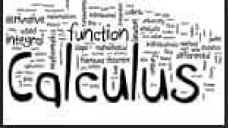


Calculus

Why Calculus ??

Every ML engineer some basic calculus in order to understand how functions change over time (derivatives), and to calculate the total amount of a quantity that accumulates over a time period (integrals). The language of calculus will allow you to speak precisely about the properties of functions and better understand their behavior.

Calculus for Machine learning

- Machine learning uses derivatives in optimization problems. Optimization algorithms like gradient descent use derivatives to decide whether to increase or decrease weights in order to maximize or minimize some objective (e.g. a model's accuracy or error functions). Derivatives also help us approximate nonlinear functions as linear functions (tangent lines), which have constant slopes. With a constant slope we can decide whether to move up or down the slope (increase or decrease our weights) to get closer to the target value (class label).

Calculus

Where we use Calculus in Machine learning

Gradient computations: Are generally fed into numerical optimization algorithms and calculus is readily used to compute these, especially in the case of neural networks where we use chain rule to arrive at the Backpropagation Algorithm.

Numerical Optimization: This is used to train models, given a dataset, that will be used to perform anything from inference to data generation to sequential decision making.

Bayesian Methods using probability density functions:

- One will use integrals and even need to approximately compute these integrals to perform the desired Bayesian computations

Calculus

Where we use Calculus in Machine learning

Variational Inference and related techniques: The whole framework of Variational Inference is based on variational calculus, so certainly this is another use of calculus.

Generative Adversarial Networks: A lot of the fundamental theory is based on differential game theory, which itself uses quite a bit of variational calculus, so calculus again finds some great importance

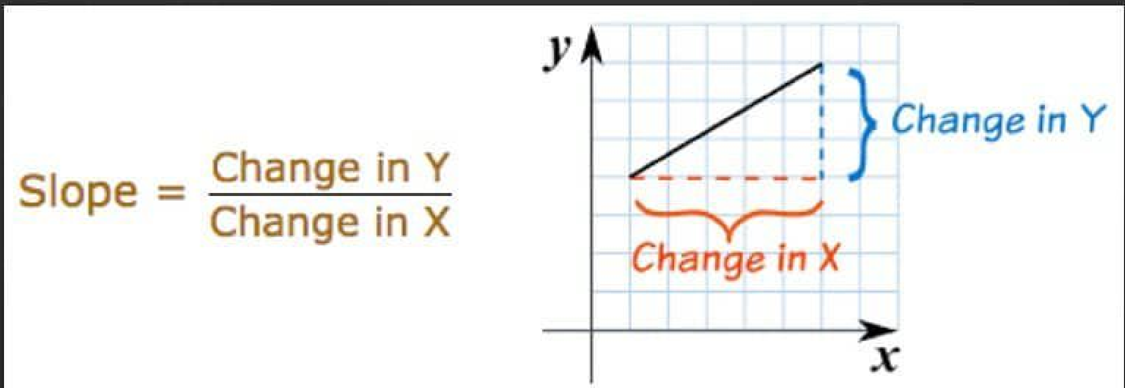
Calculus

Derivatives

A derivative can be defined in two ways
1. Instantaneous rate of change (Physics)
2. Slope of a line at a specific point (Geometry)

Here we go with Geometry which is easy to understand and of course we are dealing with machine learning not with physics learning.

It is all about slope(how much does y or f(x) change given a specific change in x?)



Calculus

Derivatives

Using this we can easily calculate the slope between two points. But how do we find the slope at a point? There is nothing to measure!. Derivatives help us answer this question. A derivative outputs an expression we can use to calculate the instantaneous rate of change, or slope, at a single point on a line.

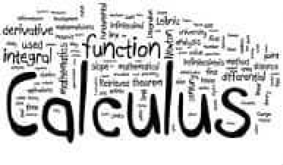
Let us Find a Derivative!

1. Define the function let it be $f(x)=x^2$
2. Increment x by a very small value $h(h=\Delta x)$

$$f(x+h)=(x+h)^2$$

3. Apply the slope formula

$$\frac{f(x+h)-f(x)}{h}$$



Calculus

Derivatives

4. Simplify the equation

$$\frac{x^2+2xh+h2-x^2}{h}$$

$$\frac{2xh+h^2}{h}=2x+h$$

5. Set h to 0 (the limit as h heads toward 0)

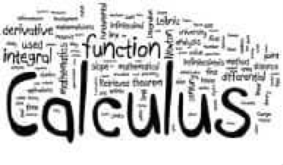
$$2x+0=2x$$

Result: the derivative of x2 is 2x, In other words, the slope at x is 2x.

We write dx instead of "h heads towards 0". And "the derivative of" is commonly written d/dx : $\frac{d}{dx}x^2 = 2x$

The derivative of x2 equals 2x or simply "d dx of x2 equals 2x

What does this mean? It means that, for the function x2, the slope or "rate of change" at any point is 2x.



Calculus

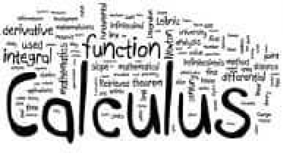
Chain rule

- The chain rule is a formula for calculating the derivatives of composite functions. Composite functions are functions composed of functions inside other function(s).
- Given a composite function f(x)=A(B(x)), the derivative of f(x) equals the product of the derivative of A with respect to B(x) and the derivative of B with respect to x.
- composite function derivative= outer function derivative * inner function derivative.
- For example, given a composite function f(x), where:

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

The chain rule tells us that the derivative of f(x) equals:

$$\frac{df}{dx} = \frac{dh}{dg} \cdot \frac{dg}{dx}$$



Calculus

Chain rule

- for example Say f(x) is composed of two functions h(x)=x^3 and g(x)=x^2. And that: f(x)=h(g(x))
- The derivative of f(x) would equal:

$$\frac{df}{dx} = \frac{dh}{dg} \cdot \frac{dg}{dx}$$

Steps

1. Solve for the inner derivative of g(x)=x^2

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

2. Solve for the outer derivative of h(x)=x3, using a placeholder b to represent the inner function x2

$$\frac{dh}{db} = 3b^2$$

Chain rule

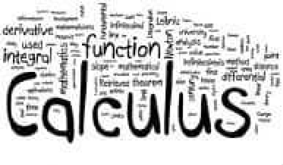
3. Swap out the placeholder variable (b) for the inner function (g(x)).

$$= 3(x^2)^2$$

$$= 3x^4$$

4. Return the product of the two derivatives.

$$= 3x^4 \cdot 2x = 6x^5$$



Calculus

Multiple functions

In the previous post we assumed a composite function containing a single inner function. But the chain rule can also be applied to higher-order functions like

$$f(x)=A(B(C(x)))$$

The chain rule tells us that the derivative of this function equals:

$$\frac{df}{dx} = \frac{dA}{dB} . \frac{dB}{dC} . \frac{dC}{dx}$$

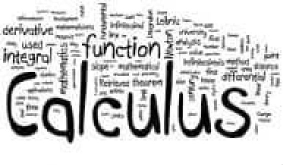
We can also write this derivative equation f′ notation:

$$f'=A'(B(C(x))\cdot B'(C(x))\cdot C'(x)$$

Steps

1. Given the function $f(x)=A(B(C(x)))$, lets assume:

$$A(x) = \sin(x)$$
$$B(x) = x^2$$
$$C(x) = 4x$$



Calculus

Multiple functions

2 . The derivatives of these functions would be:

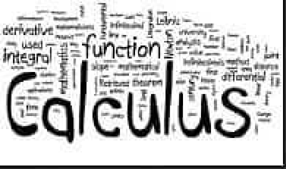
$$A'(x) = \cos(x)$$
$$B'(x) = 2x$$
$$C'(x) = 4$$

3. We can calculate the derivative of f(x) using the following formula:

$$f'(x)=A'((4x)^2)\cdot B'(4x)\cdot C'(x)$$

3. We then input the derivatives and simplify the expression:

$$\begin{aligned} f'(x) &= \cos((4x)^2)\cdot 2(4x)\cdot 4 \\ &= \cos(16x2)\cdot 8x\cdot 4 \\ &= \cos(16x2)32x \end{aligned}$$



Calculus

Multiple functions

In the previous post we assumed a composite function containing a single inner function. But the chain rule can also be applied to higher-order functions like

$$f(x)=A(B(C(x)))$$

The chain rule tells us that the derivative of this function equals:

$$\frac{df}{dx} = \frac{dA}{dB} \cdot \frac{dB}{dC} \cdot \frac{dC}{dx}$$

We can also write this derivative equation f′ notation:

$$f'=A'(B(C(x))\cdot B'(C(x))\cdot C'(x)$$

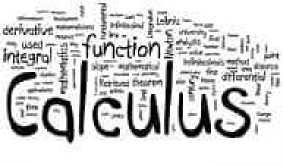
Steps

1. Given the function $f(x)=A(B(C(x)))$, lets assume:

$$A(x) = \sin(x)$$

$$B(x) = x^2$$

$$C(x) = 4x$$



Calculus

Multiple functions

2 . The derivatives of these functions would be:

$$A'(x) = \cos(x)$$

$$B'(x) = 2x$$

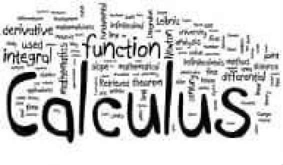
$$C'(x) = 4$$

3. We can calculate the derivative of f(x) using the following formula:

$$f'(x)=A'((4x)^{\wedge}2)\cdot B'(4x)\cdot C'(x)$$

3. We then input the derivatives and simplify the expression:

$$\begin{aligned} f'(x) &= \cos((4x)^{\wedge}2)\cdot 2(4x)\cdot 4 \\ &= \cos(16x2)\cdot 8x\cdot 4 \\ &= \cos(16x2)32x \end{aligned}$$



Calculus

Gradients

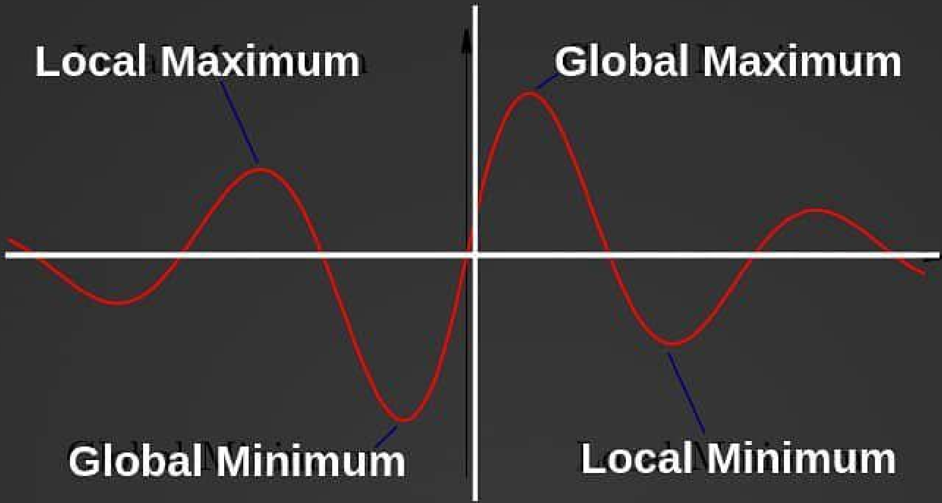
A gradient is a vector that stores the partial derivatives of multivariable functions. It helps us calculate the slope at a specific point on a curve for functions with multiple independent variables. In order to calculate this more complex slope, we need to isolate each variable to determine how it impacts the output on its own. To do this we iterate through each of the variables and calculate the derivative of the function after holding all other variables constant. Each iteration produces a partial derivative which we store in the gradient.

Partial derivatives

In functions with 2 or more variables, the partial derivative is the derivative of one variable with respect to the others.

Calculus

Maxima and Minima



- we can have multiple local maximas and multiple local minimas. But we only have one global minima(lowest of all local minima) and one global maxima(highest of all local minima).
- The slope at maxima or minima is zero.
- we use derivatives to calculate the slope of a point and compare with neighboring points to say whether it is maxima or minima.
- Based on the initialization of weights it takes a minama or maxima.