Homework Assignment 1

2-INF-150: Machine Learning, Fall 2024

Deadline: 4.11.2024, 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. Regression with an exponential model. Consider a regression problem with the set of hypotheses:

$$H = \{h_{a,b} : x \to ae^{-x/b}\}.$$

Note that a and b are model parameters. Propose an appropriate training algorithm. Choice of an appropriate loss function is up to you, but clearly describe and justify your choice.

2. Machine Learning Theory. Consider a regression problem with the set of hypotheses:

$$H = \{h_b : x \to 2x + b\}.$$

a) Describe an algorithm that for the training set $(x_1, y_1), \ldots, (x_t, y_t)$ finds the hypothesis minimizing the error function $J(b) = \frac{1}{t}\sum_{i=1}^{t}(h_b(x_i) - y_i)^2$.

   (Your algorithm should be MUCH simpler than the algorithms used for the traditional linear regression.)

b) Consider a probabilistic distribution $P_{x,y}$ defined as follows:

   - distribution of x is uniform over the interval $[0, 100]$
   - for given x, $\Pr(y = 2x + 3 \mid x) = 0.3$ and $\Pr(y = 2x - 2 \mid x) = 0.7$ (there are no other values of y that can occur in combination with x).

   Compute bias of the set of hypotheses H if we assume that the data will be independent samples from $P_{x,y}$.

c) For distribution $P_{x,y}$ from part b) and number of training examples $t = 1$, compute expected training and expected testing error.

d) For distribution $P_{x,y}$ from part b) and general value of t, compute expected training and testing error. How are these errors related to the bias computed in part b) as $t \to \infty$? (You can illustrate this by drawing a graph.)

   Based on your results, would you be able to give a guidance as to an appropriate size of the training set in this scenario?