Information Theory, Kernels & Applications

Zoltán Szabó
(Dept. of Statistics, LSE)

Research interests

- stat.ML: information theory, kernels, scalable computation and their applications.

Research interests

- stat.ML: information theory, kernels, scalable computation and their applications.
- Dissemination:
  - machine learning (ICML, NeurIPS, AISTATS, UAI, JMLR),
  - control (IFAC WC), computer vision (CVPR, ECCV), and signal processing (EUSIPCO, LVA/ICA) venues.

Kernels: why?

1. Generalized inner product: various data types,
   - strings, time series, trees, groups and specifically rankings, sets, probability measures, various generative models, fuzzy domains, graphs, ...

Kernels: why?

1. Generalized inner product: various data types,
   - strings, time series, trees, groups and specifically rankings, sets, probability measures, various generative models, fuzzy domains, graphs, ...
2. Rich function class:
   - denseness in $C_b(\mathcal{X})$,
   - represents distributions and independence.

Kernels: why?

1. Generalized inner product: various data types,
   - strings, time series, trees, groups and specifically rankings, sets, probability measures, various generative models, fuzzy domains, graphs, ...
2. Rich function class:
   - denseness in $C_b(\mathcal{X})$,
   - represents distributions and independence.
3. Computationally tractable: $k(x_i, x_j)$.

Kernels: why?

1. Generalized inner product: various data types,
   - strings, time series, trees, groups and specifically rankings, sets, probability measures, various generative models, fuzzy domains, graphs, ...
2. Rich function class:
   - denseness in $C_b(\mathcal{X})$,
   - represents distributions and independence.
3. Computationally tractable: $k(x_i, x_j)$.
4. Hilbert structure of RKHSs: statistical analysis.

Kernels: why?

1. Generalized inner product: various data types,
   - strings, time series, trees, groups and specifically rankings, sets, probability measures, various generative models, fuzzy domains, graphs, ...
2. Rich function class:
   - denseness in $C_b(\mathcal{X})$,
   - represents distributions and independence.
3. Computationally tractable: $k(x_i, x_j)$.
4. Hilbert structure of RKHSs: statistical analysis.
5. vRKHSs: encode dependency among output coordinates.

Applications

- economics (consumer behavior, production function), finance (portfolio optimization, risk estimation), gene analysis, aerosol prediction, climate data analysis, Bayesian inference and adaptive sampling, criminal data analysis, emotion and gesture recognition, emotion transfer, convoy trajectory reconstruction, NLP (automatic word puzzle generation, Wikification, corrupted word determination, interpretable topic discrimination), collaborative filtering, hyperspectral anomaly detection, outlier-robust image registration, inpainting of natural images, blind source separation (with ECG and sound signals), cost-sensitive classification, quantile regression, density level set estimation, hypothesis testing, analysis of safe aircraft departure trajectories, safety-critical control.

Applications

- ITE toolbox (BitBucket):
  - 53 entropy, independence, divergence, association measures and kernels of probability distributions,
  - (by at least an order) the largest package in the domain,
  - $>$ 85 successful projects worldwide.

Some maths results

6. Consistent distribution regression (17-year open).

Some maths results

5. Linear time & interpretable hypothesis testing:
   - best paper award @ NeurIPS (3/3240).
6. Consistent distribution regression (17-year open).

Some maths results

4. Validness of HSIC and MMD (2008-).
5. Linear time & interpretable hypothesis testing:
   - best paper award @ NeurIPS (3/3240).
6. Consistent distribution regression (17-year open).

Some maths results

3. Scaling kernel methods:
   - 10-year test-of-time award (NIPS-2017),
   - exponential improvement: kernels & derivatives.
4. Validness of HSIC and MMD (2008-).
5. Linear time & interpretable hypothesis testing:
   - best paper award @ NeurIPS (3/3240).
6. Consistent distribution regression (17-year open).

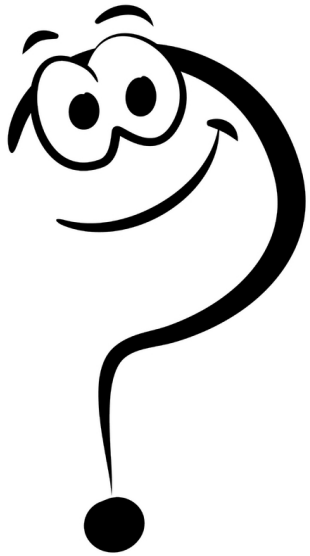
Some maths results

2. Outlier-robust mean embedding & MMD estimators:
   - mean estimation in $\mathbb{R}^d$, optimal sub-G deviation bounds ('71-, $\mathcal{O}\left(N^{24}\right)$),
   - RKHS & algorithms: $\mathcal{O}(N^2) - \mathcal{O}(N^3)$ scaling.
3. Scaling kernel methods:
   - 10-year test-of-time award (NIPS-2017),
   - exponential improvement: kernels & derivatives.
4. Validness of HSIC and MMD (2008-).
5. Linear time & interpretable hypothesis testing:
   - best paper award @ NeurIPS (3/3240).
6. Consistent distribution regression (17-year open).

Some maths results

1. Hard shape-constrained prediction in RKHSs.
2. Outlier-robust mean embedding & MMD estimators:
   - mean estimation in $\mathbb{R}^d$, optimal sub-G deviation bounds ('71-, $\mathcal{O}\left(N^{24}\right)$),
   - RKHS & algorithms: $\mathcal{O}(N^2) - \mathcal{O}(N^3)$ scaling.
3. Scaling kernel methods:
   - 10-year test-of-time award (NIPS-2017),
   - exponential improvement: kernels & derivatives.
4. Validness of HSIC and MMD (2008-).
5. Linear time & interpretable hypothesis testing:
   - best paper award @ NeurIPS (3/3240).
6. Consistent distribution regression (17-year open).

Thank you for the attention!



Website: https://zoltansz.github.io/