



THE UNIVERSITY OF GEORGIA
DEPARTMENT OF STATISTICS

Colloquium Series

Bharath Sriperumbudur

Department of Statistic, Pennsylvania State University

3:45 PM, Thursday, September 24, 2020

Zoom link: <https://zoom.us/j/7979236528>

Approximate Kernel Principal Component Analysis: Computational vs. Statistical Trade-off

Kernel principal component analysis (KPCA) is a popular non-linear dimensionality reduction technique, which generalizes classical linear PCA by finding functions in a reproducing kernel Hilbert space (RKHS) such that the function evaluation at a random variable X has a maximum variance. Despite its popularity, kernel PCA suffers from poor scalability in big data scenarios as it involves solving a $n \times n$ eigensystem leading to the computational complexity of $O(n^3)$ with n being the number of samples. To address this issue, in this work, we consider a random feature approximation to kernel PCA which requires solving an $m \times m$ eigenvalue problem and therefore has a computational complexity of $O(m^3 + nm^2)$, implying that the approximate method is computationally efficient if $m < n$ with m being the number of random features. The goal of this work is to investigate the trade-off between computational and statistical behaviors of approximate KPCA, i.e., whether the computational gain is achieved at the cost of statistical efficiency. We show that the approximate KPCA is both computationally and statistically efficient compared to KPCA in terms of the error associated with reconstructing a kernel function based on its projection onto the corresponding eigenspaces.

For more information, please contact us at:

Phone: 706.542.5232 E-Mail: stat@uga.edu

Parking is available in the North Campus Parking Deck.

For a UGA Campus map, please see: <http://aviary.camplan.uga.edu/CampusMap/Default.aspx>