Google Developer Groups

ML Math Club: Calculs for optimization

```
lookup.KeyValue
f.constant(['em
=tf.constant([0
.lookup.StaticV

_buckets=5)
```


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
7. Q&A

Google Developer Groups

Google Developer Groups

Recap

Where we left off last session?

In the last session we have seen the secrets of matrix transformations and eigenvalues.

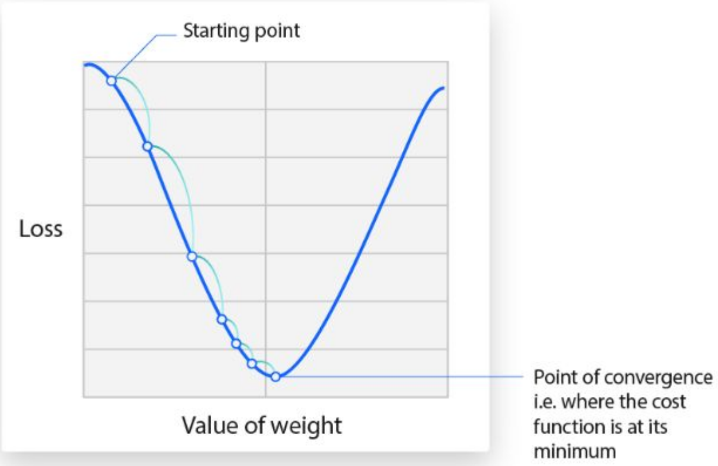
what is mathematical optimization ?

“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 key to training neural networks by minimizing the loss function to achieve the best performance.



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

What else can we do with optimization?

Optimization is a branch of applied mathematics which is useful in many different fields. Here are a few examples:

• Manufacturing

• Production •

• Transportation

• Scheduling

• Finance

etc.

slido

?

ⓘ Click **Present with Slido** or install our [Chrome extension] to activate this poll while presenting.

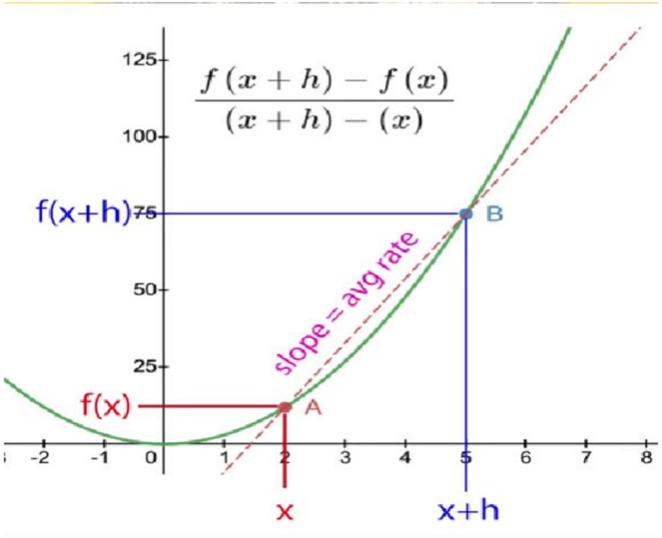
Google Developer Groups

Derivatives and Partial Derivatives

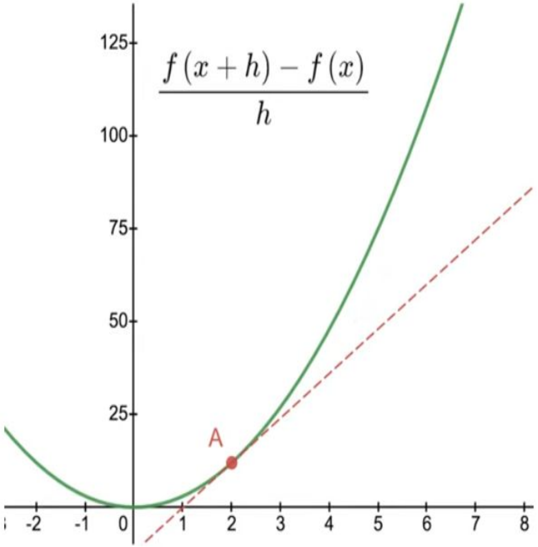
Derivatives :

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/zusk0tkwn6?lang=fr



https://www.desmos.com/calculator/ic9r3mgsxu?lang=fr
https://www.youtube.com/watch?v=-bSm-sjr22g&t=28s&ab_channel=ChrisOzarka

Partial Derivatives

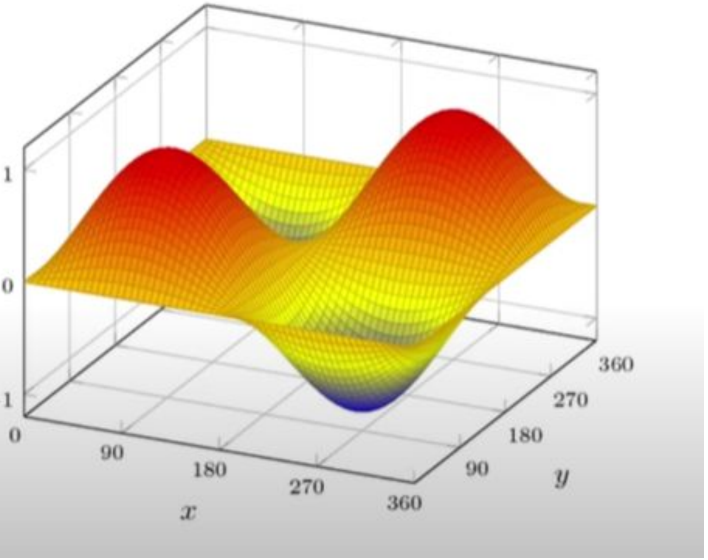
A partial derivative measures the rate of change of a function with respect to one of its variables while keeping all other variables constant.

notation

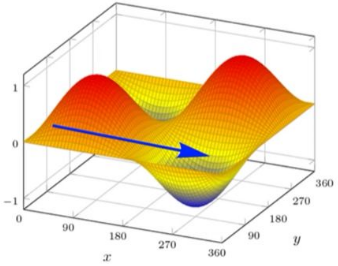
$$\frac{\partial f}{\partial x}(x,y) = \lim_{h\to 0} \frac{f(x+h,y)-f(x,y)}{h},$$
$$\frac{\partial f}{\partial y}(x,y) = \lim_{h\to 0} \frac{f(x,y+h)-f(x,y)}{h}$$

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

f(x, y)

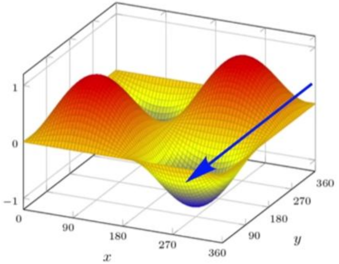


f(x, y)



erivative **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?lang=fr

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 :

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

Example: Let's compute together the gradient of this function:

https://www.desmos.com/3d/out3uua3js?lang=fr

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

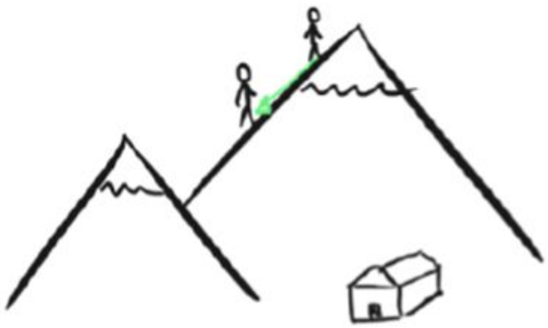
```
lookup.KeyValue
f.constant(['em
=tf.constant([0
.lookup.StaticV

_buckets=5)
```

Understanding Gradient Descent

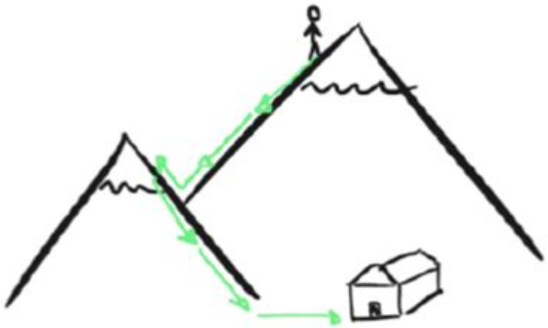
Imagine being lost in the mountains. Your goal is to reach the shelter at the lowest point in the valley. Without a map, you don't know the shelter's exact location—you must find it on your own.

solution

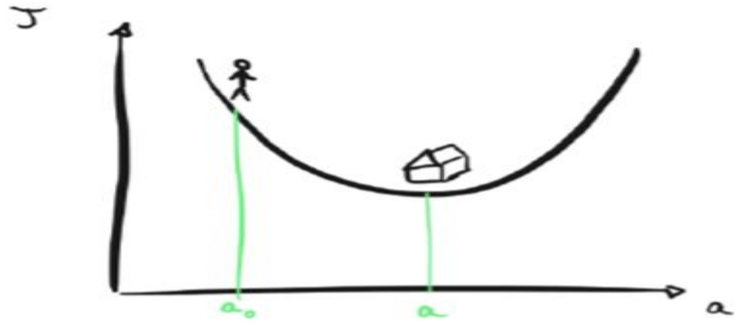
Step 1: Find the steepest slope



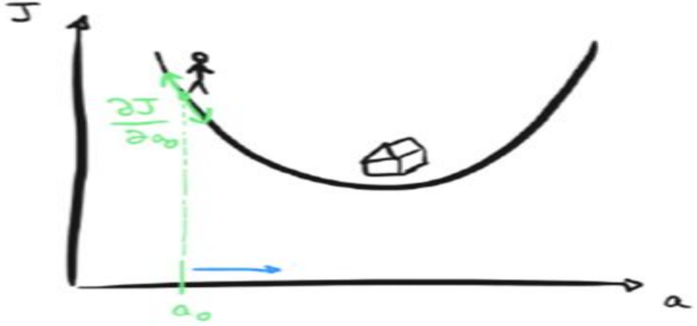
Step 2: Walk some distance in that direction



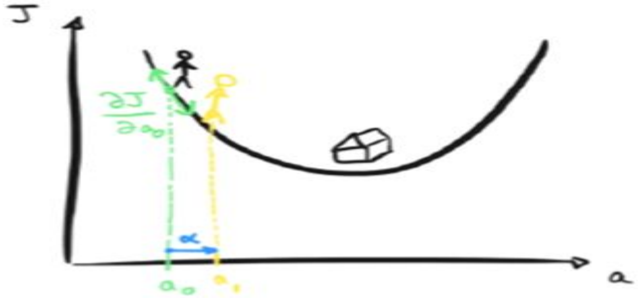
Step 3: Repeat steps 1 and 2 in a loop



Step 1: Calculate the gradient



Step 1: Calculate the gradient



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:

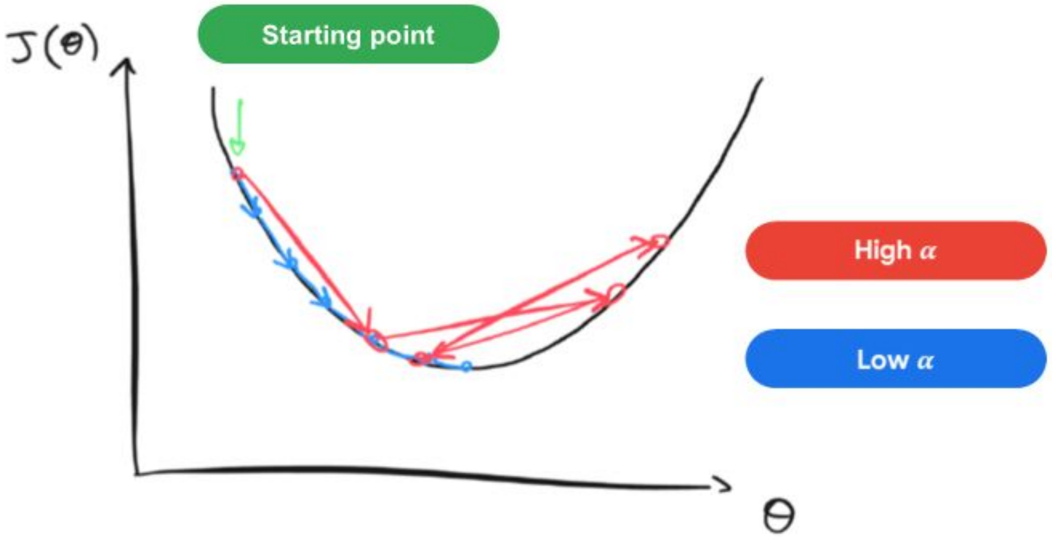
$$a = a - \alpha \frac{\partial J(a,b)}{\partial a}$$

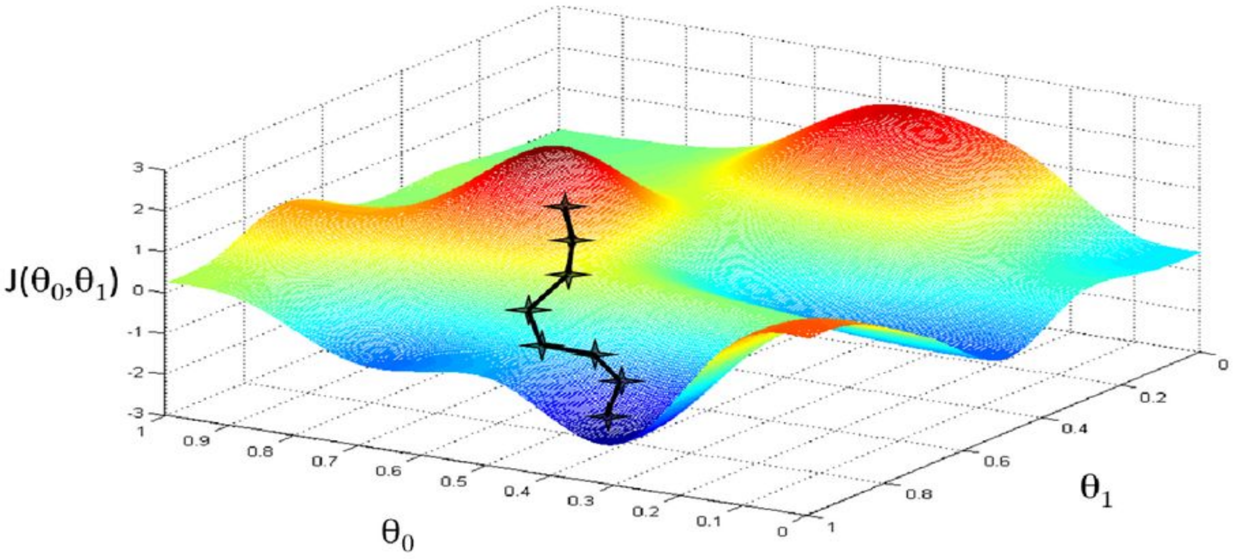
$$b = b - \alpha \frac{\partial J(a,b)}{\partial b}$$

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

Step 1: Calculate the gradient

Effect of Learning Rate (α) on Gradient Descent





Gradient Descent

https://github.com/greyhatguy007/Machine-Learning-Specialization-Coursera/blob/main/C1%20-%20Supervised%20Machine%20Learning%20-%20Regression%20and%20Classification/week1/Optional%20Labs/C1_W1_Lab05_Gradient_Descent_Soln.ipynb
https://www.engineerknow.com/2021/12/gradient-descent-with-interactive.html

QUIZ!

Google Developer Groups

?