Calculus for ML

We will talk about...

- Machine Learning Use Cases
- Derivatives
- Chain Rule
- Definite Integrals

Machine Learning Use Cases

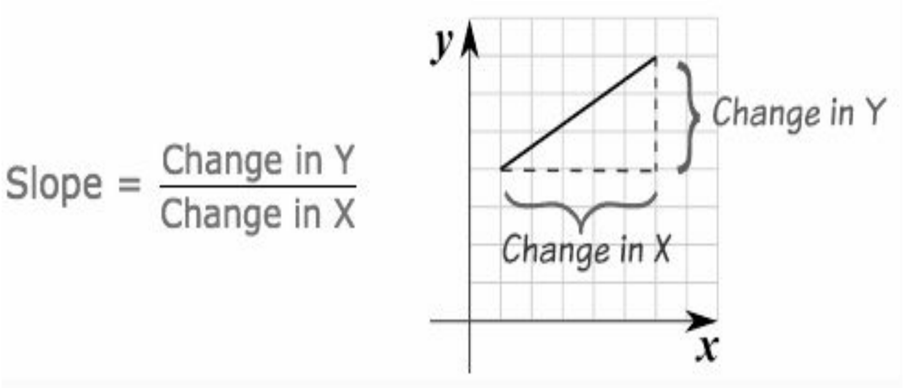
- Numerical Optimization
- Gradient Computations
- Probability Density Functions
- Variational Inference and Related Techniques (Out of scope)

Derivatives

A derivative can be defined in two ways:

- Instantaneous rate of change (Physics)
- Slope of a line at a specific point (Geometry)

A derivative outputs an expression we can use to calculate the instantaneous rate of change, or slope, at a single point on a line.



Derivatives (Contd.)

The derivative of a function $y = f(x)$ with respect to x is defined as

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

provided that the limit exists.

Derivative of y = f(x) with respect to x is represented as dy/dx or f'(x)

Chain Rule

The chain rule is a formula for calculating the derivatives of composite functions. Composite functions are functions composed of functions inside another function(s).

Given a composite function $f(x)$ = h(g(x)), the derivative of $f(x)$ equals the product of the derivative of h with respect to g(x) and the derivative of g with respect to x.

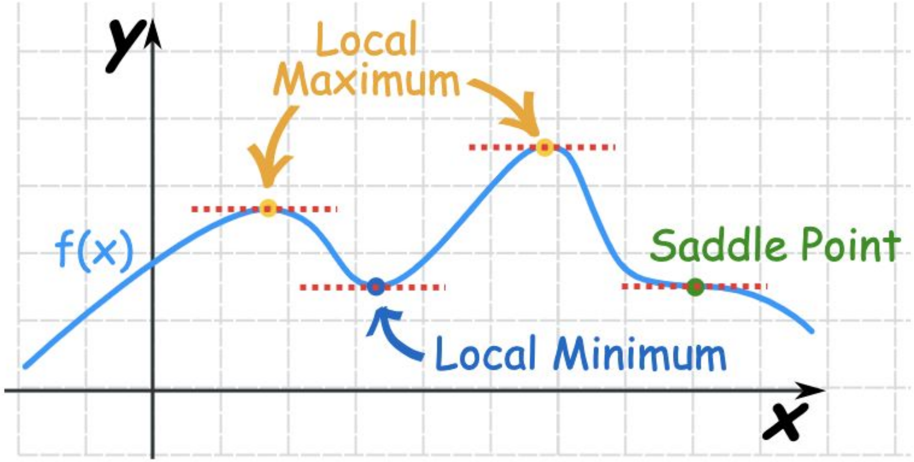
Composite function derivative = Outer function derivative $*$ Inner function derivative

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

$$df/dx = df/dg * dg/dx$$

Maxima and Minima using Derivatives

In a smoothly changing function a maximum or minimum is always where the function flattens out (except for a saddle point).



Where does it flatten out?
Where the slope is zero.

Where is the slope zero?
The first derivative will tell us!
$f'(x) = 0 \Rightarrow x = x_0, x_1$

How do we know if it's a maximum or minimum?
The second derivative will tell us!
If $f''(x_0) < 0 \Rightarrow x_0$ is point of maxima
If $f''(x_0) > 0 \Rightarrow x_0$ is point of minima

Partial Derivatives

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

Let's take an example: $\mathbf{z = f(x, y)}$

If we change x, but hold all other variables constant, how does $f(x, y)$ change?

That's one partial derivative. The next variable is y.

If we change y but hold x constant, how does $f(x, y)$ change?

Partial Derivatives (Contd.)

$$f(x,y) = 2x^2 + 3xy + 6x + 7y$$

$$\frac{\partial f}{\partial x} = 4x + 3y + 6$$

$$\frac{\partial^2 f}{\partial x^2} = 4$$

$$\frac{\partial f}{\partial y} = 3x+7$$

$$\frac{\partial^2 f}{\partial y^2} = 0$$

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

Jacobian Matrix

Consider the function $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\mathbf{f}(x, y) = \begin{bmatrix} x^2 y \\ 5x + \sin y \end{bmatrix}.$$

Then we have

$$f_1(x, y) = x^2 y$$

and

$$f_2(x, y) = 5x + \sin y$$

and the Jacobian matrix of $\mathbf{F}$ is

$$\mathbf{J_f}(x, y) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x} & \dfrac{\partial f_1}{\partial y} \\ \\ \dfrac{\partial f_2}{\partial x} & \dfrac{\partial f_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 2xy & x^2 \\ 5 & \cos y \end{bmatrix}$$

and the Jacobian determinant is

$$\det(\mathbf{J_f}(x, y)) = 2xy \cos y - 5x^2.$$

$$J(x_1, x_2, x_3, \cdots x_n) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \frac{\partial f_3}{\partial x_3} & \cdots & \frac{\partial f_3}{\partial x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \frac{\partial f_n}{\partial x_3} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

Derivatives Formulas Recap

Constant Rule: $\frac{d}{dx}(c) = 0$

Constant Multiple Rule: $\frac{d}{dx}[cf(x)] = cf'(x)$

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

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

Difference Rule: $\frac{d}{dx}[f(x) - g(x)] = f'(x) - g'(x)$

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

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

Chain Rule: $\frac{d}{dx}f(g(x)) = f'(g(x))g'(x)$

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

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

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

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

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

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

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

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

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

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

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

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

Definite Integrals

The integral of $f(x)$ corresponds to the computation of the area under the graph of $f(x)$. The area under $f(x)$ between the points $x=a$ and $x=b$ is denoted as follows:

$$A(a, b) = \int_a^b f(x)\, dx.$$