Problem set: Gradient descent

Exercise (i) (7P) Enter the missing code snippets in the jupyter notebook. Partial credit will be awarded.

Definition 0.1. For a matrix $Q \in \mathbb{R}^{m \times d}$ its **operator norm** is defined as

$$\|Q\|_{op} := \sup_{x \in \mathbb{R}^d : \|x\| \leq 1} \|Qx\|,$$

but for convenience we will mostly write $\|Q\|$ (so no subscript) but mean the operator norm.

Lemma 0.1. The operator norm of a matrix Q fulfills

$$\|Qx\| \leq \|Q\|_{op} \|x\|.$$

In particular, every linear map (given by a matrix Q) is Lipschitz continuous with Lipschitz constant $\|Q\|$.

Exercise (ii) (2P) Prove that the quadratic function

$$f(x) = \frac{1}{2} x^T Q x + b^T x + c$$

is **smooth** with parameter $\|Q\|$. Note that without loss of generality Q can be assumed to symmetric, however, this should not change anything.

Exercise (iii) (1P) Suppose that we have observations (x_i, y_i) which are **centered**, meaning that $\sum_{i=1}^n x_i = 0 = \sum_{i=1}^n y_i$. Let (b^*, w^*) be the global minimum of the least squares objective

$$f(b, w) = \sum_{i=1}^n \left(b + w^T x_i - y_i\right)^2.$$

Prove that $b^* = 0$.