Meeting minutes on research project "nonparametric learning with high-dimensional matrix-valued features"

Discussants: Chanwoo Lee, Lexin Li, Miaoyan Wang, Helen Zhang

Date: June 5, 2020

[**Interpretability**] Visualization and feature selection are important problems relevant to predictive data analysis. Common approach for variable selection is based on ℓ-1 or ℓ-0 penalties. However, one need to take the union of selected features using a sequence of SVM classifiers, which is computationally expensive in high dimensions. Can we find an efficient, interpretable way to identify the most predictive features from matrix-valued inputs? In the same vein, visualization of feature matrices and conditional probabilities $P(y = 1|\boldsymbol{X})$ should be considered for good interpretation.

[**Exhaustiveness of SDR estimation**] A related technical issue with regards to interpretability is the exhaustiveness of SDR estimation. Our current theoretical result has yet to guarantee the exhaustiveness for the bilinear central subspace estimation. Non-exhaustiveness is a drawback for SIR-based method, especially with binary-valued outcomes. There are some work addressing the non-exhaustiveness in SDR (unsure whether they are specifically for SVM-based method). We should extend the results into our matrix-valued SDR framework.

[**Inference/uncertainty in estimation**]. Developing uncertainty quantification (such as confidence interval) for the point-wise probability estimates will have high impact. This problem is challenging in the full non-parametric setting; it could be possible, though, with some additional assumptions on data such as multivariate normal as in LDA. Incorporating inference into the current framework would greatly highlight our contribution over other approaches such as neural networks. Easiest (universal) approach for non-parametric inference is bootstrapping, but we would prefer a solution that is more problem-specific.

[**Symmetry in inputs**] Can our method adapt to symmetric matrices as inputs? How much modifications are needed to both theory and algorithm? This question is also related to the interpretability in applications.

[**Two goals in one framework**] Our current work addresses two goals simultaneously, one for probability estimation and another for sufficient dimension reduction. These two goals are related but also distinctive: (1) Both goals are "equally hard" when the outcome is binary. This is because, for binary outcomes, the conditional probability solely determines the conditional distribution. (2) The probability estimation focuses on prediction whereas the sufficient dimension reduction focuses on space selection. Perhaps we should choose a prevailing theme and present both goals as a coherent approach for matrix/tensor data analysis (as opposite to two disentangled tasks).

[**Higher-order tensors**] Most current results extend naturally to high-order tensors. Pre-

senting the results in terms of matrices highlights the key ideas in our work (low-rank projection kernel, nonparametric learning, bilinear sufficient space). These ideas apply equally to higher-order tensor-valued inputs, and the (theoretical) benefit to present in tensors seems marginal. On the other hand, presenting the work using tensors is probably more eye-catching? A subtlety in tensor learning is that we need to be careful in different notions of low-rankness, such as CP vs. Tucker. It is worthwhile comparing our method with CNN, which also allows tensor-based features.

[**Tuning parameter selection**] The current algorithm has at least two tuning parameters (r, α), where r is the structural dimension and α is the regularity parameter in the classifier. Tuning parameters in a computationally efficient way is a practice issue to address.