Homework 5
due December 4 (Friday), 16-00.

<div align="center">

Skoltech
Machine learning course
November-December 2015

</div>

Recommendations: all solutions should be short, mathematically strict (unless qualitative explanation is needed), precise with respect to the stated question and clearly written.

1. Let $H(Z) = -\sum_z p(z) \ln(z)$ be the entropy of random variable Z (where z denotes any of its values) and let $H(Y|X)$ be the conditional entropy of random variable Y given random variable X. Prove that mutual information between random variables X and Y $MI(X, Y) = H(Y) - H(Y|X)$ can be equivalently defined as Kullback-Leibler divergence between $p(x, y)$ and $g(x, y) = p(x)p(y)$ distribution functions (here $p(x, y) = p(X = x, Y = y)$ and $p(x), p(y)$ are marginal distributions of $p(x, y)$).

2. Prove that $K(x, z) = e^{-\gamma \|x-z\|^2}$ is a valid Mercer kernel, where $\|u\| = \sqrt{\langle u, u \rangle}$.

3. Prove that weights $\{w_i^{m+1}\}_{i=1}^N$ on $m + 1$ iteration of discrete AdaBoost satisfy the property that m-th base learner $h^m(x)$ always gives weighted error rate equal to 0.5 on the training set with these weights.