

Foundations of Machine Learning in Python

Moritz Wolter

July 12, 2022

High Performance Computing and Analytics Lab, Uni Bonn

Derivatives and Gradients

Optimization

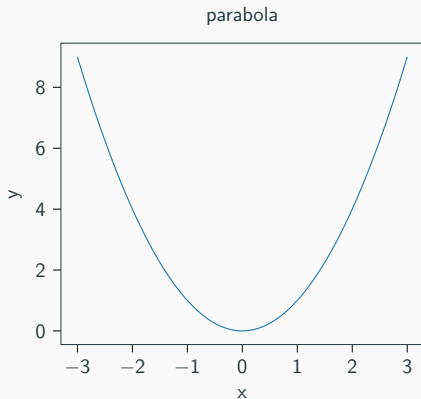
Traditionally, optimization means minimizing using a cost function $f(x)$. Given the cost, we must find the cheapest point x^* on the function, or in other words,

$$x^* = \min_{x \in \mathbb{R}} f(x) \quad (1)$$

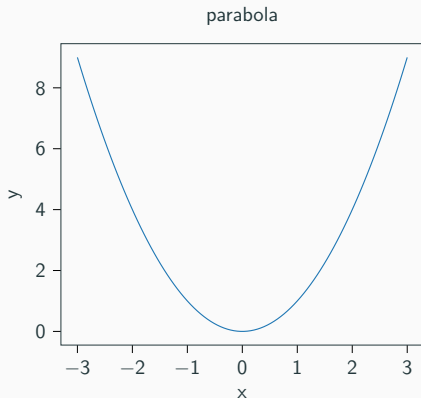
Functions

Functions are mathematical mappings. Consider for example the quadratic function, $f(x) : \mathbb{R} \rightarrow \mathbb{R}$:

$$f(x) = x^2 \quad (2)$$



Where is the minimum?



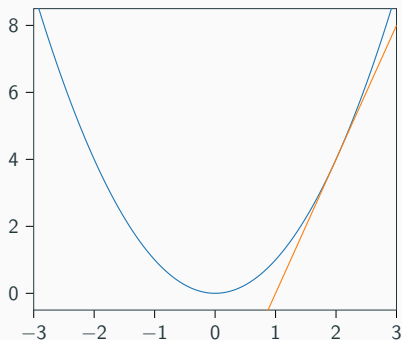
In this case we immediately see it's at zero. Finding it algorithmically requires derivative information.

Derivatives and Gradients

The derivative

$$\frac{df(x)}{dx} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \quad (3)$$

parabola with derivative at two



Derivation of the parabola derivative

$$\lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} \Leftrightarrow \lim_{h \rightarrow 0} \frac{x^2 + 2x + h^2 - x^2}{h} \quad (4)$$

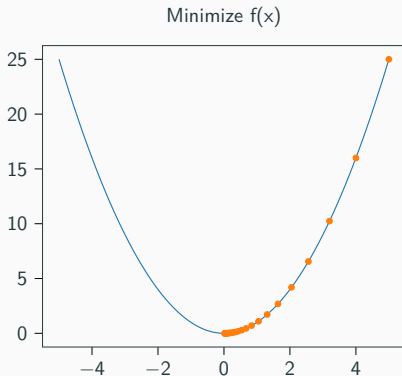
$$\Leftrightarrow \lim_{h \rightarrow 0} \frac{2xh + h^2}{h} \Leftrightarrow \lim_{h \rightarrow 0} \frac{h(2x + h)}{h} \quad (5)$$

$$\Leftrightarrow \lim_{h \rightarrow 0} 2x + h \Leftrightarrow 2x \quad (6)$$

Steepest descent on the parabola

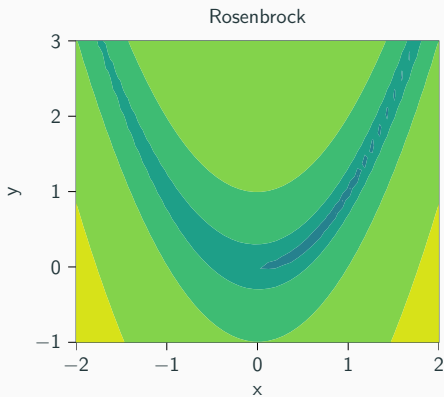
$$x_n = x_{n-1} - \alpha \cdot \frac{df}{dx} \quad (7)$$

Working with $x_0 = 5$ and $\alpha = 0.1$ for 25 steps leads to:



Multidimensional problems

$$f(x, y) = (a - x)^2 + b(y - x^2)^2 \quad (8)$$



The gradient

$$\nabla \mathbf{x} \quad (9)$$

TODO

Gradient descent on the Rosenbrock function

Rosenbrock Optimization

Rosenbrock Optimization

Optimization

TODO