Google Developer Groups

ML Math Club: Calculus for Optimization

Moataz Ben Trad

Data Science student | AI Enthusiastic

ML/AI Mentor in gdg on campus ISAMM

Content

1. Introduction to Optimization
2. Derivatives and Partial Derivatives
3. Gradient
4. Jacobians
5. Gradient Descent
6. Exercises and Coding Examples

Do not edit
How to change the design

Join at slido.com
#4587614

Presenting with animations, GIFs or speaker notes? Enable our Chrome extension

slido

Google Developer Groups

Introduction to Optimization

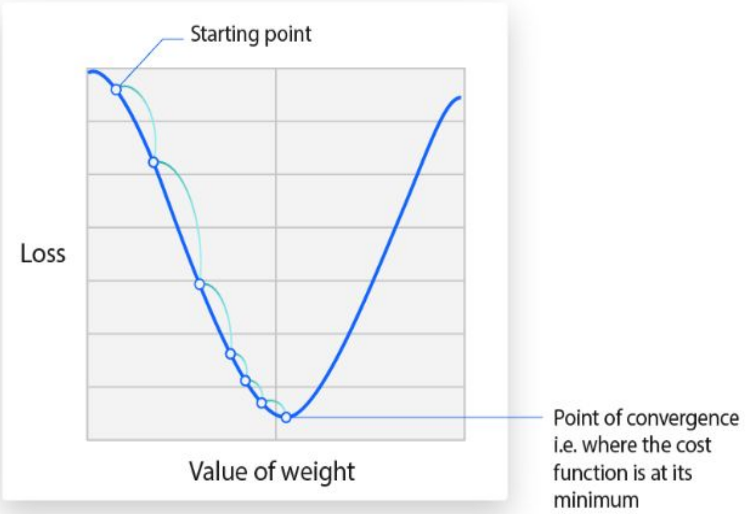
What is Mathematical Optimization ?

“Optimization” comes from the same root as “optimal”, which means best. When you optimize something, you are “making it best”. But “best” can vary. If you’re a runner, you might want to maximize your speed and distance, and also minimize your risk of injury. Both maximizing and minimizing are types of optimization problems…

Why Optimization Matters in ML?

Optimization is at the core of many machine learning tasks. Here are some key areas where optimization techniques are applied:

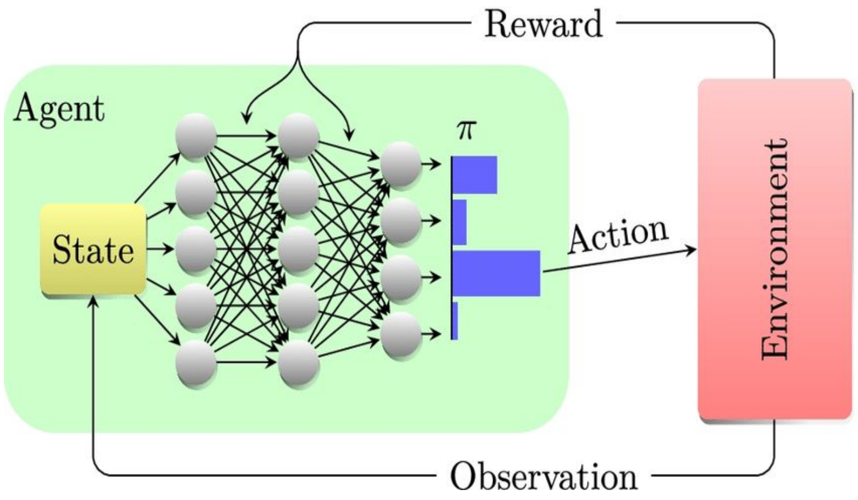
Model Training: Minimize the loss function to improve model accuracy.



https://www.ibm.com/fr-fr/topics/neural-networks

Why Optimization Matters in ML?

Reinforcement Learning: Optimization is used to maximize rewards in reinforcement learning, where an agent learns by interacting with an environment.



https://www.researchgate.net/publication/360910430_Quantum_reinforcement_learning_the_maze_problem/figures?lo=1

What does "optimization" mean?

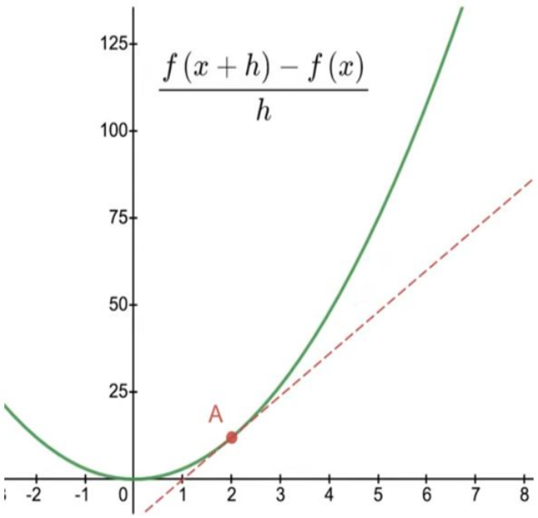
Google Developer Groups

Derivatives
and
Partial Derivatives

Derivatives

Rate of Change of a function with respect to its variable.
The derivative at a point represents the **slope of the tangent** line to the curve at that point.

Notation: $f'(x)$ ou $\dfrac{\mathrm{d}f}{\mathrm{d}x}(x)$.



https://www.desmos.com/calculator/ic9r3mgsxu?

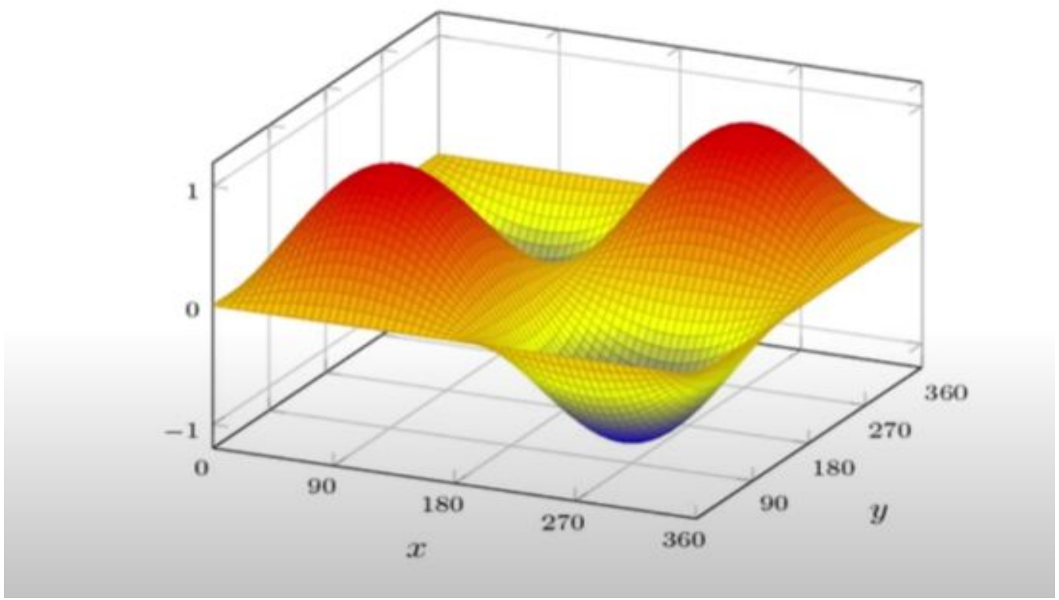
Partial Derivatives

A partial derivative measures the rate of change of a function with respect to one of its variables while keeping all other variables constant.

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

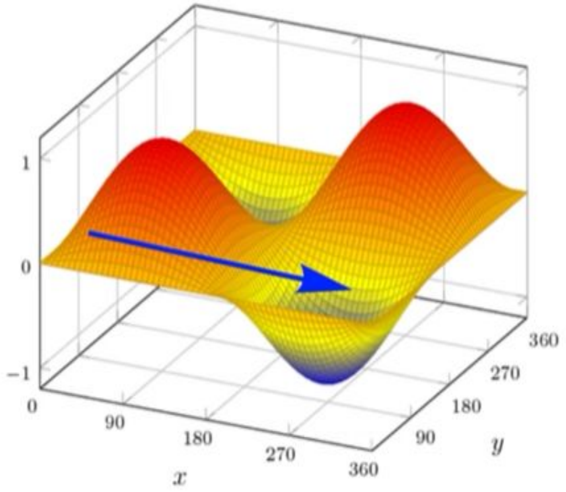
https://relcalc.espaceweb.usherbrooke.ca/relcalc-3/sec-Derivees-Partielles.html

f(x, y)



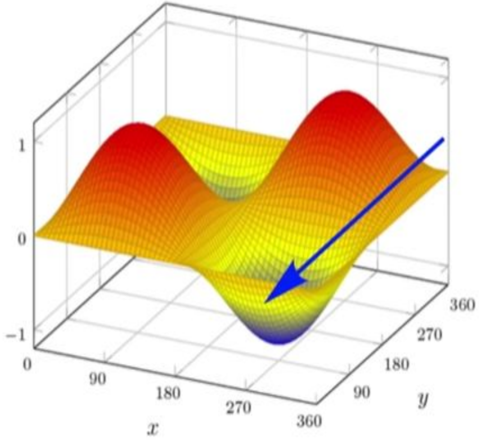
$$f(x, y)$$

$$\frac{\partial}{\partial x}$$



partial derivative with respect to x

$f(x, y)$



$$\frac{\partial}{\partial y}$$

partial derivative with respect to y

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

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

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

these are the rates of change for the function

https://www.desmos.com/3d/out3uua3js

Do not edit

How to change the design

Given the function $f(x,y) = 3x^2y + 2xy^2 + x$, what is the partial derivative of f with respect to x and y

Google Developer Groups

Gradient & Jacobian

Gradient

The gradient of a function is a vector of its partial derivatives.

Explanation: Points in the direction of the greatest increase of the function and its magnitude represents the rate of increase.

notation :

$$\nabla f(p) = \left(\frac{\partial f}{\partial x_1}(p), \ldots, \frac{\partial f}{\partial x_n}(p) \right)$$

Jacobian

The Jacobian matrix of a vector-valued function $\mathbf{f}(x_1, x_2, \ldots, x_n)$ is a matrix of its first-order partial derivatives.

Notation:

$$\mathbf{J}_\mathbf{f} = \begin{bmatrix} \dfrac{\partial \mathbf{f}}{\partial x_1} & \cdots & \dfrac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^\mathsf{T} f_1 \\ \vdots \\ \nabla^\mathsf{T} f_m \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Example:

Consider a function $\mathbf{f} : \mathbf{R}^2 \to \mathbf{R}^2$, with $(x, y) \mapsto (f_1(x, y), f_2(x, y))$, given by

$$\mathbf{f}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} f_1(x, y) \\ f_2(x, y) \end{bmatrix} = \begin{bmatrix} x^2 y \\ 5x + \sin y \end{bmatrix}.$$

The Jacobian matrix of $\mathbf{f}$ is:

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

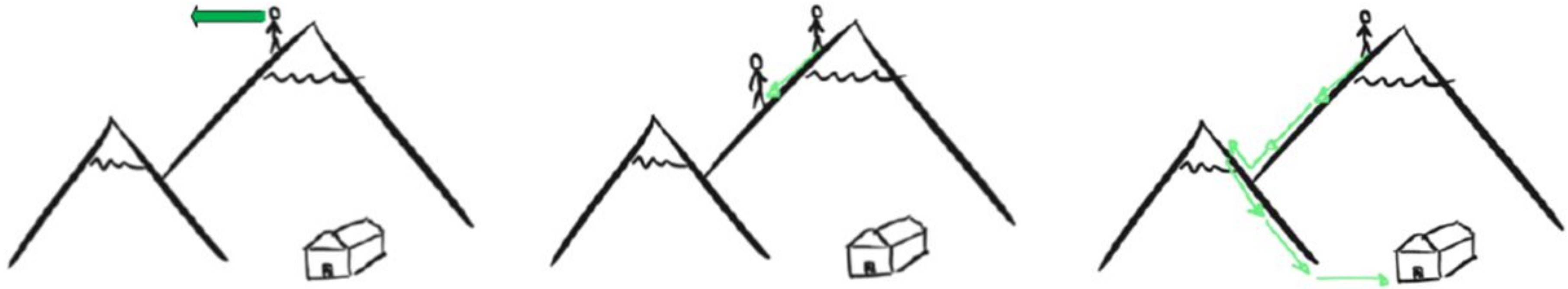
Google Developer Groups

Gradient Descent

Understanding Gradient Descent

Imagine being lost in the mountains. Your goal is to reach the shelter at the lowest point in the valley. Without a map, you don’t know the shelter’s exact location—you must find it on your own.

solution

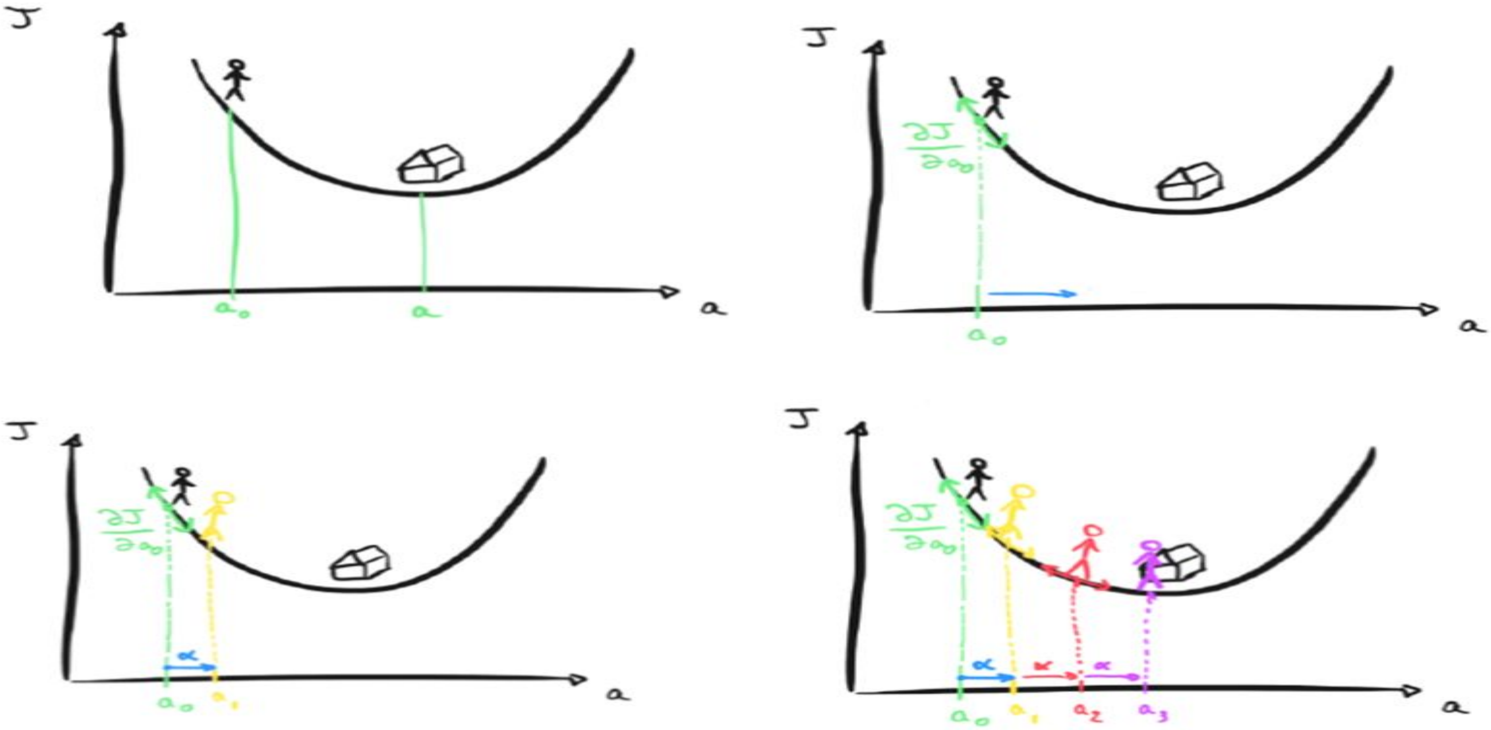


Step 1: Find the steepest slope

Step 2: Walk some distance in that direction

Step 3: Repeat steps 1 and 2 in a loop

Step 1: Calculate the gradient



Our ultimate goal: Find the parameters a and b that minimize $J(a, b)$.

To achieve this, we will start by choosing a and b randomly (starting from an arbitrary point on the mountain) and then iteratively use gradient descent to update our parameters in the direction that lowers the Cost Function the most.

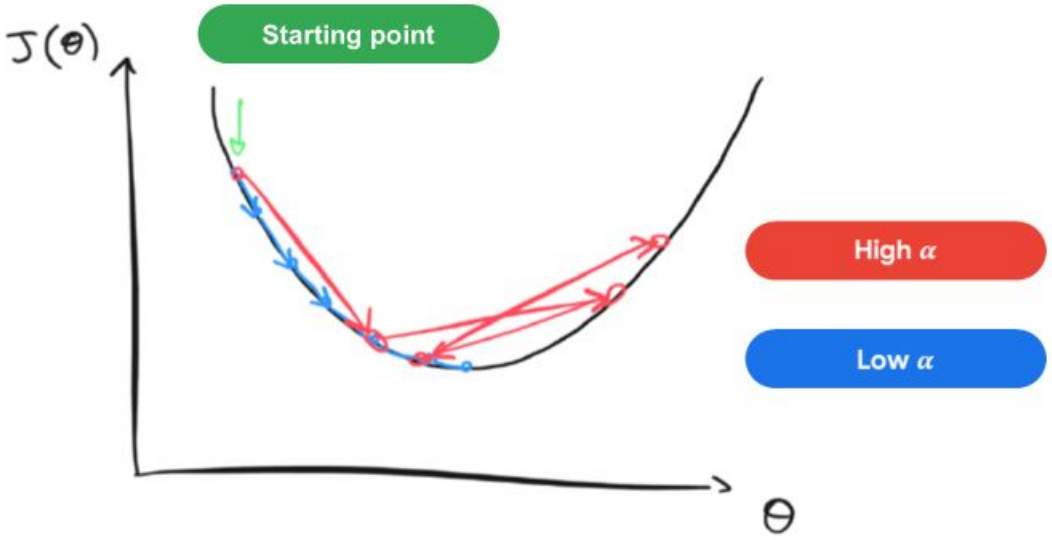
Repeat in a loop:

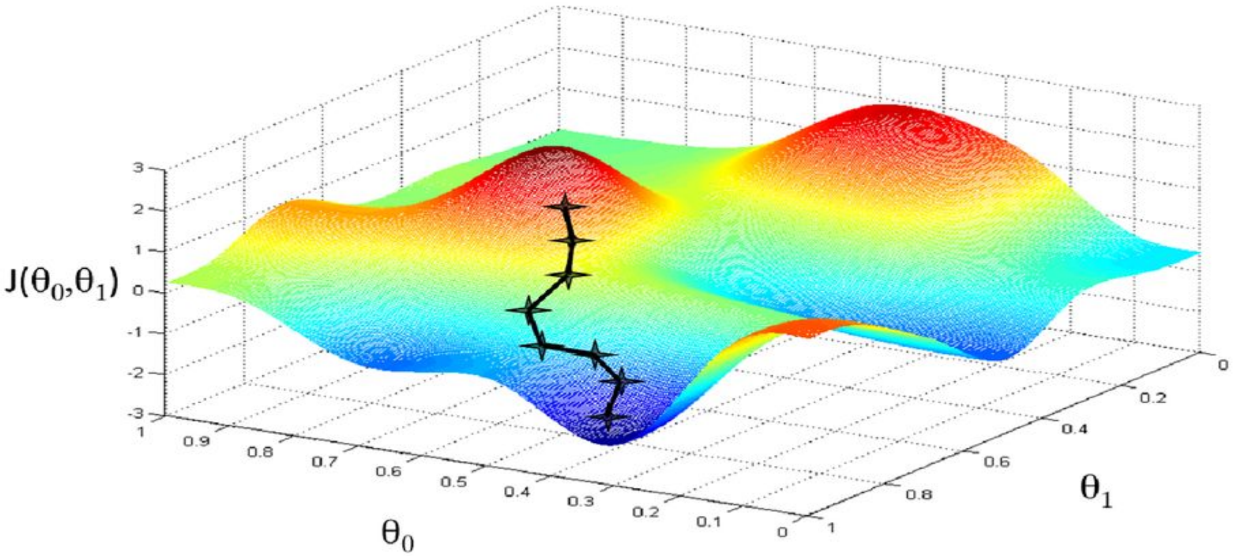
$$a = a - \alpha \frac{\partial J(a, b)}{\partial a}$$

$$b = b - \alpha \frac{\partial J(a, b)}{\partial b}$$

$$J : \ Loss\ Function$$
$$\alpha : Learning\ rate$$
$$a : \ Slope$$
$$b : \ Intercept$$

Effect of Learning Rate (α) on Gradient Descent





Gradient Descent

QUIZ!

Google Developer Groups

If the learning rate is too small, gradient descent will converge very slowly?

What happens if the learning rate is too high?

Do not edit
How to change the design

Presenting with animations, GIFs or speaker notes? Enable our Chrome extension

slido

?