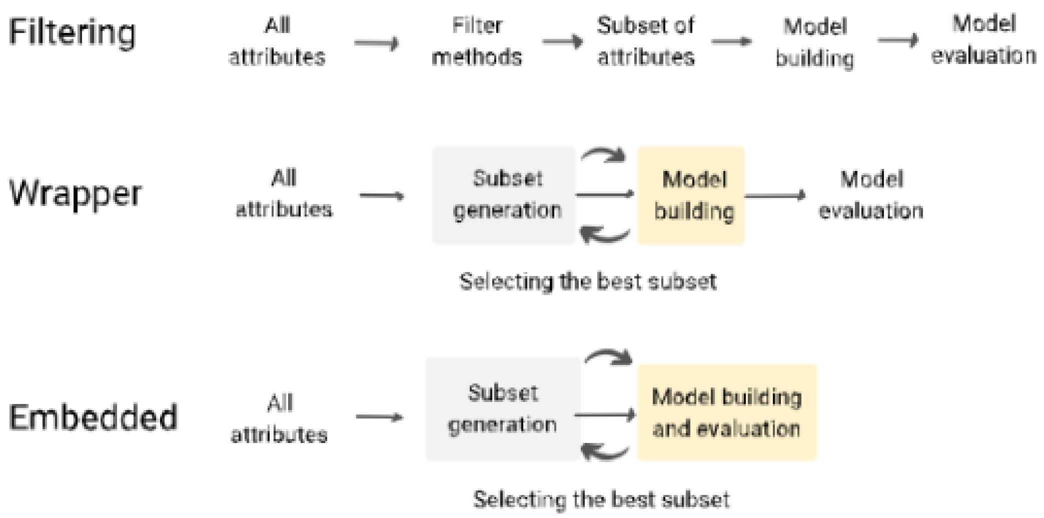
- **Regularization:**

Regularization is a technique that is used to prevent machine learning models from overfitting the training data. Overfitting occurs when the model learns the training data too well and is unable to generalize to new data. Regularization adds a penalty to the loss function that discourages the model from fitting the training data too closely. Derivatives can be used to calculate the penalty term.

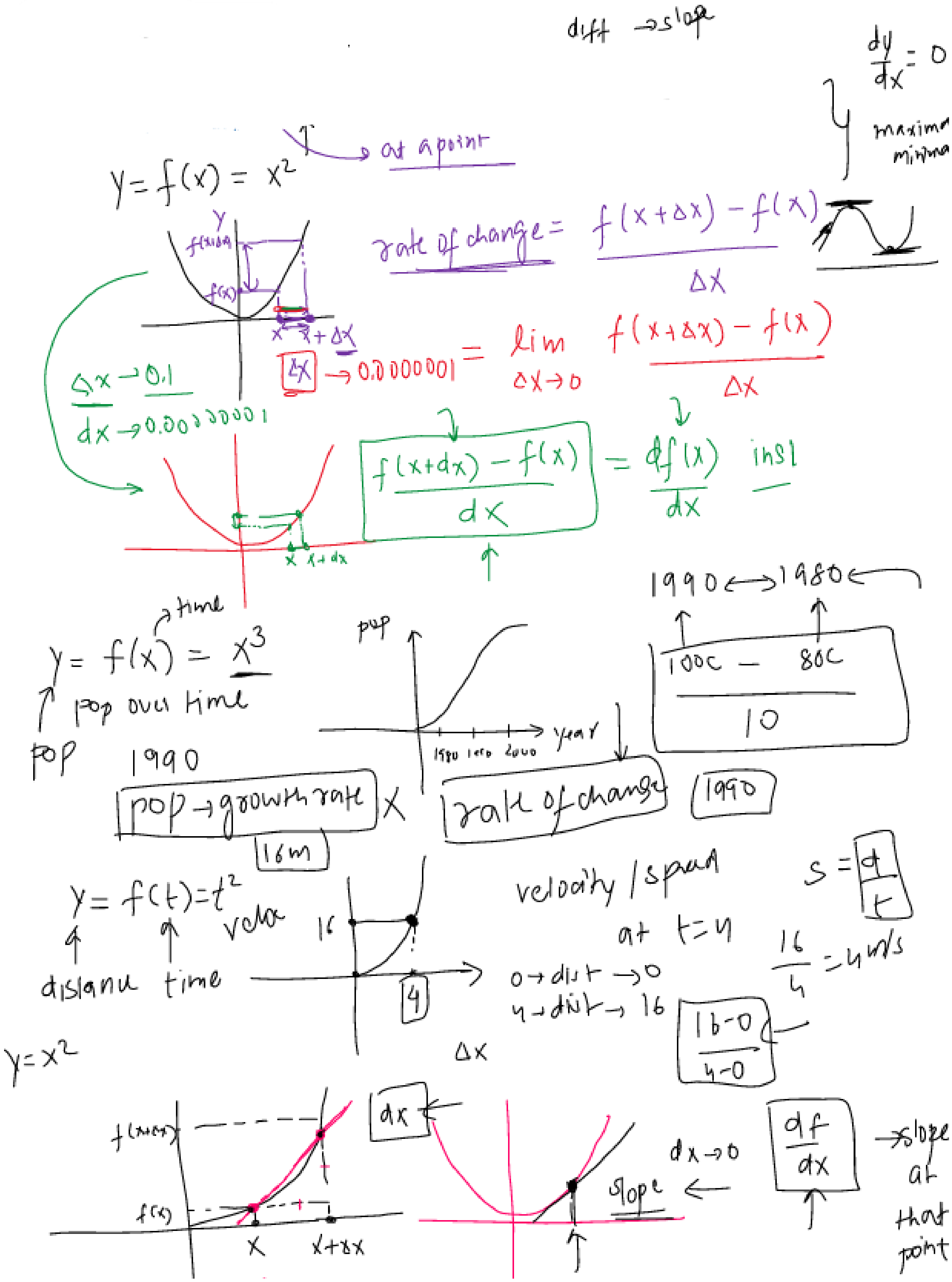


- **Feature selection:**

  Feature selection is the process of selecting a subset of features that are most relevant to the problem that the model is trying to solve. Derivatives can be used to calculate the importance of each feature. The features with the highest importance are the most likely to be relevant to the problem.


Filtering
All attributes
Filter methods
Subset of attributes
Model building
Model evaluation
Wrapper
All attributes
Subset generation
Model building
Model evaluation
Selecting the best subset
Embedded
All attributes
Subset generation
Model building and evaluation
Selecting the best subset

Explanation

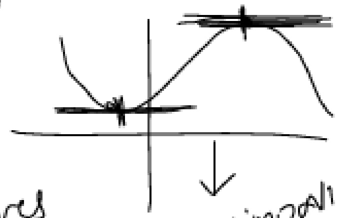


$x \quad x+\delta x$

point

diff $=0$

derivative



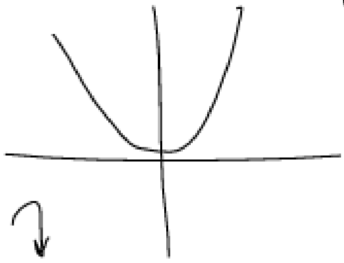
derivatives

optimization

$y = x^2 + 2x$

$\boxed{\frac{dy}{dx}} = 2x + 2 = \boxed{0} \qquad x = -1$

$x = -\frac{2}{2} = -1$



$x^2 \rightarrow \boxed{2x}$

$y = f(x) = x^2$

$\frac{df(x)}{df} = \frac{f(x+dx) - f(x)}{dx}$

$dx \rightarrow 0$

$f(x+dx) = (x+dx)^2 \qquad = \frac{(x+dx)^2 - x^2}{dx}$

$f(x) = x^2$

$= \frac{\cancel{x^2} + (dx)^2 + 2x\,dx - \cancel{x^2}}{dx}$

$\frac{df}{dx} = 2x$

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

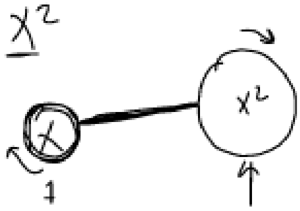
0

intuition

$\boxed{-2}$

$y = f(x) = x^2$

$\boxed{\frac{dy}{dx}}$



$L(\beta_0)$

$\frac{dL}{d\beta_0}$

$L \rightarrow \uparrow\uparrow \rightarrow \downarrow$

β_0 change

$\beta_0 \rightarrow \uparrow\downarrow$

random

Derivative of a constant?

Derivative of a Constant

$$\frac{d}{dx}\left[c\right] = 0$$

The derivative of a constant is always zero.

Let's consider a constant function f(x) = c, where c is a constant value. In this case, the function f(x) does not depend on the variable x and remains constant for all values of x.

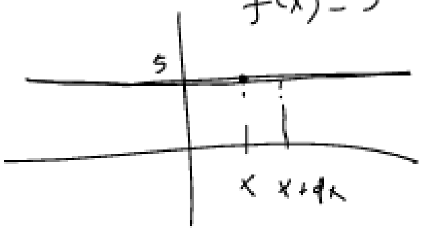
To find the derivative of f(x), we apply the differentiation rules. However, since the function does not change with respect to x, the derivative of a constant term is zero. Mathematically, we can express it as:

f'(x) = d/dx [c] = 0

So, the derivative of a constant function is always zero, regardless of the value of the constant. This result holds true for any constant value in the function.

$\frac{d}{dx}(c) = 0$ 0 $y = 5$

derivative of a const = 0



$f(x) = 5$ $f(x+dx) = 5$

$$\frac{dy}{dx} = \frac{f(x+dx) - f(x)}{dx}$$

$$\frac{dy}{dx} = 0 \quad \frac{5-5}{}$$

$$y = \boxed{5}$$

$$\frac{dy}{dx} = 0$$

COMMON DERIVATIVES

$$\frac{d}{dx}(x) = 1$$

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

$$\frac{d}{dx}(\cos x) = -\sin x$$

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

$$\frac{d}{dx}(\sec x) = \sec x \tan x$$

$$\frac{d}{dx}(\csc x) = -\csc x \cot x$$

$$\frac{d}{dx}(\cot x) = -\csc^2 x$$

$$\frac{d}{dx}(\sin^{-1} x) = \frac{1}{\sqrt{1 - x^2}}$$

$$\frac{d}{dx}(\cos^{-1} x) = -\frac{1}{\sqrt{1 - x^2}}$$

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

$$\frac{d}{dx}(a^x) = a^x \ln(a)$$

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

$$\frac{d}{dx}(\ln(x)) = \frac{1}{x}, x > 0$$

$$\frac{d}{dx}(\ln|x|) = \frac{1}{x}$$

$$\frac{d}{dx}(\log_a(x)) = \frac{1}{x \ln(a)}$$

What is Power Rule?

The power rule is a fundamental rule in calculus that allows us to find the derivative of a function that is raised to a power. Specifically, it states that if we have a function of the form f(x) = x^n, where n is a constant, then the derivative of f(x) with respect to x is given by:

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

f'(x) = n * x^(n-1)

Here, n is the exponent of x, and the derivative gives us the rate at which the function is changing with respect to x.

Let's look at a few examples to illustrate the power rule:

1. **f(x) = x^2:**

   Applying the power rule, we have:

   f'(x) = 2 * x^(2-1) = 2 * x^1 = 2x

2. **f(x) = x^3:**

   Applying the power rule, we have:

   f'(x) = 3 * x^(3-1) = 3 * x^2 = 3x^2

3. **f(x) = x^0:**

   Applying the power rule, we have:

   f'(x) = 0 * x^(0-1) = 0 * x^(-1) = 0

Note that the power rule applies only when the exponent is a constant. If the exponent is a function of x, then the power rule does not directly apply, and other rules like the chain rule may be needed to find the derivative.

The power rule is a fundamental tool in differential calculus and is used extensively in finding derivatives of polynomials, exponential functions, and many other types of functions.

The Power Rule Emerges

The binomial formula for any exponent 'n' is as follows:

$$(a+b)^n = \sum_{k=0}^{n} \left[\overbrace{{}^nC_k}^{\text{Binomial Coefficient}} a^{n-k} b^k \right]$$

When we plug in 'x' and 'dx' in place of 'a' and 'b' respectively, we get the following result:

$$y + dy = (x + dx)^n = {}^nC_0 x^n dx^0 + {}^nC_1 x^{n-1} dx^1 \underbrace{+ {}^nC_2 x^{n-2} (dx)^2 + \ldots + {}^nC_n x^0 (dx)^n}_{\text{Negligible}}$$

Discarding $(dx)^2$ and higher order terms, we get:

$$\Rightarrow y + dy = x^n + n x^{n-1} dx$$

Subtracting $y \; (= x^n)$ from both sides:

$$\Rightarrow dy = n x^{n-1} dx$$

Dividing both sides by dx:

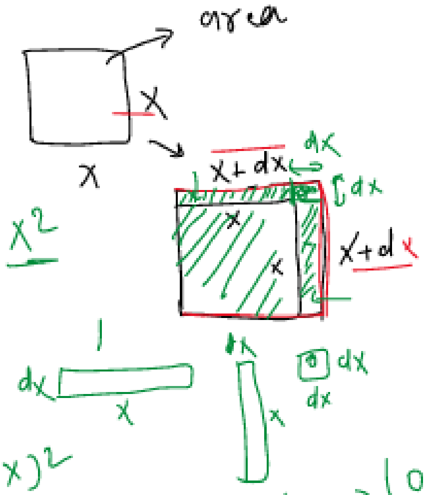
$$\Rightarrow \boxed{\frac{dy}{dx} = n x^{n-1}}$$

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

$y = \boxed{f(x) = x^2}$ x^2

$$\frac{dy}{dx} = \frac{f(x+dx) - f(x)}{dx}$$

x^2

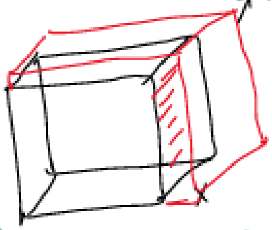


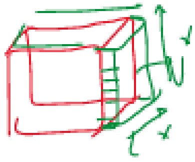
$$f(x+dx) = x^2 + x\,dx + x\,dx + (dx)^2$$

$$= x^2 + 2x\,dx$$

$dx \to (0.000001)^2$

$$\frac{\cancel{x^2} + 2x\,dx - \cancel{x^2}}{dx} = \frac{2x\,\cancel{dx}}{\cancel{dx}} = 2x$$





$$\frac{x^2dx}{x^2dx}$$

$y = f(x) = x^3$

$$\frac{dy}{dx} = \frac{f(x+dx) - f(x)}{dx}$$

$$\frac{\cancel{x^3} + x^2dx + x^2dx + x^2dx - \cancel{x^3}}{\cancel{dx}}$$

$$\frac{3x^2\,\cancel{dx}}{\cancel{dx}} = 3x^2$$

$$\frac{dy}{dx} = 3x^2$$

power rule

$x^4 \to 4x^3$

$x^5 \to 5x^4$

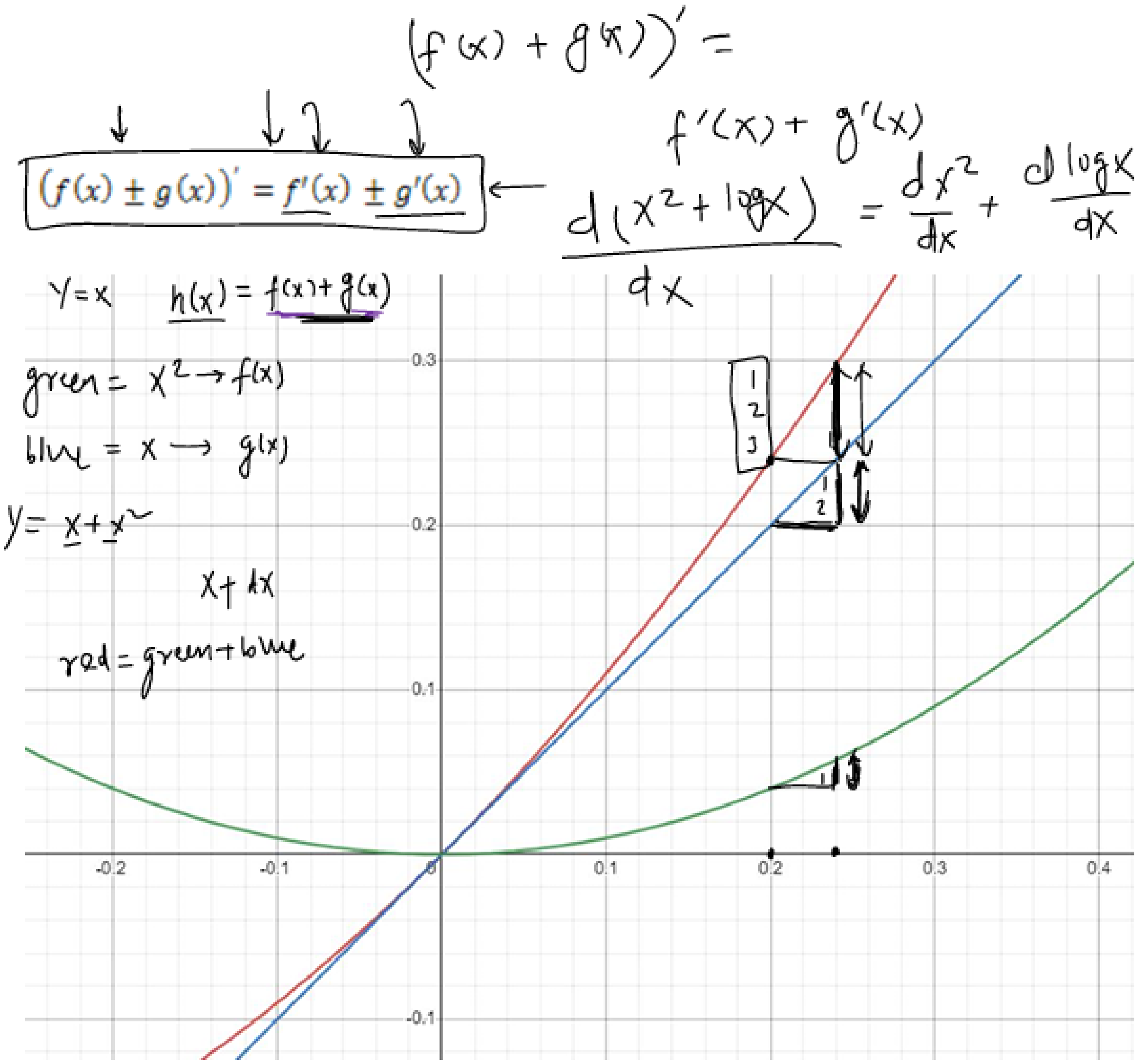
$x^6 \to 6x^5$

$x^n \to \boxed{nx^{n-1}}$

Sum Rule

The sum rule states that the derivative of the sum of two functions is the sum of the derivatives of the two functions. This means that if we have two functions, f(x) and g(x), then the derivative of **f(x) + g(x) is f'(x) + g'(x)**.

In other words, to **find the derivative of the sum of two functions**, we simply take the derivatives of each function separately and add them together.



Example to illustrate the sum rule:

- Suppose we have two functions:

```
f(x) = 2x^2 and g(x) = 3x + 1
```

- To find the derivative of their sum, we can apply the sum rule:

```
(d/dx) [f(x) + g(x)] = (d/dx) [2x^2 + (3x + 1)] = (d/dx) [2x^2] + (d/dx) [3x + 1]
```

- Now, we find the derivatives of each function individually:

```
f'(x) = (d/dx) [2x^2] = 4x
g'(x) = (d/dx) [3x + 1] = 3
```

- Finally, we add the derivatives together to obtain the derivative of the sum:

```
f'(x) + g'(x) = 4x + 3
```

- So, the derivative of the sum **f(x) + g(x) is 4x + 3** according to the sum rule.

The sum rule is a fundamental concept in calculus and is used in more complex cases where we have to find the derivative of a sum of multiple functions.

Product Rule

The product rule is a fundamental rule in calculus that allows us to **find the derivative of a product of two functions**. It is used when we have a function f(x) multiplied by another function g(x).

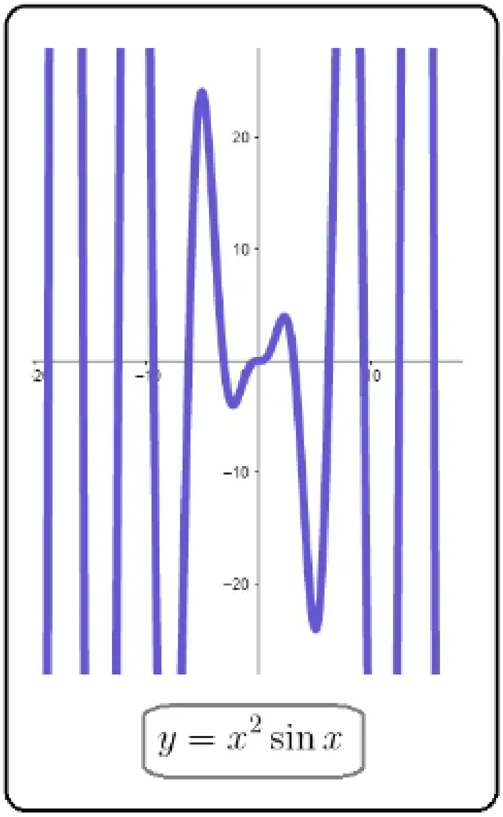
The product rule states that the derivative of the product of two functions is equal to the derivative of the first function multiplied by the second function, plus the first function multiplied by the derivative of the second function.

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

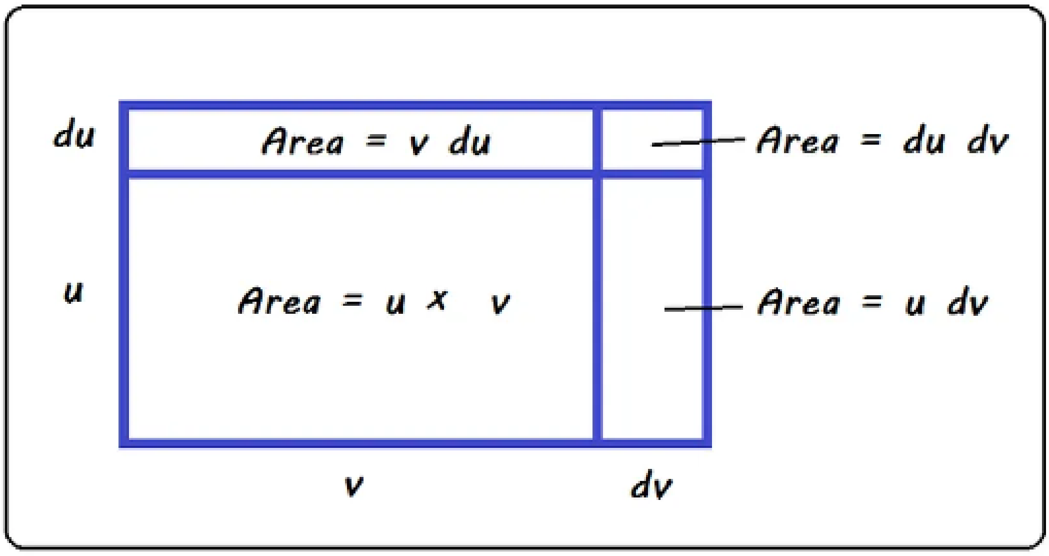
To apply the product rule, we follow these steps:

- Identify the two functions: Let's say we have f(x) and g(x).
- Calculate the derivatives of each function: Find the derivatives of f(x) and g(x) separately, denoted as f'(x) and g'(x), respectively.
- Apply the product rule: Plug in the derivatives into the product rule formula: f'(x) * g(x) + f(x) * g'(x).
- Simplify the expression if possible: If further algebraic simplification is possible, simplify the expression obtained from applying the product rule.

Suppose you’d like to multiply two functions:



Let’s take two generic functions, u and v, and represent them as sides of a rectangle. Their product is the area of the rectangle. The slope of a function tells us the rate at which that function is increasing or decreasing. We are interested in how the area changes as each of the sides of the rectangle change.



The increase in area is given by: v du + u dv + du dv. Suppose u and v are functions of x. We want to represent the change in the combined function, uv, as x changes. As we did with the trig functions, we can express this change as a ratio showing the slope of the function.

$$\text{Slope of } uv = \frac{d\ uv}{dx} = \frac{v\ du + u\ dv + du\ dv}{dx} = v\frac{du}{dx} + u\frac{dv}{dx}$$

The last product in the numerator, du dv, doesn't make the final cut. This is because the term (du dv)/dx comprises only infinitesimals. It has no effect on the result.

Example to illustrate the product rule:

- Suppose we have two functions:

```
f(x) = x^2 and g(x) = 3x + 1
```

- To find the derivative of their product, we can apply the product rule:

```
(d/dx) [f(x) * g(x)] = (d/dx) [x^2 * (3x + 1)]
```

- Now, we find the derivatives of each function individually:

```
f'(x) = (d/dx) [x^2] = 2x

g'(x) = (d/dx) [3x + 1] = 3
```

- Finally, we apply the product rule:

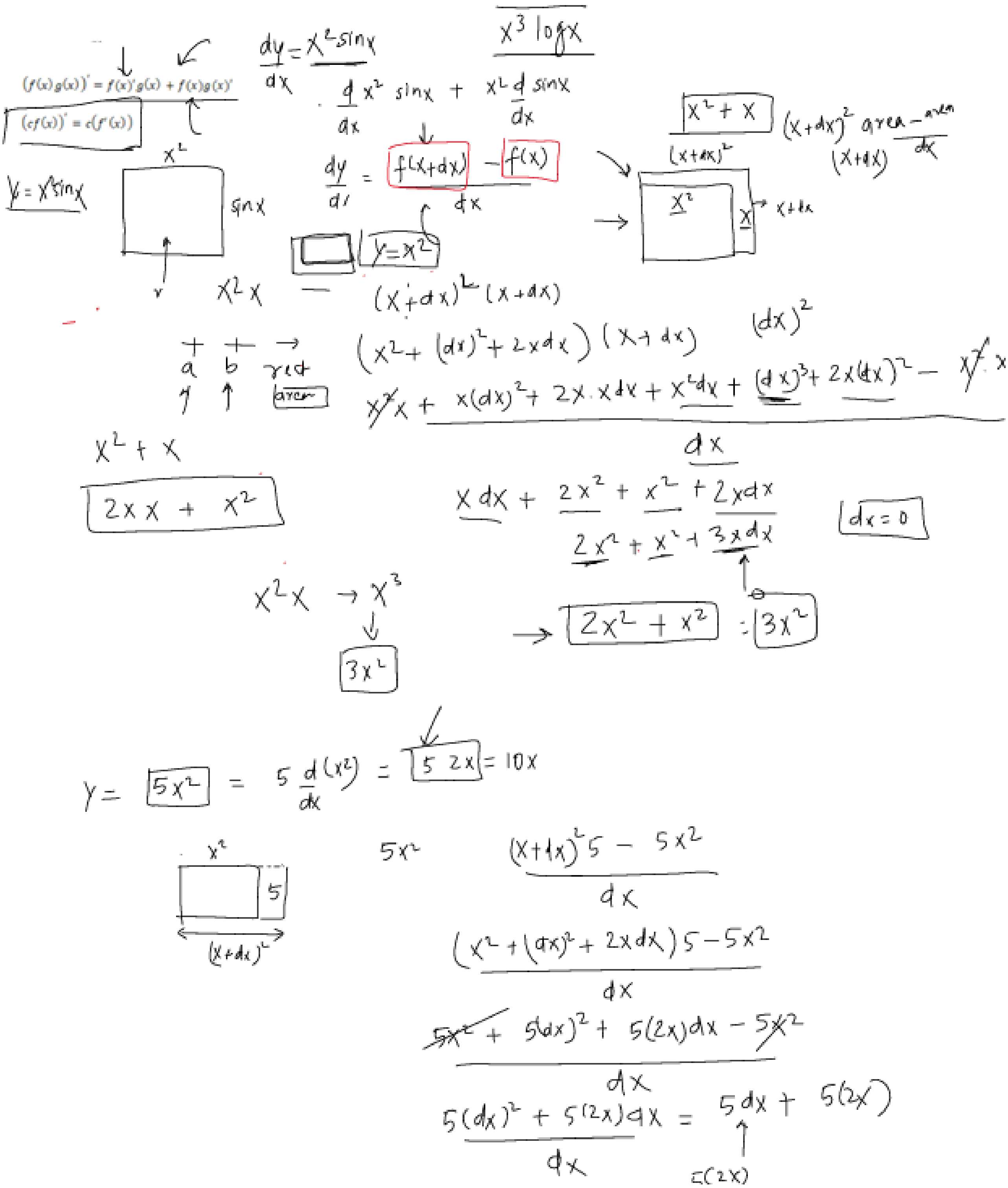
```
(d/dx) [f(x) * g(x)] = 2x * (3x + 1) + x^2 * 3
```

- Simplifying the expression gives us the derivative of the product:

```
(d/dx) [f(x) * g(x)] = 6x^2 + 2x + 3x^2 = 9x^2 + 2x
```

So, the derivative of the product **f(x) * g(x) is 9x^2 + 2x** according to the product rule.

The product rule is a valuable tool in calculus and is used to find derivatives in cases where we have a product of functions.



Quotient Rule

The quotient rule is a fundamental rule in calculus used to find the derivative of a quotient of two functions. It is applied when we have a function f(x) divided by another function g(x). The quotient rule states that the derivative of the quotient of two functions is equal to the numerator's derivative multiplied by the denominator, minus the numerator multiplied by the derivative of the denominator, all divided by the square of the denominator.

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

To apply the quotient rule, we follow these steps:

1. Identify the two functions: Let's say we have f(x) as the numerator and g(x) as the denominator.
2. Calculate the derivatives of each function: Find the derivatives of f(x) and g(x) separately, denoted as f'(x) and g'(x), respectively.
3. Apply the quotient rule formula: Plug in the derivatives into the quotient rule formula: [f'(x) * g(x) - f(x) * g'(x)] / [g(x)]^2.
4. Simplify the expression if possible: If further algebraic simplification is possible, simplify the expression obtained from applying the quotient rule.

Example to illustrate the quotient rule:

- Suppose we have two functions:

  ```
  f(x) = 3x^2 and g(x) = x + 1
  ```

- To find the derivative of their quotient, we can apply the quotient rule:

  ```
  (d/dx) [f(x) / g(x)] = [3x^2 * (x + 1) - (3x^2) * 1] / [(x + 1)]^2
  ```

- Simplifying the expression gives us the derivative of the quotient:

  ```
  (d/dx) [f(x) / g(x)] = [3x^3 + 3x^2 - 3x^2] / [(x + 1)]^2
  ```

- Simplifying further, we have:

  ```
  (d/dx) [f(x) / g(x)] = (3x^3) / [(x + 1)]^2
  ```

So, the derivative of the quotient **f(x) / g(x) is (3x^3) / [(x + 1)]^2** according to the quotient rule.

The quotient rule is a powerful tool in calculus and is used to find derivatives in cases where we have a quotient of functions.

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

$$\frac{d}{dx}\frac{x^2}{\sin x} =$$

$$x^2 (\sin x)^{-1}$$

a b

$$\frac{\frac{d}{dx}x^2 \sin x - x^2 \frac{d}{dx}\sin x}{(\sin x)^2}$$

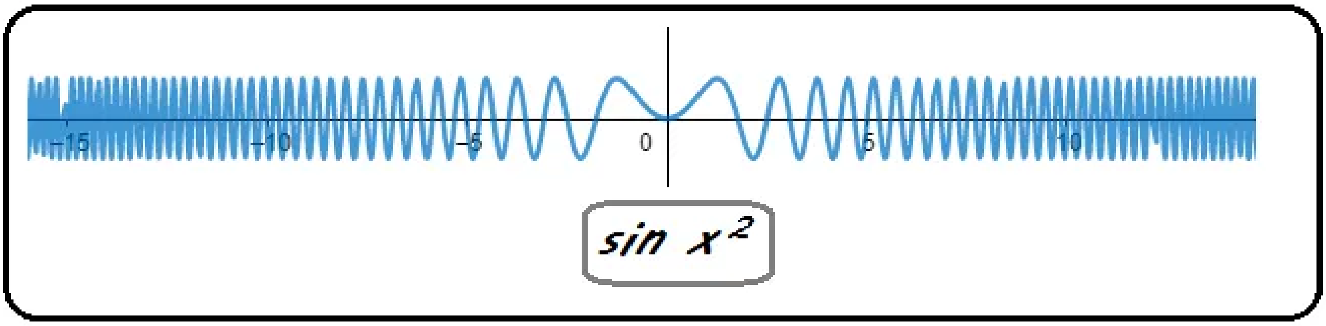
$$\boxed{\frac{dy}{dx} = \frac{2x\sin x - x^2\cos x}{(\sin x)^2}}$$

Chain Rule

The chain rule is a fundamental rule in calculus used to find the derivative of composite functions. It is applied when we have a function composed with another function. The chain rule allows us to determine how changes in the input of the outer function affect the output by considering the derivative of the outer function and the derivative of the inner function.

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

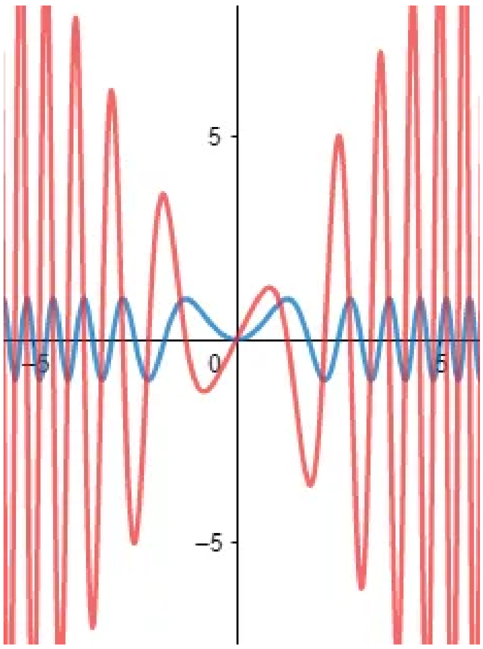
In other words, the derivative of the composite function is equal to the derivative of the outer function evaluated at the inner function, multiplied by the derivative of the inner function with respect to x.



$$\boxed{\frac{d}{dx}u(v(x)) = \frac{du}{\cancel{dv}} \cdot \frac{\cancel{dv}}{dx} = \frac{du}{dx}}$$

Applying this process to example:

$$\begin{aligned}\frac{d}{dx}\sin x^2 &= \frac{d}{d\ x^2}\sin x^2 \cdot \frac{d\ x^2}{dx} \\ &= 2x\cos x^2\end{aligned}$$



To apply the chain rule, we follow these steps:

1. Identify the inner and outer functions: Let's say we have a composite function y = f(g(x)).
2. Calculate the derivative of the inner function: Find the derivative of the inner function g(x) with respect to x, denoted as du/dx.

3. Calculate the derivative of the outer function: Find the derivative of the outer function f(u) with respect to u, denoted as dy/du.

4. Apply the chain rule formula: Multiply the derivatives obtained in steps 2 and 3 to get the derivative of the composite function dy/dx.

Example to illustrate the chain rule:

- Suppose we have the composite function:

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

- We can break it down into the inner and outer functions:

```
g(x) = x^2 + 1
f(u) = u^3
```

- To find the derivative dy/dx using the chain rule:

1. Calculate the derivative of the inner function:

```
du/dx = 2x
```

2. Calculate the derivative of the outer function:

```
dy/du = 3u^2
```

3. Apply the chain rule formula:

```
dy/dx = dy/du * du/dx = 3u^2 * 2x = 6x(x^2 + 1)^2
```

So, the derivative dy/dx of the composite function is 6x(x^2 + 1)^2 according to the chain rule.

The chain rule is a powerful tool in calculus and is used to find derivatives when we have functions composed with other functions. It allows us to handle more complex scenarios and is widely used in various areas of mathematics, physics, and engineering.

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

$f(x) \quad g(x)$

$y = f(g(x)) = \sin(x^2)$

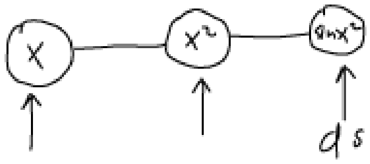
deep learning

$g(x) = x^2$

$f(g(x)) = \sin(x^2)$

$y = f(g(x))$

$$\frac{dy}{dx} = \boxed{\frac{df}{dg}\frac{dg}{dx}} \rightarrow$$



$$\frac{d\sin(x^2)}{dx^2} \times \frac{dx^2}{dx}$$

$$\cos x^2 \; 2x = \boxed{2x\cos x^2}$$

$y = \sin x^2$

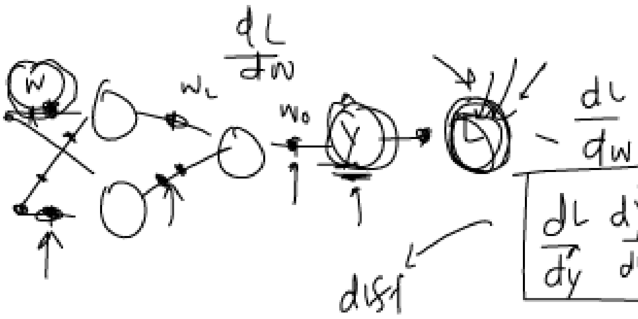
$\frac{dy}{dx}$

$\sin(\log(x^2))$

$x \rightarrow x^2 \rightarrow \log x^2 \rightarrow \sin(\log x^2)$

$$\frac{d\sin(\log x^2)}{d\log x} \; \frac{d\log x^2}{dx^2} \; \frac{dx^2}{dx}$$

$$\cos(\log x^2)\,\frac{1}{\cancel{x^2}}\,2\cancel{x} = \boxed{\frac{1}{x}\cos(\log x^2)}$$



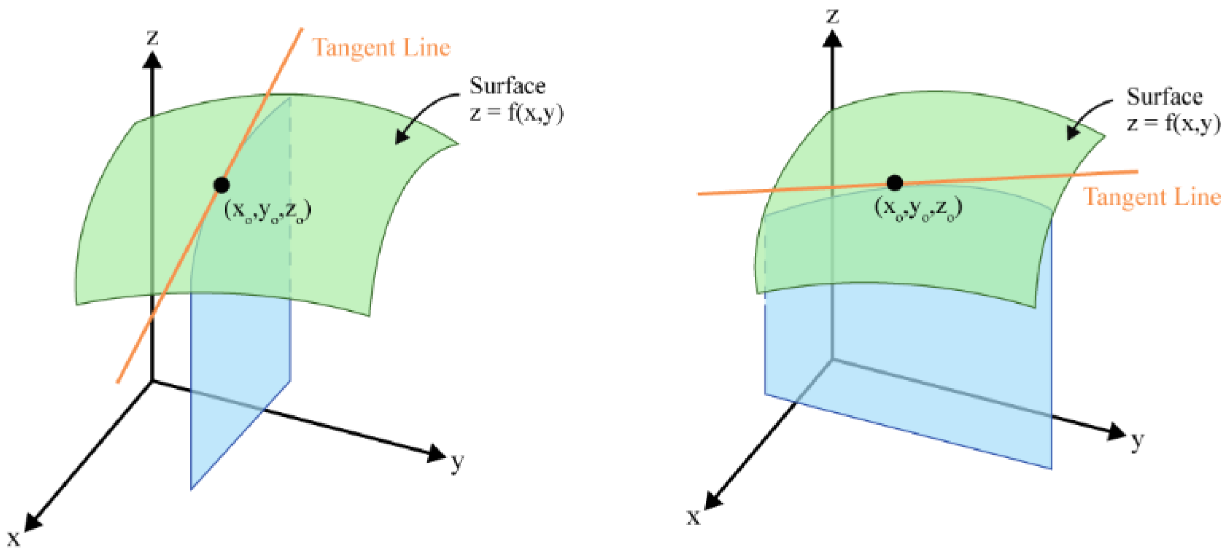
$$\boxed{\frac{dL}{dy}\,\frac{d\hat{y}}{dw_0}\,\frac{dw_0}{dw_2} \times \ldots}$$

Partial differentiation

In multivariable calculus, partial differentiation is a way of differentiating a function of several variables with respect to one of those variables, while treating the others as if they were constants.

For example, if we have a function f(x, y) of two variables, then the partial derivative of f with respect to x is:

```
∂f/∂x = lim_{h->0} (f(x + h, y) - f(x, y)) / h
```



Partial differentiation and **partial derivative** are two terms that are often used interchangeably in multivariable calculus. However, there is a subtle difference between the two terms.

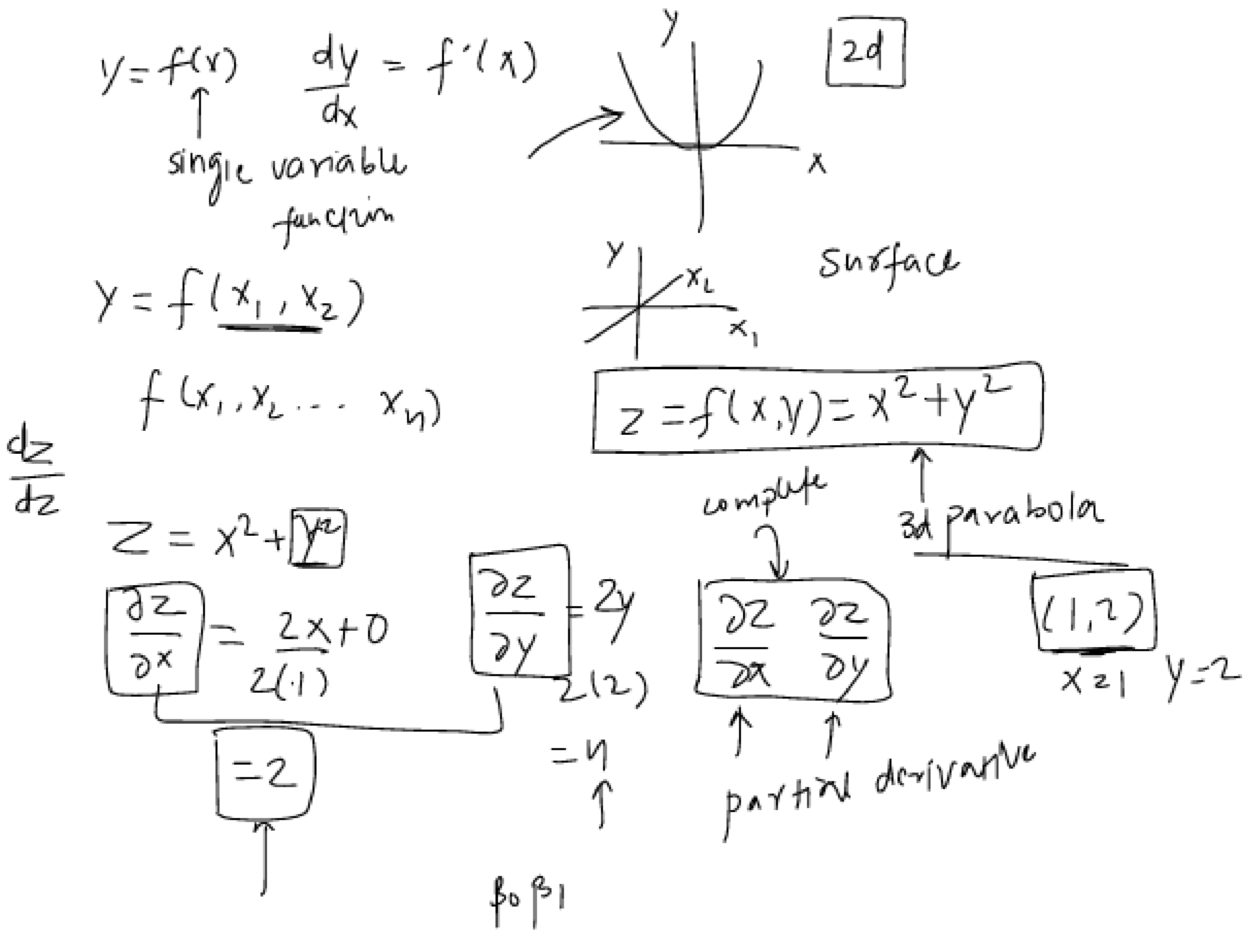
- Partial differentiation is the process of differentiating a function of several variables with respect to one of those variables, while treating the others as if they were constants.
- Partial derivative is the result of partial differentiation. It is a number that represents the rate of change of the function with respect to one of the variables, holding the other variables constant.

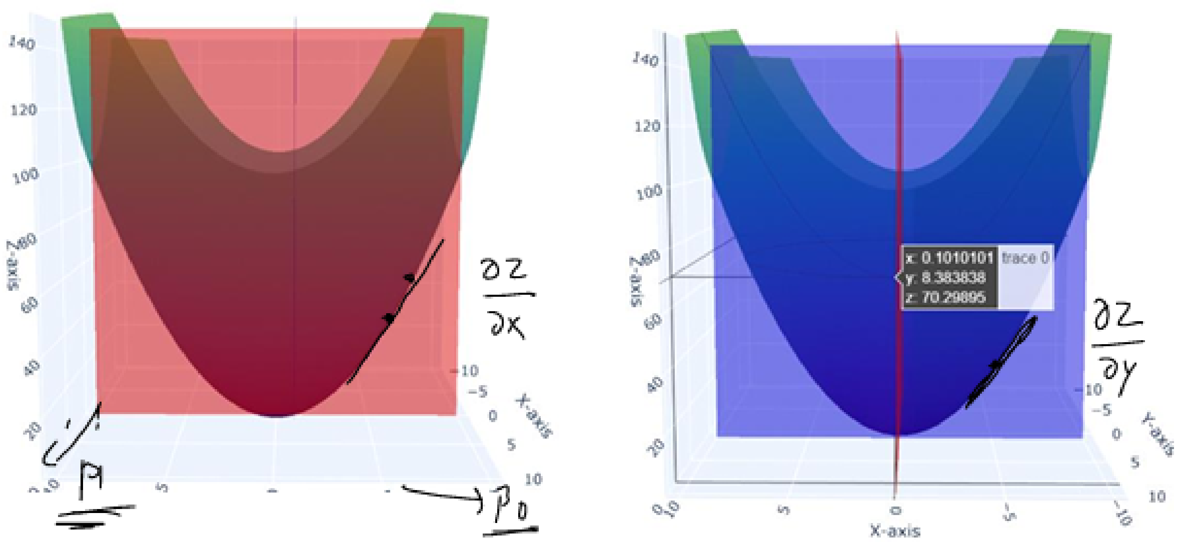
This is the same as the regular derivative of f(x, y), except that we are treating y as a constant.

Partial differentiation can be used to find the slope of a tangent plane to a surface, to find the maximum or minimum of a function, and to solve differential equations.

Here are some of the uses of partial differentiation:

- **Finding the slope of a tangent plane:** The slope of a tangent plane to a surface at a point is the partial derivative of the function with respect to the variable that is perpendicular to the tangent plane.
- **Finding the maximum or minimum of a function:** The partial derivatives of a function can be used to find the critical points of the function. Critical points are points where the derivative of the function is equal to 0 or undefined. The function can have a maximum, minimum, or neither at a critical point.
- **Solving differential equations:** Partial differential equations are equations that involve the partial derivatives of a function. Partial differentiation can be used to solve these equations.





Higher order Derivatives

Higher-order derivatives refer to the process of taking the derivative of a function multiple times. Just as we can find the first derivative of a function to measure its rate of change, we can continue this process to find second, third, and higher derivatives.

The n-th derivative of a function f(x) is denoted as $f^{(n)}(x)$ or d^nf/dx^n. The superscript '$^{(n)}$' represents the order of the derivative, indicating how many times the function has been differentiated.

To find higher-order derivatives, we apply the rules of differentiation repeatedly. Here are a few examples:

1. First derivative:

```
f'(x) = d/dx [f(x)]
```

2. Second derivative:

```
f''(x) = d/dx [f'(x)]
```

3. Third derivative:

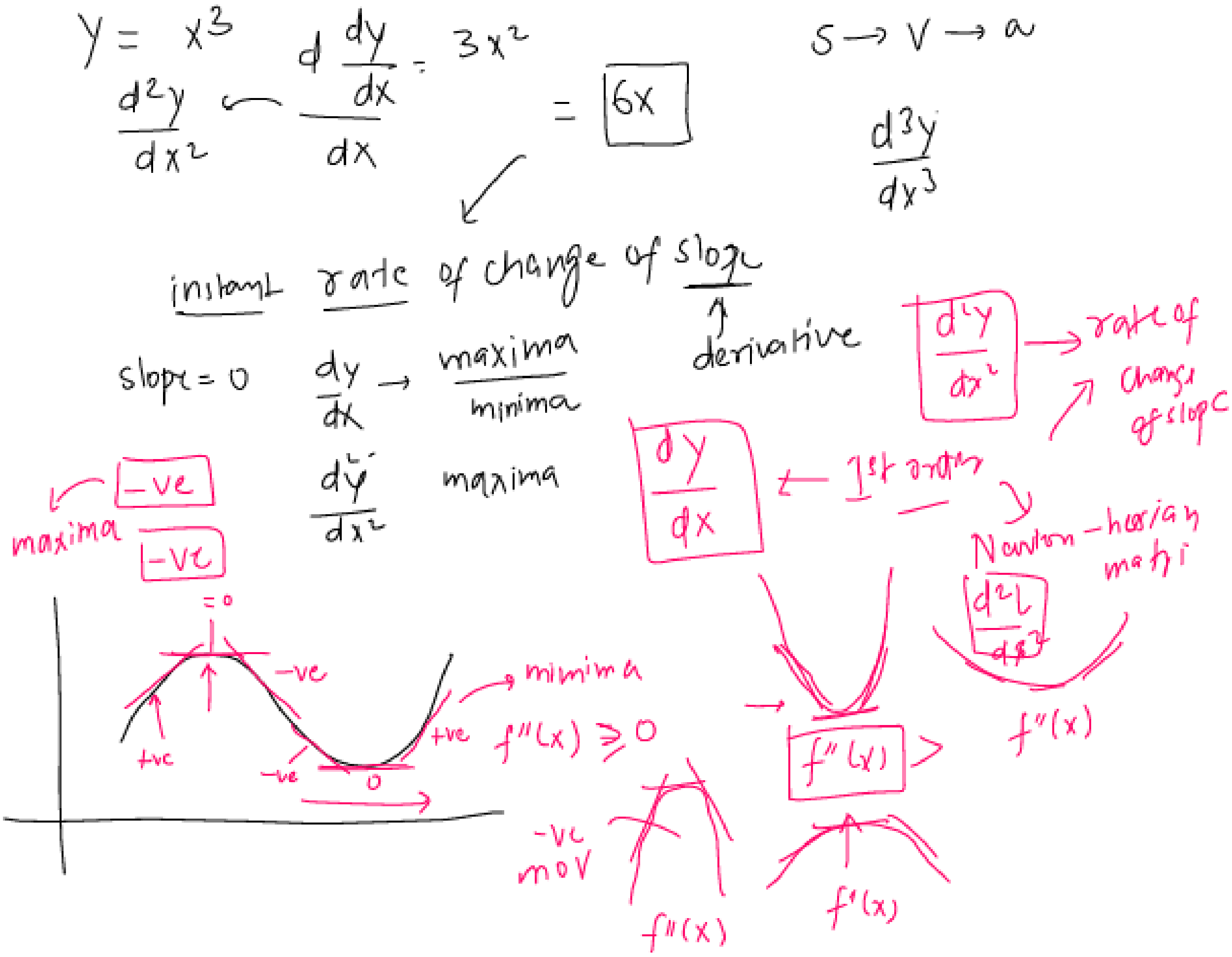
```
f'''(x) = d/dx [f''(x)]
```

We can also express higher-order derivatives using the prime notation. For example, $f^{(4)}(x)$ represents the fourth derivative of f(x).

Higher-order derivatives have important applications in various fields, including physics, engineering, and optimization. They provide information about the curvature, inflection points, and higher-order behavior of functions. Additionally, they are utilized in Taylor series expansions, which approximate functions as a sum of their derivatives at a specific point.

Here are some of the uses of higher order derivatives:

- **Determining the behavior of functions:** Higher order derivatives can be used to determine whether a function is increasing, decreasing, concave up, concave down, or has inflection points.
- **Solving differential equations:** Higher order derivatives can be used to solve differential equations.
- **Machine learning:** Higher order derivatives can be used in machine learning algorithms to improve the accuracy of predictions.



Matrix differentiation

Matrix differentiation refers to the process of taking derivatives with respect to matrices or vector variables. It extends the concepts of differentiation from scalar functions to functions that involve matrices or vectors.

When differentiating with respect to a matrix or vector, we consider each element of the matrix or vector as an independent variable. The rules of matrix differentiation involve manipulating and calculating derivatives based on these elements.

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x}^T \left(\mathbf{A} + \mathbf{A}^T\right) \tag{44}$$

Proof: *By definition*

$$\alpha = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j \tag{45}$$

Differentiating with respect to the k*th element of* $\mathbf{x}$ *we have*

$$\frac{\partial \alpha}{\partial x_k} = \sum_{j=1}^{n} a_{kj} x_j + \sum_{i=1}^{n} a_{ik} x_i \tag{46}$$

for all $k = 1, 2, \ldots, n$, *and consequently,*

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x}^T \mathbf{A}^T + \mathbf{x}^T \mathbf{A} = \mathbf{x}^T \left(\mathbf{A}^T + \mathbf{A}\right) \tag{47}$$

q.e.d.

Here are some key concepts and rules related to matrix differentiation:

1. **Scalar-Vector Differentiation**:

   - If y is a column vector and x is a vector, then the derivative of y with respect to x, denoted as dy/dx or ∇y, is a row vector containing the derivatives of each element of y with respect to the elements of x.
   - Each element of the derivative is calculated separately using standard differentiation rules.

2. **Scalar-Matrix Differentiation**:

   - If y is a scalar and X is a matrix, then the derivative of y with respect to X, denoted as ∇y or (∂y/∂X), is a matrix of the same size as X.
   - Each element of the derivative matrix is calculated separately by taking the derivative of y with respect to the corresponding element of X.

3. **Matrix-Vector Differentiation**:

   - If Y is a matrix and x is a vector, then the derivative of Y with respect to x, denoted as (∂Y/∂x), is a third-order tensor.
   - Each element of the tensor represents the derivative of a corresponding element of Y with respect to an element of x.

4. **Matrix-Matrix Differentiation**:

- If Y is a matrix and X is a matrix, then the derivative of Y with respect to X, denoted as (∂Y/∂X), is a fourth-order tensor.
- Each element of the tensor represents the derivative of a corresponding element of Y with respect to an element of X.

Matrix differentiation involves applying these rules while considering the specific layout and dimensions of matrices and vectors. It is a crucial tool in fields such as optimization, machine learning, and econometrics, where functions often involve matrix or vector variables.

$$\frac{d}{dx}\boxed{Ax} \leftarrow \frac{d}{dx}Cx = C \quad \text{constant} \qquad \frac{d}{dx}5x = 5 \qquad \frac{d}{dx}Ax \qquad A^T = A$$

matrix

$$\frac{d}{dx}Ax \rightarrow \boxed{A}$$

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}x_1 + \boxed{a_{12}x_2} \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix} \begin{matrix} \rightarrow f_1(x_1,x_2) \\ \rightarrow f_2(x_1,x_2) \end{matrix} \rightarrow \begin{bmatrix} \frac{df_1}{dx_1} & \frac{df_1}{dx_2} \\ \frac{df_2}{dx_1} & \frac{df_2}{dx_2} \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \rightarrow A \qquad \frac{d}{dx}Ax = A$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad A^T = A \qquad \begin{bmatrix} a_{11} & a_x \\ a_x & a_{22} \end{bmatrix}^T$$

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

$$\frac{dy}{dx} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_x \\ a_x & a_{12} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \begin{bmatrix} a_{11} & a_x \\ a_x & a_{22} \end{bmatrix} \leftarrow$$

$$\underbrace{\underset{1\times 2}{\begin{bmatrix} x_1 & x_2 \end{bmatrix}} \begin{bmatrix} a_{11}x_1 + a_x x_2 \\ a_x x_1 + a_{22}x_2 \end{bmatrix}}_{2\times 1} \qquad f(x_1, x_2)$$

$$a_{11}x_1^2 + a_x x_2^2 + a_x x_1^2 + a_{11}x_2^2 \qquad \begin{bmatrix} \frac{df}{dx_1} \leftarrow \\ \frac{df}{dx_2} \end{bmatrix}$$

$$\begin{bmatrix} 2a_{11}x_1 + 2a_x x_1 \\ 2a_x x_2 + 2a_{22}x_2 \end{bmatrix} = 2\begin{bmatrix} a_{11}x_1 + a_x x_1 \\ a_x x_2 + a_{22}x_2 \end{bmatrix} \leftarrow$$

$$\left\{ \text{matrix calculus} \right\} \qquad 2\begin{bmatrix} a_{11} & a_x \\ a_x & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\rightarrow 2x^2 \qquad \rightarrow \boxed{4x} \leftarrow \qquad \rightarrow (2Ax)^T \quad 2X^T A^T = \boxed{2X^T A}$$

Use of Calculus and differentiation in Machine Learning and Artificial Intelligence

Calculus, particularly differentiation, plays a fundamental role in machine learning and artificial intelligence. Differentiation enables us to optimize models, understand their behavior, and make predictions.

Here's how calculus and differentiation are used in machine learning and artificial intelligence:

1. **Optimization**: Machine learning often involves optimizing models to minimize a cost or loss function. This optimization process aims to find the optimal values for the model's parameters that minimize the error. Calculus is used to calculate the gradients, which indicate the direction of steepest ascent or descent of the loss function. Techniques like gradient descent use these gradients to update the model's parameters iteratively, moving towards the minimum of the loss function.

2. **Backpropagation**: Backpropagation is a widely used algorithm in training artificial neural networks. It utilizes the chain rule of calculus to efficiently calculate the gradients for each parameter in the network. By propagating the errors backward from the output layer to the input layer, backpropagation enables the adjustment of weights and biases in a neural network to improve its performance.

3. **Activation Functions**: Activation functions are crucial in neural networks as they introduce non-linearities, allowing the network to learn complex relationships. The choice of activation function affects the shape of the network's decision boundaries and influences how information flows through the network. The derivatives of activation functions are used during backpropagation to update the network's weights and biases.

4. **Regularization**: Regularization techniques, such as L1 and L2 regularization, are used to prevent overfitting and improve the generalization of machine learning models. These techniques involve adding regularization terms to the loss function, which help control the complexity of the model. Calculus is used to find the derivatives of these regularization terms and incorporate them into the overall gradient calculations.

5. **Sensitivity Analysis**: Sensitivity analysis in machine learning involves measuring how changes in the input variables affect the output predictions. Calculus comes into play by calculating the partial derivatives of the output with respect to the input variables. These derivatives quantify the sensitivity of the model's predictions to changes in the input features.

Calculus and differentiation provide the mathematical foundation for optimization, training, and fine-tuning of machine learning models. They enable us to leverage the power of gradients, adjust model parameters, and make accurate predictions.

In [ ]:

```

```