A short mathematics preliminaries

Gradient example

Note that in practice when we differentiate f with respect to x then y should be treated as a constant.

Example

Let $f(x, y) = \exp(2xy) + x^2y^3 - x\sin(y)$. Compute ∇f.

Solution.

We have

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

and

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

Thus,

$$\nabla f = \left(2y\exp(2xy) + 2xy^3 - \sin(y), 2x\exp(2xy) + 3x^2y^2 - x\cos(y)\right).$$

Gradient descent example 1D

Let $f(x) = x^2$. This function has its (unique) minimum at $x = 0$.

We will illustrate the gradient descent method by choosing an initial $x_0 := 5$, and we will use a learning rate $\alpha := 0.1$.

The first few steps yields

$$\begin{aligned} x_1 &= 5 - 0.1 \cdot f'(5) = 5 - 0.1 \cdot 10 = 4 \\ x_2 &= 4 - 0.1 \cdot f'(4) = 3.2, \\ x_3 &= 3.2 - 0.1 \cdot f'(3.2) = 2.56. \\ &\dots \\ x_{100} &\approx 1.09 \cdot 10^{-9}. \end{aligned}$$

The multivariable case of gradient descent involves the gradient instead of the derivative.