Foundations of Machine Learning in Python

Moritz Wolter

July 7, 2022

High Performance Computing and Analytics Lab, Uni Bonn

Overview

Derivatives and Gradients

Optimization

Optimization

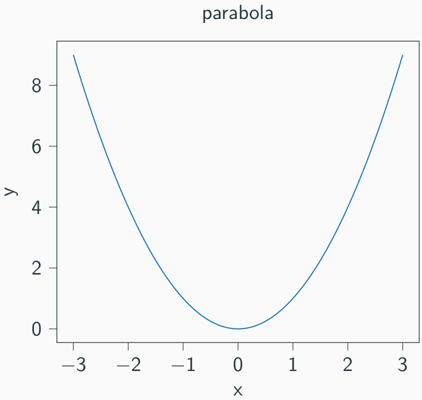
Traditionally, optimization means minimizing using a cost function $f(x)$. Given the cost, we must find the cheapest point x^* on the function, or in other words,

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

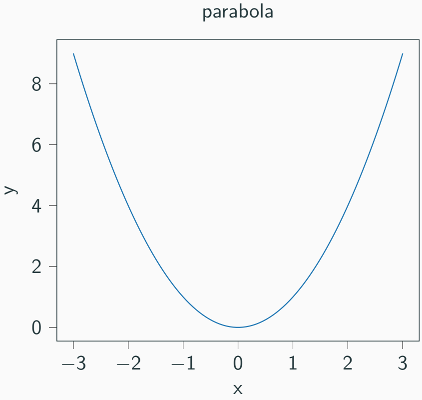
Functions

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

$$f(x) = x^2 \tag{2}$$



Where is the minimum?



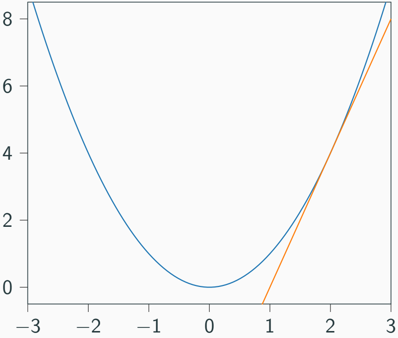
In this case we immediately see it's at zero. Finding it algorithmically requires derivate information.

Derivatives and Gradients

The derivative

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

parabola with derivative at two



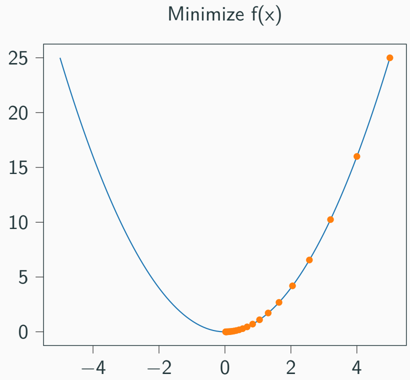
Derivation of the parabola derivative

TODO

Steepest descent on the parabola

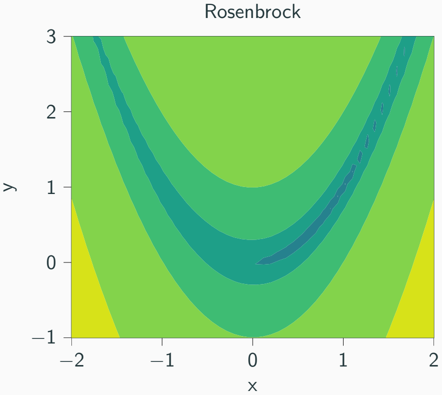
$$x_n = x_{n-1} - \alpha \cdot \frac{df}{dx} \tag{4}$$

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



Multidimensional problems

$$f(x, y) = (a - x)^2 + b(y - x^2)^2 \tag{5}$$



The gradient

$$\nabla \mathbf{x} \tag{6}$$

Gradient descent

TODO

Optimization

Frame 2

TODO