

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

MATTHIEU LERASLE, ZOLTÁN SZABÓ*, TIMOTHÉE MATHIEU, GUILLAUME LECUÉ

Center of Applied Mathematics (CMAP), École Polytechnique, France

Email: zoltan.szabo@polytechnique.edu

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/>