A linear-time adaptive nonparametric two-sample test[*]

Zoltán Szabó (CMAP, École Polytechnique)

Abstract

In my talk, I am going to focus on the two-sample testing problem: we are given two sets of observations, and our goal is to check whether the two sets are statistically indistinguishable (this may be thought of as a generalization of a t-test). We propose to choose features underlying the test in an adaptive way so as to maximize the distinguishability of the distributions by optimizing a lower bound on the test power. The constructed nonparametric t-test is based on a novel representation of probability distributions by analytic functions, and runs in linear time with respect to the number of samples. The efficiency of the method is illustrated in natural language processing (distinguishing articles from two categories) and computer vision (differentiating positive and negative emotions). The presented linear-time test achieves comparable performance to state-of-the-art quadratic-time alternatives, while returning human-interpretable features that explain the test results.

- Paper: Wittawat Jitkrittum, Zoltán Szabó, Kacper Chwialkowski, Arthur Gretton. Interpretable Distribution Features with Maximum Testing Power. In NIPS-2016, pages 181-189.
- Code: `https://github.com/wittawatj/interpretable-test`