Sparse Subspace Clustering for the Popularity Adjusted Block Model

John Koo, Minh Tang, Michael Trosset

Abstract

TODO

1 Introduction

1.1 Notation

P denotes the edge probability matrix for the PABM. $A_{ij} \overset{\text{indep}}{\sim} Bernoulli(P_{ij})$ for $i > j$, and $A_{ji} = A_{ij}, A_{ii} = 0 \; \forall i, j \leq n$ to make A the edge weight matrix for a hollow, unweighted, and undirected graph. X is an ASE of A while Y is constructed using the popularity vectors $\{\lambda^{(kl)}\}_K$ and the projection matrix Π as an ASE of P. $Z = XQ - Y$ for $Q = \arg\min_{Q \in \mathbb{O}(p,q)} ||XQ - Y||_F$. Let $x_i^\top, y_i^\top, z_i^\top$ be the rows of X, Y, Z. $X^{(n)}$ represents the full X matrix for a sample of size n. $X^{(n,k)}$ represents the k^{th} block of $X^{(n)}$. Similarly, $P^{(k,l)}$ is the kl^{th} block of P, $P^{(n)}$ specifies that P is $n \times n$, and $P^{(n,k,l)}$ is the kl^{th} block of $P^{(n)}$.

2 Main Results

Theorem 1. The subspace detection property holds for Y with probability at least $1 - \sum_k^K n_k e^{-\sqrt{K(n_k-1)}}$.

This falls out of Theorem 2.8 from Soltanolkotabi and Candés [2]. The subspaces in Y are orthogonal, so $\text{aff}(S_k, S_l) = 0 \; \forall k, l \leq K$.

Property 2. By Rubin-Delanchy et al. [1], $\max_i ||Q_n x_i^{(n)} - y_i^{(n)}|| = \max_i ||z_i^{(n)}|| = \delta^{(n)} = O_P\Big(\frac{(\log n)^c}{n^{1/2}}\Big)$. Then $||Z^{(n)}||_F \to 0$, $\delta^{(n)} \to 0$, and $r(X^{(n,l)} Q^{(n,l)}) \to r(Y^{(n,l)})$. Here we assume $r(Y^{(n,l)}) > 0 \; \forall n > K + 1$ and $l \leq K$.

Theorem 3. ***TODO*** Let $r_k^{(n)} = r(U^{(n,k)})$ and $\hat{r}_k^{(n)} = r(\hat{U}^{(n,k)})$. Then $|\hat{r}_k^{(n)} - r_k^{(n)}| = O_P(a_n)$. ($a_n \to 0$.)

Alternatively, suppose $r_k^{(n)} > \alpha$ for some $\alpha > 0$. Then $P(\hat{r}_k^{(n)} > \alpha - a_n) > 1 - \epsilon(n)$.

Property 4. $P(\mu(Y^{(n,k)}) = 0) = 1$ [2].

This also holds for $\mu(U^{(n,k)})$ where U is the matrix of eigenvectors of P.

Theorem 5. Let $P^{(n)} = U^{(n)}\Lambda^{(n)}(U^{(n)})^\top$ be the spectral decomposition of P. Let $A^{(n)} = \hat{U}^{(n)}\hat{\Lambda}^{(n)}(\hat{U}^{(n)})^\top$ be the approximate spectral decomposition of $A^{(n)}$ where $\hat{U}^{(n)} \in \mathbb{R}^{n\times K^2}$. Then for some $a, c > 0$, $P(\mu(\hat{U}^{(n,l)}) \leq 4(\log n_l(n_k+1) + \log K + t)\frac{1}{K^2} a \frac{(\log n)^c}{n\sqrt{\rho_n}}) \geq 1 - \frac{1}{K^2}\sum_{k\neq l}\frac{4e^{-2t}}{(n_k+1)n_l}$.

Equivalently, we can say $\mu(\hat{U}^{(n,l)}) = O_P(\frac{\log n_l(n_k+1)+\log K+t}{K^2}\frac{(\log n)^c}{n\sqrt{\rho_n}})$
$= O_P(\frac{(\log n)^{c'} t}{nK^2\sqrt{\rho_n}})$ if $n, t \gg K$.

Theorem 6. ***TODO*** $P(\hat{\mu}_k^{(n)} > \hat{r}_k^{(n)}) =???$

Corollary. If $n \geq M$, $\exists\lambda > 0$ such that the LASSO subspace detection property holds for $X^{(n)}$ with parameter λ.

This falls out of Theorem 6 of Wang and Xu [3] and Theorem 6 of this paper.

References

[1] Patrick Rubin-Delanchy, Joshua Cape, Minh Tang, and Carey E. Priebe. A statistical interpretation of spectral embedding: the generalised random dot product graph, 2017.

[2] Mahdi Soltanolkotabi and Emmanuel J. Candés. A geometric analysis of subspace clustering with outliers. *Ann. Statist.*, 40(4):2195–2238, 08 2012. doi: 10.1214/12-AOS1034. URL https://doi.org/10.1214/12-AOS1034.

[3] Yu-Xiang Wang and Huan Xu. Noisy sparse subspace clustering. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 89–97, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR. URL http://proceedings.mlr.press/v28/wang13.html.