Spring 2016

ECE598YW: Information-theoretic methods in high-dimensional statistics

Final reading project

1 Rule

As the final project, please choose one of the following topics and write an essay summarizing the given paper(s) and give a 20-minute presentation in class (May 9th, Monday, time and location TBA). In preparing the essay, please note the following:

- Clearly summarize the statistical/mathematical model, the technical assumptions and the main results (e.g., upper bounds, lower bounds, etc).
- Read the proof, sketch the main argument and present the key ideas.
- Discus in your opinion the most innovative/challenging part of the paper and why it is cool (if you think it is obvious, it is perfectly fine too as long as you can make a convincing case).
- It is also valuable to discuss the part of the paper which you find difficult to understand, or are suspicious about, or can be improved.

2 Reading list

1. Confidence interval in high-dimensional regression: [ZZ14] (see also the follow-up work [JM14, CG15])
2. Gaussian graphical models: [RSZZ15] – *Ravi Kiran*
3. MLE with convex parameter set in high dimensions: [Cha14]
4. Estimating sparse covariance matrices [CZ12]
5. Estimating bandable covariance matrices [CZZ10] – *Subhadeep*
6. Risk bounds for binary graphical models [SW12] – *Yanjun*
7. Detecting sparse mixtures via f-divergences [JW07]
8. Minimax risk bounds for high-dimensional linear regression [Ver12b]
9. Model selection and adaptivity [BBM99]
10. Metric entropy of ℓ_p-balls in $\mathbb{R}^d$ with respect to ℓ_q-norm [Sch84] – *Ashok*
11. Duality of metric entropy [AMS04] – *Matt*
12. Concentration of sample covariance matrix [Ver12a] – *Yingxiang*

13. Oracle inequalities for high-dimensional matrices [GL10] – *Siddhartha*
14. Density estimation and mutual information [HO97] – *Aolin*
15. Predictive density estimation [LB04]
16. Sharp risk bounds for learning Gaussian mixtures [HP15] (read the full version http://arxiv.org/abs/1404.4997)
17. Functional estimation in high-dimensional Gaussian location model [CL11]
18. Nonparametric functional estimation [LNS99] – *Jaeho*
19. Sparse PCA [JL09a] (see also the earlier paper [JL09b]) – *Pengkun*
20. Analysis of LASSO and Dantzig selector [BRT09] – *Pan*
21. Estimating spare inverse covariance matrix via convex optimization [CLL11] – *Harsh Gupta*
22. Community detection and submatrix localization [CX14] – *Joseph*
23. Information theory and aggregation [LB06] – *Daewon*
24. Adaptive and sharp minimax procedure for linear regression with random design [CS15]
25. General analysis of penalized M-estimator: [NRWY12] (proofs are in supplement https://projecteuclid.org/euclid.ss/1356098555#supplemental) – *Jiaqi*
26. Entropy estimation with sublinear sample complexity [Pan04] – *Yuheng*

References

[AMS04] Shiri Artstein, Vitali Milman, and Stanisław J Szarek. Duality of metric entropy. *Annals of mathematics*, pages 1313–1328, 2004.

[BBM99] Andrew Barron, Lucien Birgé, and Pascal Massart. Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, 113(3):301–413, 1999.

[BRT09] Peter J Bickel, Ya'acov Ritov, and Alexandre B Tsybakov. Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, pages 1705–1732, 2009.

[CG15] T Tony Cai and Zijian Guo. Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *arXiv preprint arXiv:1506.05539*, 2015.

[Cha14] Sourav Chatterjee. A new perspective on least squares under convex constraint. *The Annals of Statistics*, 42(6):2340–2381, 2014.

[CL11] T. T. Cai and M. G. Low. Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. *The Annals of Statistics*, 39(2):1012–1041, 2011.

[CLL11] Tony Cai, Weidong Liu, and Xi Luo. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.

[CS15] Emmanuel Candes and Weijie Su. SLOPE is adaptive to unknown sparsity and asymptotically minimax. *arXiv preprint arXiv:1503.08393*, 2015.

[CX14] Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *arXiv preprint arXiv:1402.1267*, 2014.

[CZ12] T. T. Cai and H. H. Zhou. Optimal rates of convergence for sparse covariance matrix estimation. *The Annals of Statistics*, 40(5):2389–2420, 2012.

[CZZ10] T. T. Cai, C.-H. Zhang, and H. H. Zhou. Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics*, 38(4):2118–2144, 2010.

[GL10] Stéphane Gaïffas and Guillaume Lecué. Sharp oracle inequalities for the prediction of a high-dimensional matrix. *arXiv preprint arXiv:1008.4886*, 2010.

[HO97] David Haussler and Manfred Opper. Mutual information, metric entropy and cumulative relative entropy risk. *The Annals of Statistics*, 25(6):2451–2492, 1997.

[HP15] Moritz Hardt and Eric Price. Tight bounds for learning a mixture of two gaussians. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 753–760. ACM, 2015. http://arxiv.org/abs/1404.4997.

[JL09a] Iain M. Johnstone and Arthur Yu Lu. On consistency and sparsity for principal components analysis in high dimensions [with comments]. *Journal of the American Statistical Association*, 104(486):682–703, 2009.

[JL09b] Iain M Johnstone and Arthur Yu Lu. Sparse principal components analysis. *arXiv preprint arXiv:0901.4392*, 2009.

[JM14] Adel Javanmard and Andrea Montanari. Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1):2869–2909, 2014.

[JW07] L. Jager and J. A. Wellner. Goodness-of-fit tests via phi-divergences. *The Annals of Statistics*, 35(5):2018–2053, 2007.

[LB04] Feng Liang and Andrew Barron. Exact minimax strategies for predictive density estimation, data compression, and model selection. *IEEE Trans. Inf. Theory*, 50(11):2708–2726, 2004.

[LB06] Gilbert Leung and Andrew R Barron. Information theory and mixing least-squares regressions. *Information Theory, IEEE Transactions on*, 52(8):3396–3410, 2006.

[LNS99] Oleg Lepski, Arkady Nemirovski, and Vladimir Spokoiny. On estimation of the L_r norm of a regression function. *Probability Theory and Related Fields*, 113(2):221–253, 1999.

[NRWY12] Sahand N. Negahban, Pradeep Ravikumar, Martin J. Wainwright, and Bin Yu. A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Statistical Science*, 27(4):538–557, 11 2012.

[Pan04] Liam Paninski. Estimating entropy on m bins given fewer than m samples. *IEEE Transactions on Information Theory*, 50(9):2200–2203, 2004.

[RSZZ15] Zhao Ren, Tingni Sun, Cun-Hui Zhang, and Harrison H Zhou. Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *The Annals of Statistics*, 43(3):991–1026, 2015.

[Sch84] Carsten Schütt. Entropy numbers of diagonal operators between symmetric banach spaces. *Journal of approximation theory*, 40(2):121–128, 1984.

[SW12] N. Santhanam and M.J. Wainwright. Information-theoretic limits of selecting binary graphical models in high dimensions. *IEEE Trans. Inf. Theory*, 58(7):4117–4134, July 2012.

[Ver12a] Roman Vershynin. How close is the sample covariance matrix to the actual covariance matrix? *Journal of Theoretical Probability*, 25(3):655–686, 2012.

[Ver12b] Nicolas Verzelen. Minimax risks for sparse regressions: Ultra-high dimensional phenomenons. *Electronic Journal of Statistics*, 6:38–90, 2012.

[ZZ14] Cun-Hui Zhang and Stephanie S Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242, 2014.