Machine Learning I: Foundations
Exercise Sheet 5

Prof. Marius Kloft TA:Billy Joe Franks

19.05.2020

Deadline: 02.06.2020

1) (MANDATORY) 10 Points

In this exercise we will try to understand gradient descent, its initialization value, and its learning rate schedule. Consider only functions

$$f : \mathbb{R} \to \mathbb{R}.$$

a) Find a constant learning rate schedule ($\lambda_i = c$), a convex function f (with a global minimum), and an initialization value x_0 such that the global minimum is never reached if you apply gradient descent with λ_i on f starting at x_0. Prove that the function is convex. Prove that the global minimum is never reached.

b) For your choice of convex function and initialization value, is it possible to choose a learning rate schedule (constant or otherwise) such that the global minimum is reached? Prove your claim.

c) Give an example of an unbounded function ($f(\mathbb{R}) = \mathbb{R}$), a learning rate schedule, and an initialization value for which gradient descent converges to a plateau (a critical point that is neither a minimum nor a maximum). The initialization value can not be chosen as this plateau.

d) Consider $f(x) = x^3$, is it possible to find a learning rate schedule which converges to the plateau at $x = 0$ for any initialization value? Prove your claim.

2) Let $k(\cdot, \cdot)$ be a kernel on $\mathbb{R}^d$. Let $\phi(\cdot)$ be the kernel mapping, i.e. $\langle \phi(x), \phi(y) \rangle = k(x, y)$. Let $x_1, \dots, x_n \in \mathbb{R}^d$, $a = [a_1, \dots, a_n]^T \mathbb{R}^n$ and $b = [b_1, \dots, b_n]^T \mathbb{R}^n$. Let $K \in \mathbb{R}^{n \times n} = [k(x_i, x_j]_{i,j}$ be the kernel matrix.

Prove that

$$\left\langle \sum_{i=1}^{n} a_i \phi(x_i), \sum_{j=1}^{n} b_j \phi(x_j) \right\rangle = a^T K b$$

3) For a matrix $X \in \mathbb{R}^{m \times n}$ let $X_{i,:} = [X_{i,1}, \ldots, X_{i,n}]$ be the i-th row vector and $X_{:,i} = [X_{1,i}, \ldots, X_{m,i}]^T$ be the i-th column vector. For $X \in \mathbb{R}^{m \times n}$ and $Y \in \mathbb{R}^{n \times q}$ show that

$$XY = [X_{i,:}Y_{:,j}]_{i,j}$$

and

$$XY = \sum_{i=1}^{n} X_{:,i}Y_{i,:}.$$

Be sure to note the orientations of the vectors, some of these are row vectors and others are column vectors.

4) Solve programming task 5.