# Theoretical homework 1

Innopolis, Algorithms of machine learning, Autumn 2016.

due 23:59 16.09.2016.

1. Suppose you have two subordinates that propose forecasting algorithms. One subordinate applied his algorithm once to the data and obtained classification accuracy 0.8. Another subordinate tried 100 different algorithms and achieved accuracy 0.95. What are the risks of preferring the solution of the 2nd subordinate to the solution of the 1st subordinate?

2. Suppose $x$ is some random feature with continious cumulative distribution function $F(u) = P(x \leq u)$. Prove that transformed feature $z = F(x)$ is uniform random variable distributed on $[0, 1]$.

3. Suppose we need to estimate the center of distribution of real numbers $z_1, z_2, ...z_D$ in terms of their robustness to outliers (rare values which have are very distant from other values and have very low probability).

   (a) Will arithmetic average be robust to outliers? Why?

   (b) Will median be robust to outliers? Why?

4. Consider real numbers $z_1, z_2, ...z_N$. Find such constant approximation $\mu$ of these numbers, so that

   (a) the sum of absolute deviations from these points to $\mu$ $\sum_{n=1}^{N} |z_n - \mu|$ is minimized.

   (b) the sum of square deviations from these points to $\mu$ $\sum_{n=1}^{N} (z - \mu)^2$ is minimized.