Reading list and project guidelines for EE378C

1 Project guideline

In the final project, please choose one of the following papers (or several papers in the same topic, or propose your own upon the advisor's approval), write an essay summarizing the given paper(s), and give a 20-minute presentation in class. In preparing the essay, please note the following:

- Clearly summarize the statistical/mathematical model, the technical assumptions and the main results. If the paper contains multiple key results, feel free to present the part you feel most interesting;
- Read the proof, sketch the main argument, and present the key ideas;
- If (part of) the paper is about lower bounds, discuss their technique with the tools you learned in class. If the technique was covered in class, comment on how it is applied in the paper; if not, discuss whether the tools covered in the class could provide an alternative proof; if your answer is no, summarize the innovative points in their arguments;
- Discuss 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 encouraged to propose your own research project related to the course. In this case, in addition to the essay and presentation, you also need to submit a proposal at the end of the third week. Feel free to reach out to instructors any time for feedbacks and topic suggestions.

2 Reading list

2.1 Communication complexity

1. Public vs private coins:

   Ilan Newman, and Mario Szegedy. "Public vs. private coin flips in one round communication games." In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pp. 561-570, 1996.

2. Coupling arguments for communication complexity:

   Alexander Razborov. "On the distributed complexity of disjointness." *Theoretical Computer Science*, 106: 385-390, 1992.

3. Statistical lower bounds via communication complexity:

   Eric Blais, Clément L. Canonne, and Tom Gur. "Distribution testing lower bounds via reductions from communication complexity." *ACM Transactions on Computation Theory* 11, no. 2 (2019): 1-37.

2.2 Asymptotic equivalence and non-equivalence

1. Limits of experiments:

   Le Cam, L. "Limits of experiments." *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability.* Vol. 1. University of California Press Berkeley-Los Angeles, 1972.

2. Asymptotic equivalence results:

   Lawrence D. Brown, Andrew V. Carter, Mark G. Low, and Cun-Hui Zhang. "Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift." *The Annals of Statistics*, vol. 32, no. 5 (2004): 2074 - 2097.

3. Asymptotic non-equivalence results for small densities:

   Kolyan Ray and Johannes Schmidt-Hieber. "Asymptotic nonequivalence of density estimation and Gaussian white noise for small densities." *Ann. Inst. H. Poincar Probab. Statist.* 55 (4) 2195 - 2208, November 2019.

2.3 Statistical/computational tradeoff

1. The seminal paper on sparse PCA:

   Quentin Berthet, and Philippe Rigollet. "Complexity theoretic lower bounds for sparse principal component detection." In *Conference on Learning Theory*, 2013.

2. SOS lower bound for planted clique:

   Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K. Kothari, Ankur Moitra, and Aaron Potechin. "A nearly tight sum-of-squares lower bound for the planted clique problem." *SIAM Journal on Computing* 48, no. 2 (2019): 687-735.

3. A statistical/computational tradeoff independent of planted clique:

   Yuchen Zhang, Martin J. Wainwright, and Michael I. Jordan. "Optimal prediction for sparse linear models? Lower bounds for coordinate-separable M-estimators." *Electronic Journal of Statistics* 11 (2017): 752-799.

4. Generalization of planted clique - secret leakage:

   Matthew Brennan, and Guy Bresler. "Reducibility and statistical-computational gaps from secret leakage." *Conference on Learning Theory*, 2020.

2.4 f-divergence and joint range

1. Joint range of f-divergences:

   Peter Harremoës and Igor Vajda. "On pairs of f-divergences and their joint range". *IEEE Transactions on Information Theory* 57.6 (2011): 3230-3235.

2. Sharp inequalities for f-divergences:

   Adityanand Guntuboyina, Sujayam Saha, and Geoffrey Schiebinger. "Sharp inequalities for f-divergences." *IEEE transactions on information theory* 60, no. 1 (2013): 104-121.

2.5 Two-point methods

1. Local minimax rate of convergence:

   David L. Donoho and Richard C. Liu. "Geometrizing rates of convergence, II." *The Annals of Statistics*, 19: 668-701, 1991.

2. Posted-price auction:

   Robert Kleinberg, and Tom Leighton. "The value of knowing a demand curve: Bounds on regret for online posted-price auctions." In *44th Annual IEEE Symposium on Foundations of Computer Science*, 2003.

3. Bandits with bounded regret:

   Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet. "Bounded regret in stochastic multi-armed bandits." *Conference on Learning Theory*, 2013.

4. Duality arguments in combinatorial bandits:

   Xi Chen, Yanjun Han, and Yining Wang. "Adversarial Combinatorial Bandits with General Non-linear Reward Functions." *arXiv preprint* arXiv:2101.01301 (2021).

2.6 Testing simple against composite hypotheses

1. Sparse covariance estimation:

   T. Tony Cai, and Harrison H. Zhou. "Optimal rates of convergence for sparse covariance matrix estimation." *The Annals of Statistics* 40, no. 5 (2012): 2389-2420.

2. Instance-optimal identity testing:

   Gregory Valiant, and Paul Valiant. "An automatic inequality prover and instance optimal identity testing." *SIAM Journal on Computing* 46, no. 1 (2017): 429-455.

3. Local goodness-of-fit tests:

   S. Balakrishnan, and L. Wasserman. "Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates." *Annals of Statistics* 47, no. 4 (2019): 1893-1927.

2.7 Testing two composite hypotheses

1. Dualizing Le Cam's method:

   Yury Polyanskiy, and Yihong Wu. "Dualizing Le Cam's method, with applications to estimating the unseens." *arXiv preprint* arXiv:1902.05616 (2019).

2. Nonparametric entropy estimation:

   Yanjun Han, Jiantao Jiao, Tsachy Weissman, and Yihong Wu. "Optimal rates of entropy estimation over lipschitz balls." *Annals of Statistics*, 48(6): 3228-3250, 2020.

3. Multi-reference alignment:

   Afonso S. Bandeira, Philippe Rigollet, and Jonathan Weed. "Optimal rates of estimation for multi-reference alignment." *arXiv preprint* arXiv:1702.08546 (2017).

2.8 Variants of Fano's inequality

1. Distance-based Fano's inequality:

   John C. Duchi, and Martin J. Wainwright. "Distance-based and continuum Fano inequalities with applications to statistical estimation." *arXiv preprint* arXiv:1311.2669 (2013).

2. Fano's inequality based on f-informativity:

   Xi Chen, Adityanand Guntuboyina, and Yuchen Zhang. "On Bayes risk lower bounds." *The Journal of Machine Learning Research* 17, no. 1 (2016): 7687-7744.

2.9 Aggregation and model selection

1. Theory of model selection:

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

2. Theory of aggregation:

   Alexandre B. Tsybakov, "Optimal rates of aggregation." *Learning theory and kernel machines.* Springer, Berlin, Heidelberg, 2003. 303313.

3. Optimal competitive factor in proper and improper learning:

   Olivier Bousquet, Daniel Kane, and Shay Moran. "The optimal approximation factor in density estimation." In *Conference on Learning Theory*, 2019.

2.10 Covering and packing bounds

1. Metric entropy of classes of convex sets:

   Richard M. Dudley, Hiroshi Kunita, and Franois Ledrappier. *Ecole d'Ete de Probabilites de Saint-Flour XII*, 1982. (Section 7.3)

2. Metric entropy of ℓ_p-balls in $\mathbb{R}^d$ with respect to ℓ_q-norm:

   Carsten Schtt. "Entropy numbers of diagonal operators between symmetric Banach spaces." *Journal of approximation theory* 40, no. 2 (1984): 121-128.

3. Duality of metric entropy:

   Shiri Artstein, Vitali Milman, and Stanisaw J. Szarek. "Duality of metric entropy." *Annals of mathematics* (2004): 1313-1328.

2.11 Statistical bounds based on global Fano

1. Hellinger covering:

   Lucien Birgé. "Approximation dans les espaces métriques et théorie de l'estimation." *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 65, no. 2 (1983): 181-237. (In French; could also refer to Chapter 18 of http://www.stat.yale.edu/~yw562/teaching/it-stats.pdf)

2. KL covering:

   Yuhong Yang, and Andrew Barron. "Information-theoretic determination of minimax rates of convergence." *Annals of Statistics* (1999): 1564-1599.

3. Estimation and prediction in sparse linear regression:

   Garvesh Raskutti, Martin J. Wainwright, and Bin Yu. "Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls." *IEEE transactions on information theory* 57, no. 10 (2011): 6976-6994.

4. Log-concave density estimation beyond Donsker regime:

   Gil Kur, Yuval Dagan, and Alexander Rakhlin. "Optimality of maximum likelihood for log-concave density estimation and bounded convex regression." *arXiv preprint* arXiv:1903.05315 (2019).

2.12 Other geometric arguments

1. Redundancy and superefficiency:

   Andrew Barron, and Nicolas Hengartner. "Information theory and superefficiency." *Annals of statistics* 26, no. 5 (1998): 1800-1825.

2. Sample amplification:

   Brian Axelrod, Shivam Garg, Vatsal Sharan, and Gregory Valiant. "Sample amplification: Increasing dataset size even when learning is impossible." In *International Conference on Machine Learning*, pp. 442-451, 2020.

2.13 Compression-based arguments

1. Linear convergence with first-order oracle in convex optimization:

   Nemirovskij, A. S., and Yudin, D. B. (1983). "Problem complexity and method efficiency in optimization". (Only need to read Chapter 4; electronic copy of the full book is available at https://www2.isye.gatech.edu/~nemirovs/Nemirovskii_Yudin_1983.pdf)

2. Complexity of finding stationary points with p-th order oracle:

   Carmon, Yair, John C. Duchi, Oliver Hinder, and Aaron Sidford. "Lower bounds for finding stationary points I." *Mathematical Programming* (2019): 1-50.

3. Compression-based upper bounds:

   Ashtiani, Hassan, Shai Ben-David, Nicholas JA Harvey, Christopher Liaw, Abbas Mehrabian, and Yaniv Plan. "Nearly tight sample complexity bounds for learning mixtures of Gaussians via sample compression schemes." In NeurIPS 2018.

4. Lower bound of Johnson–Linderstrass:

   Larsen, Kasper Green, and Jelani Nelson. "Optimality of the Johnson-Lindenstrauss lemma." *IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2017.

2.14 Sequential experiments

1. Stochastic optimization with zeroth-order oracle:

   Ohad Shamir. "On the complexity of bandit and derivative-free stochastic convex optimization." In *Conference on Learning Theory*, pp. 3-24. PMLR, 2013.

2. Stochastic batch optimization:

   John Duchi, Feng Ruan, and Chulhee Yun. "Minimax bounds on stochastic batched convex optimization." In *Conference On Learning Theory*, pp. 3065-3162. PMLR, 2018.

3. Active learning:

   R.M. Castro, and R.D. Nowak. "Minimax bounds for active learning." *IEEE Transactions on Information Theory*, 54(5), pp. 2339-2353, 2008.

4. Nonparametric bandit:

   Andrea Locatelli, and Alexandra Carpentier. "Adaptivity to smoothness in $\mathcal{X}$-armed bandits." In *Conference on Learning Theory*, pp. 1463-1492. PMLR, 2018.

2.15 Ability of adaptation

1. A general constrained risk inequality:

   John C. Duchi, and Feng Ruan. "A constrained risk inequality for general losses." *arXiv preprint* arXiv:1804.08116 (2018).

2. Lepski's adaptation trick:

   Oleg V. Lepskii. "Asymptotically minimax adaptive estimation. i: Upper bounds. optimally adaptive estimates." *Theory of Probability and Its Applications*, 36(4):682-697, 1992.

3. Inability of adaptive tests:

   Vladimir Spokoiny. "Adaptive and spatially adaptive testing of a nonparametric hypothesis." 1996.

4. Adaptation to loss functions:

   Yanjun Han. "On the high accuracy limitation of adaptive property estimation." *AISTATS*, 2021.

2.16 Communication/privacy constrained estimation

1. Strong data-processing inequality:

   Mark Braverman, Ankit Garg, Tengyu Ma, Huy L. Nguyen, and David P. Woodruff. "Communication lower bounds for statistical estimation problems via a distributed data processing inequality." In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pp. 1011-1020. 2016.

2. Communication complexity:

   John Duchi, and Ryan Rogers. "Lower bounds for locally private estimation via communication complexity." *Conference on Learning Theory*, 2019.

3. Quantized Fisher information:

   Leighton Pate Barnes, Yanjun Han, and Ayfer Ozgur. "Lower bounds for learning distributions under communication constraints via Fisher information." *Journal of Machine Learning Research* 21.236 (2020): 1-30.

2.17 Memory constrained estimation

1. Memory-constrained parity learning:

   Ran Raz. "Fast learning requires good memory: A time-space lower bound for parity learning." *Journal of the ACM (JACM)* 66.1 (2018): 1-18.

   Or a clearer follow-up paper:

   Sumegha Garg, Ran Raz, and Avishay Tal. "Extractor-based time-space lower bounds for learning." *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing.* 2018.

2. Memory-constrained learning with small noise:

   Vatsal Sharan, Aaron Sidford, and Gregory Valiant. "Memory-sample tradeoffs for linear regression with small error." *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing.* 2019.

3. Memory-constrained uniformity testing:

   Ilias Diakonikolas, Themis Gouleakis, Daniel M. Kane, and Sankeerth Rao. "Communication and memory efficient testing of discrete distributions." In *Conference on Learning Theory*, 2019.

4. Memory constraints in belief propagation:

   Vishesh Jain, Frederic Koehler, Jingbo Liu, and Elchanan Mossel. "Accuracy-memory tradeoffs and phase transitions in belief propagation." In *Conference on Learning Theory*, pp. 1756-1771. PMLR, 2019.

2.18 Max-min vs min-max formulation

1. Sequential Radamacher complexity:

   Alexander Rakhlin, and Karthik Sridharan. "Online nonparametric regression with general loss functions." *arXiv preprint* arXiv:1501.06598 (2015).

2. Batched bandit with data-driven batch sizes:

   Zijun Gao, Yanjun Han, Zhimei Ren, and Zhengqing Zhou. "Batched multi-armed bandits problem." *NeurIPS*, 2019.

3. Semi-min-max framework:

   Jayadev Acharya, Clément L. Canonne, Yanjun Han, Ziteng Sun, and Himanshu Tyagi. "Domain compression and its application to randomness-optimal distributed goodness-of-fit." In *Conference on Learning Theory*, pp. 3-40. PMLR, 2020.

2.19 Network information theory

1. Reverse hypercontractivity:

   Jingbo Liu and Ayfer Özgür. "Capacity upper bounds for the relay channel via reverse hypercontractivity." *IEEE Transactions on Information Theory* 66.9 (2020): 5448-5455.

2. Capacity of relay channel:

   Yikun Bai, Xiugang Wu, and Ayfer Ozgur. "Information Constrained Optimal Transport: From Talagrand, to Marton, to Cover." *arXiv preprint* arXiv:2008.10249 (2020).

3. Strengthened cutset bound:

   Abbas El Gamal, Amin Gohari, and Chandra Nair. "Strengthened Cutset Upper Bounds on the Capacity of the Relay Channel and Applications." *arXiv preprint* arXiv:2101.11139 (2021).