

# Theoretical homework 1

due April 11 (Monday), 23:59.

Innopolis  
Machine learning course

*Recommendations: all solutions should be short, mathematically strict (unless qualitative explanation is needed), precise with respect to the stated question and clearly written. The scan of the solution should be sent to v.v.kitov@yandex.ru with your name, surname and word “innopolis” in the title. Please send before deadline, because late submissions will be downgraded by 50%.*

1. **Bootstrap properties:** In bagging from original training set of size  $N$  we generate bootstrap pseudosamples of size  $N$ . Let's fix some observation in the training set, for example  $(x_1, y_1)$  and consider one bootstrap pseudosample.
  - (a) What is the probability that this observation will not appear anywhere in the selected bootstrap pseudosample?
  - (b) What is the limit of this probability when  $N \rightarrow \infty$ ?
2. **Constant approximator:** Consider points  $z_1, z_2, \dots, z_N$ . Find such constant approximation  $\mu$  of these points, so that the sum of absolute deviations from these points to  $\mu$   $\sum_{n=1}^N |z_n - \mu|$  is minimized.
3. **Universal functional normalizer:** Prove that if particular feature  $x^i$  has arbitrary random distribution with continuous cumulative distribution function  $F(u) = P(x^i \leq u)$ , then monotonous transformation with  $F$  will give uniformly distributed feature:

$$F(x^i) \sim \text{Uniform}[0, 1]$$