Gradient Descent

In this lesson, we'll learn the principles and the math behind the gradient descent algorithm.

Gradient Calculation

In the last few videos, we learned that in order to minimize the error function, we need to take some derivatives. So let's get our hands dirty and actually compute the derivative of the error function. The first thing to notice is that the sigmoid function has a really nice derivative. Namely,

$$\sigma'(x) = \sigma(x)(1 - \sigma(x))$$

The reason for this is the following, we can calculate it using the quotient formula:

$$\begin{aligned} \sigma'(x) &= \frac{\partial}{\partial x}\frac{1}{1+e^{-x}} \\ &= \frac{e^{-x}}{(1+e^{-x})^2} \\ &= \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{1+e^{-x}} \end{aligned}$$