Differential Privacy in Scalable General Kernel Learning via K-means Nyström Random Features

KAIST Department of Mathematical Sciences

Bonwoo Lee, Jeongyoun Ahn, Cheolwoo Park

KAIST

NEURAL INFORMATION PROCESSING SYSTEMS

Differentially private kernel learning

Differential Privacy (DP)

: A standard that ensures privacy for machine learning (ML) algorithms.

- Growing data consumption of machine learning industries.
- Rising concerns on data privacy.

Kernel learning

: ML algorithms that can model various data structure via appropriate kernel.

Ex.
- Polynomial kernel $k(x,y) = (\langle x,y\rangle + c)^d$ for polynomial data.
- Graph kernel for graph data.

Undoubtably, private kernel learning method has great importance in modern ML era. However, existing methods suffer some drawbacks for practical use : time consuming, limited to special kernels or loss function, requires the knowledge on test data.

Table 1: Comparison of DP kernel ERM algorithms in terms of restrictions for privacy guarantee.

Algorithms	General kernels	Scalable	Test data free	General objective
Chaudhuri et al. (2011)	✗	✓	✓	✓
Jain and Thakurta (2013)	✓	✗	✗	✗
Hall et al. (2013)	✓	✗	✓	✗
Proposed	✓	✓	✓	✓

Goal: Private kernel learning algorithm applicable for general situations.

Challenges

Kernel learning wants to find a specific function in the (possibly) infinite dimensional space RKHS $\mathcal{H}_k$.

⇨ Infinite-dimensional amount of information are transported in learning.

⇨ Privacy protection for infinite amount of information!

= noise adding for infinitely many times

= significantly degrading the learning

Solution: Develop Private Nyström method

-Benefit: Nyström method provides a low-dimensional approximation in scalable kernel learning by subspace approximation. Its private version will allow faster, and more accurate DP kernel learning.

References

Chaudhuri, K., Monteleoni, C., and Sarwate, A. D. (2011). Differentially private empirical risk minimization. J. Mach. Learn. Res., 12(3):1069–1109.

Hall, R., Rinaldo, A., and Wasserman, L. (2013). Differential privacy for functions and functional data. J. Mach. Learn. Res., 14(1):703–727.

Jain, P. and Thakurta, A. (2013). Differentially private learning with kernels. In Proceedings of the 30th International Conference on Machine Learning, volume 28, pages 118–126

Balog, M., Tolstikhin, I., and Schölkopf, B. (2018). Differentially private database release via kernel mean embeddings. In Proceedings of the 35th International Conference on Machine Learning, volume 80, pages 414–422.

DP K -means Nyström random features

DP kernel learning using DP K-means Nyström random features.

1. First run $\epsilon/2$-DP K-means to find the K private centroids $\{z_1, \cdots z_K\}$ of data
2. Find the K-dimensional projection on the subspace in RKHS spanned by private centroids as:

$$[k(z_i,z_j)]_{K\times K} = U\Sigma U^T,\ \varphi^{(Nys)}(x) := \Sigma^{\dagger\frac{1}{2}} U^T [k(z_1,x),\cdots,k(z_K,x)]^T \in \mathbb{R}^K$$

3. Solve DP kernel learning by solving DP linear learning for K-dimensional transformed data: $\varphi^{(Nys)}(x_1),\cdots,\varphi^{(Nys)}(x_n)$.

Applications.

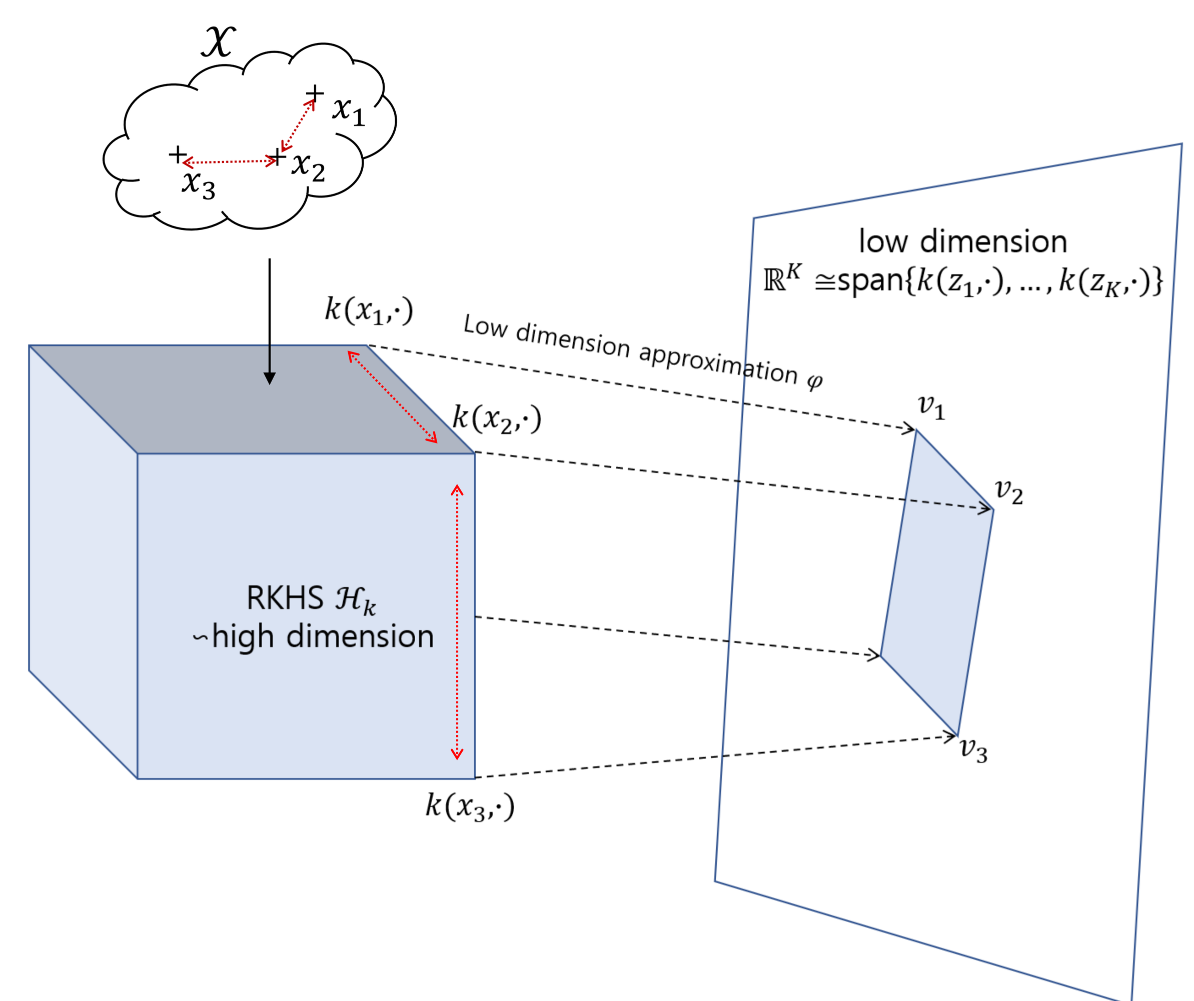
- Kernel empirical risk minimization.

$$\min_{f\in\mathcal{H}_k} \frac{1}{n}\sum_{i=1}^{n} l\left(f(x_i), y_i\right) \Rightarrow \min_{v\in\mathbb{R}^K} \frac{1}{n}\sum_{i=1}^{n} l\left(\langle v, \varphi^{(Nys)}(x)\rangle, y_i\right)$$

- Kernel maximum mean discrepancy

$$\left\|\frac{1}{n}\sum_{i=1}^{n} k(x_i,\cdot) - \frac{1}{n}\sum_{i=1}^{n} k(x_i',\cdot)\right\|_{\mathcal{H}_k} \Rightarrow \left\|\frac{1}{n}\sum_{i=1}^{n} \varphi^{(Nys)}(x_i) - \frac{1}{n}\sum_{i=1}^{n} \varphi^{(Nys)}(x_i')\right\|_2$$

(Kernel learning reduced to K-dimensional linear learning)

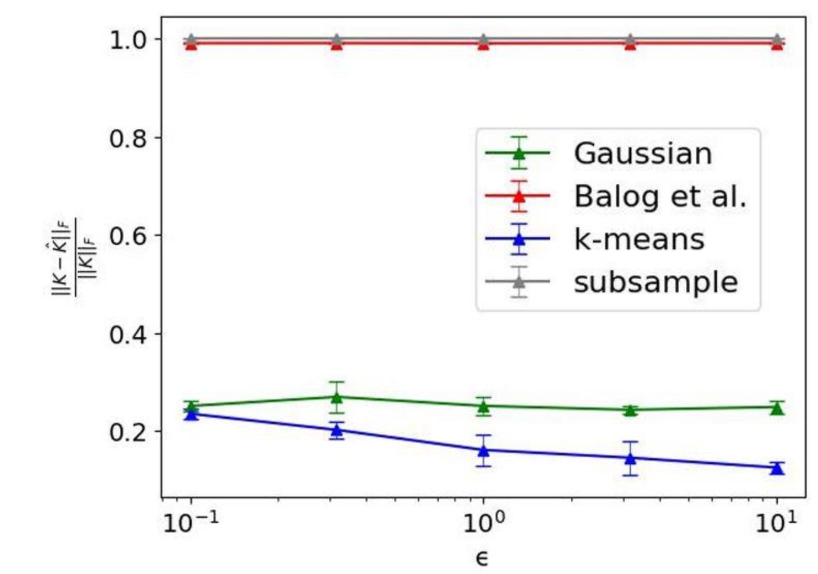


Benefits of DP K-means Nyström random features

- Robust to noise: privacy protection accompanies noise addition. However averaging included in the K -means procedure reduces the noise.
- Points close in data space is also close in RKHS.→K-means centroids can be used to generates subspace in RKHS approximating the whole data.

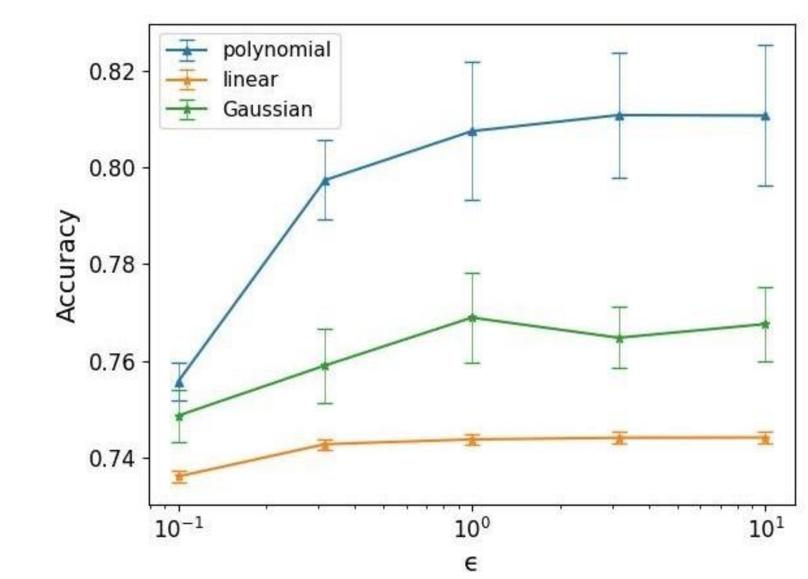
Random features quality analysis

- Quality of K-means Nyström random features: the average kernel approximation error, and kernel learning errors via Nyström random features depend on $\|K-\widehat{K}\|$ where norms are operator norm or Frobenius norm, where $\widehat{K} = [\langle\varphi^{(Nys)}(x_i), \varphi^{(Nys)}(x_j)\rangle]_{n\times n}$.
- The approximation error $\|K-\widehat{K}\|$ can be bounded by quantization error of $\{z_1,\cdots z_K\}$, which is controlled by DP K-means.



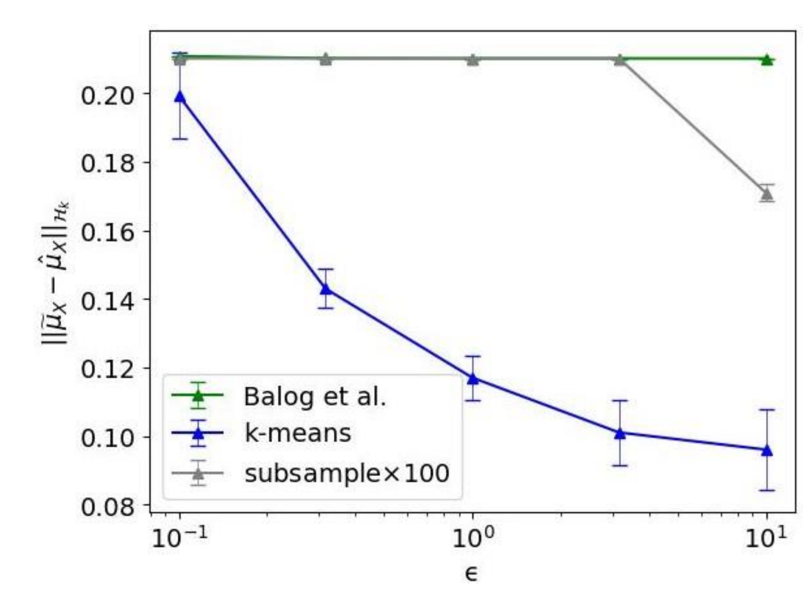
Comparison of relative approximation error between DP K -means Nyström random features and other existing approximation algorithms.

Experiments



Private binary classification of data having polynomial decision boundary.

- 1M data, 200 dimension.
- No test data knowledge.
- Existing methods can not run the kernel learning using the true kernel (=polynomial), and have low performances.



Private KME estimation.

- Adult dataset.
- Other possible variant of private Nyström method (subsample) was experimented.
- Outperforms other methods when dimension K is fixed.