Homework Assignment 2

2-INF-150: Machine Learning, Fall 2024

Deadline: 2.12.2024, 22:00
(electronic submission through classroom)

Submit answers to theoretical tasks in **one pdf file per task** in google classroom (your answer can be typeset or scanned, but please make sure that everything is legible and easy to read). **No word, jpgs, or other formats!** No late submissions are allowed. Write your solutions so that they contain all information necessary to easily understand them, but at the same time try to aim for brevity. Prove all claims, including in the cases when it is not explicitly written in the problem statement.

You can write you solutions in Slovak or English. The solutions must be your work. Do not copy from others and do not attempt to find the solutions in literature or on the internet! For more details on permissible forms of collaboration check the course web page.

1. Dual programs. In this task, we will revisit the problem of computing the distance of a point from a hyperplane. In the lecture, we have solved this problem using geometric reasoning. Since then, we have learned about dual programs and Lagrange multipliers. Here is the extended version of the strong duality theorem; compared to the version that we used in class, this one also includes equality constraints.

> Let $f : R^n \to R$ and $g : R^n \to R^m$ are convex functions, $h : R^n \to R^k$ is affine, and X is a closed convex set over R^n. Let us assume that there exists $\hat{x} \in X$ such that $g(\hat{x}) < 0$ and $h(\hat{x}) = 0$.
>
> Then if x^* minimizes $f(x)$ subject to $g(x) \leq 0$, $h(x) = 0$, $x \in X$ (primal program) and $(\lambda_g^*, \lambda_h^*)$ maximizes $L(\lambda_g, \lambda_h)$ subject to $\lambda_g \geq 0$ (dual program), then $f(x^*) = L(\lambda_g^*, \lambda_h^*)$.
>
> Here, $L(\lambda_g, \lambda_h) = \min_{x \in X}\{f(x) + \langle \lambda_g, g(x) \rangle + \langle \lambda_h, h(x) \rangle\}$, vectors λ_g, λ_h are so called Lagrange multipliers.

a) Using a quadratic program, formulate the problem of finding the distance of point y from hyperplane given by equation $\langle \theta, x \rangle - b = 0$.

   Note: Distance of point y from the hyperplane is in fact the distance from the closest point z that is located on the hyperplane.

b) Use the method of Lagrange multipliers to create a dual program.

d) Find the solution of the dual program. Can you, in addition to the distance, find point z located on the hyperplane closest to point y?

2. Taxi drivers. On the web page `https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page` you will find a public data set containing very detailed information about taxi rides in New York. Besides a boring task of finding dependencies between individual features and the sum paid for the ride, this data set can be used in much more creative ways.

a) Download the data set, study its structure, and propose one non-traditional regression problem. You should also outline why do you think that the data set can be used to answer the question that you proposed.

   Ideally, your problem will have nothing to do with taxi drivers or taxi rides, but instead would deal with cultural aspects, society, or transit in New York instead.

b) Similarly as in the previous question, propose one non-traditional classification problem.

c) **Bonus.** Choose one of the problems that you have formulated and analyse the data set. Hand in a detailed description of the methods and results, link to the source code (i.e. on github).

Note: If you need an inspiration, you can read the paper *When Bankers Go to Hail: Insights into Fed–Bank Interactions from Taxi Data*—see `https://doi.org/10.1287/mnsc.2023.4885`