
Outlier-Robust Divergence Estimation on Kernel-Endowed Domains with Median of Means*

Zoltán Szabó[†]

Abstract

Maximum mean discrepancy (MMD, also called energy distance or N-distance in statistics) is probably the most influential divergence measure on kernel-endowed domains with large number of successful applications. When the underlying kernel is unbounded (examples include polynomial, exponential, string or graph kernels), however even a single outlier can severely affect the existing MMD estimators. In order to overcome this serious sensitivity problem, I will present a new class of MMD estimators based on the median of means principle with excessive resistance properties to outliers, optimal sub-Gaussian deviation bounds under mild assumptions, and illustrations in discrimination of DNA subsequences.

- Preprint: <https://arxiv.org/abs/1802.04784>,
- Code: <https://bitbucket.org/TimotheeMathieu/monk-mmd/>

*MathoDS 3, City University of Hong Kong (CityU), Hong Kong, China, 19-23 June 2019; abstract.

[†]Joint work with Matthieu Lerasle¹, Timothée Mathieu² and Guillaume Lecué³. ¹Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud; CNRS, Université Paris Saclay, France. ²Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, France. ³CREST ENSAE ParisTech, France.