CS7545: Machine Learning Theory **Fall 2021**

Project Ideas

Lecturer: Santosh Vempala

Selected Papers by Topic:

- Mixture Models
- Topic Models
- Clustering
- Dimensionality Reduction
- ICA
- PAC learning
- Online learning
- Statistical query learning
- Brain learning
- Deep learning
- Robust learning

1 Books

- Foundations of Data Science (A. Blum, J. Hopcroft and R. Kannan)
- An Introduction to Computational Learning Theory (M. Kearns and U. Vazirani)
- Understanding Machine Learning: From Theory to Algorithms (S. Shalev-Shwartz and S. Ben-David)
- Spectral Algorithms (R. Kannan and S. Vempala)
- Introduction to Online Convex Optimization (E. Hazan)

2 Surveys (talks and articles)

- Online Learning (by N. Cesa-Bianchi)
- The Complexity of Unsupervised Learning (by S. Vempala)
- N. Cesa-Bianchi and G. Lugosi. Prediction, learning, and games. Cambridge University Press, 2006.
- A. Blum and Y. Mansour. Learning, Regret Minimization, and Equilibria. Book chapter in Algorithmic Game Theory, 2007.

3 Selected Papers by Topic

3.1 Mixture Models

- D. Achlioptas and F. McSherry. On spectral learning of mixtures of distributions. In Proc. of COLT, 2005.
- A. Anandkumar, D. Foster, D. Hsu, S. M. Kakade, and Y.K. Liu. A spectral algorithm for latent dirichlet allocation. In Advances in Neural Information Processing Systems 25, pp. 926-934, 2012.
- S. Arora and R. Kannan. Learning mixtures of arbitrary gaussians. Annals of Applied Probability, 15(1A):69-92, 2005.
- M. Belkin and K. Sinha. Polynomial learning of distribution families. In FOCS, 2010.
- S. C. Brubaker. Robust pca and clustering on noisy mixtures. In Proc. of SODA, 2009.
- S. C. Brubaker and S. Vempala. Isotropic pca and affine-invariant clustering. In M. Gr otschel and G. Katona, editors, Building Bridges Between Mathematics and Computer Science, volume 19 of Bolyai Society Mathematical Studies, 2008.

- K. Chaudhuri and S. Rao. Beyond gaussians: Spectral methods for learning mixtures of heavy-tailed distributions. In Proc. of COLT, 2008.
- S. DasGupta. Learning mixtures of gaussians. In Proc. of FOCS, 1999.
- A. Dasgupta, J. Hopcroft, J. Kleinberg, and M. Sandler. On learning mixtures of heavy-tailed distributions. In Proc. of FOCS, 2005.
- S. DasGupta and L. Schulman. A two-round variant of em for gaussian mixtures. In Proc. of UAI, 2000.
- J. Feldman and R. O'Donnell. Learning mixtures of product distributions over discrete domains. SIAM Journal on Computing, 37(5):1536-1564, 2008.
- Y. Freund and Y. Mansour. Estimating a mixture of two product distributions. In Proc. of COLT, pages 53-62, 1999.
- M. Hardt and E. Price Sharp bounds for learning a mixture of two gaussians CoRR abs/1404.4997 (2014).
- D. Hsu and S. M. Kakade. Learning mixtures of spherical gaussians: moment methods and spectral decompositions. In Proc of ITCS, pp, 11-20, 2013.
- R. Kannan, H. Salmasian, and S. Vempala. The spectral method for general mixture models. SIAM Journal on Computing, 38(3):1141-1156, 2008.
- A. T. Kalai, A. Moitra, and G. Valiant. Efficiently learning mixtures of two Gaussians. In Proc of STOC, pp. 553-562, 2010.
- A. Moitra, and G. Valiant. Settling the Polynomial Learnability of Mixtures of Gaussians. In Proc of FOCS, pp. 93-102, 2010.
- S. Vempala and G. Wang. A spectral algorithm for learning mixtures of distributions. Journal of Computer and System Sciences, 68(4):841-860, 2004.
- C. Daskalakis and G. Kamath, Faster and sample near-optimal algorithms for proper learning mixtures of Gaussians, Proc. 27th Annual Conference on Learning Theory (COLT), pp. 1183-1213, 2014.
- A. Bakshi, I. Diakonikolas, H. Jia, D. M. Kane, P. K. Kothari, and S. S. Vempala. Robustly learning mixtures of k arbitrary gaussians. arXiv preprint arXiv:2012.02119, 2020.
- A. Liu and A. Moitra. Settling the robust learnability of mixtures of gaussians. Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, 2021.

3.2 Topic models

- A. Anandkumar, D. Foster, D. Hsu, S. Kakade and Y. Liu. A Spectral Algorithm for Latent Dirichlet Allocation. NIPS 2012.
- S. Arora, R. Ge, and A. Moitra. Learning Topic Models – Going Beyond SVD. In FOCS, pp. 1-10, 2012.
- T. Bansal, C. Bhattacharyya, and R. Kannan. A provable SVD-based algorithm for learning topics in dominant admixture corpus. arXiv preprint arXiv:1410.6991 (2014).
- D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. The Journal of machine Learning research, 3, pp. 993-1022, 2003.
- C. Papadimitriou, P. Raghavan, H. Tamaki, and S. Vempala. Latent semantic indexing: A probabilistic analysis. In Proc. of PODS, pp. 159-168, 1998.

3.3 Clustering

- M. Ackerman and S. Ben-David. Measures of Clustering Quality: A working set of axioms for clustering. NIPS 2008.
- D. Arthur and S. Vassilvitskii. k-means++: The advantages of careful seeding. In Proc. of SODA, 2007.
- O. Awasthi and O. Sheffet. Improved spectral-norm bounds for clustering. In Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, pp. 37-49, 2012.
- A. Blum, M.F. Balcan and S. Vempala. A Discriminative Framework for Clustering via Similarity Functions. STOC 2008.
- A. Blum, M.F. Balcan and A. Gupta. Clustering Under Approximation Stability, JACM,Volume 60,

Issue 2, 2013.
- M. Charikar, S. Guha, É. Tardos, and D. B. Shmoys.A constant-factor approximation algorithm for the k-median problem. In Proc. of the 31st Annual ACM Symposium on Theory of Computing, pages 1-10, 1999.
- D. Cheng, R. Kannan, S. Vempala, and G. Wang. A divide-and-merge methodology for clustering. ACM Trans. Database Syst., 31(4):1499-1525, 2006.
- M. B. Cohen, S. Elder, C. Musco, C. Musco, and M. Persu. Dimensionality reduction for k-means clustering and low rank approximation. arXiv preprint arXiv:1410.6801 (2014).
- A. Dasgupta, J. Hopcroft, R. Kannan, and P. Mitra. Spectral clustering with limited independence. In Proc. of SODA, pages 1036-1045, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.
- W. Fernandez de la Vega, M. Karpinski, C. Kenyon, and Y. Rabani. Approximation schemes for clustering problems. In Proc. 35th ACM STOC, pages 50-58, 2003.
- S. Har-Peled and K. R. Varadarajan. Projective clustering in high dimensions using core-sets. In Symposium on Computational Geometry, pages 312-318, 2002.
- J. A. Hartigan and M. A. Wong. A k-means clustering algorithm. Applied Statistics, 28(1):100-108, 1979.
- P. Indyk. A sublinear time approximation scheme for clustering in metric spaces. In Proc. 40th IEEE FOCS, pages 154-159, 1999.
- R. Kannan, S. Vempala, and A. Vetta. On clusterings: Good, bad and spectral. J. ACM, 51(3):497-515, 2004.
- T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu. A Local Search Approximation Algorithm for k-Means Clustering. In Proceedings of the eighteenth annual symposium on Computational geometry, pp. 10-18. ACM, 2002.
- J. Kleinberg. An impossibility theorem for clustering. NIPS 2002.
- M. B. Cohen, S. Elder, C. Musco, C. Musco, and M. Persu. Dimensionality reduction for k-means clustering and low rank approximation. In Proceedings of the forty-seventh annual ACM symposium on Theory of computing, pp. 163-172, 2015.

3.4 Dimensionality Reduction

- N. Ailon and B. Chazelle. The fast Johnson-Lindenstrauss transform and approximate nearest neighbors. SIAM Journal on Computing, 39(1), 302-322, 2009
- R. I. Arriaga and S. Vempala. An algorithmic theory of learning: Robust concepts and random projection. In Foundations of Computer Science, 1999. 40th Annual Symposium on, pp. 616-623, 1999.
- S. Dasgupta and A. Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. Random Structures & Algorithms 22, no. 1, pp 60-65, 2003.
- W. Johnson, and J. Lindenstrauss. Extensions of Lipschitz mapping into a Hilbert space. Conference in modern analysis and probability, pp. 189-206, 1984.
- P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In Proc. 30th ACM STOC, pp. 604-613, 1998.
- S. Vempala. The Random Projection Method. AMS, 2004.
- J. Bourgain, S. Dirksen, and J. Nelson. Toward a unified theory of sparse dimensionality reduction in euclidean space. Geometric and Functional Analysis, 25(4):1009-1088, 2015.
- K. G. Larsen and J. Nelson. The Johnson-Lindenstrauss lemma is optimal for linear dimensionality reduction. arXiv preprint arXiv:1411.2404, 2014.
- S. Narayanan and J. Nelson. Optimal terminal dimensionality reduction in euclidean space. In Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, pages 1064-1069, 2019.
- K. Clarkson, R. Wang, and D. Woodruff. Dimensionality reduction for tukey regression. In International Conference on Machine Learning, pages 1262-1271. PMLR, 2019.
- A. Chowdhury, P. Drineas, D. P. Woodruff, and S. Zhou. Approximation algorithms for sparse principal component analysis. arXiv preprint arXiv:2006.12748, 2020.

- Y. Li, D. P. Woodruff, and T. Yasuda. Exponentially improved dimensionality reduction for ℓ_1: Subspace embeddings and independence testing. arXiv preprint arXiv:2104.12946, 2021.

3.5 ICA

- A. Anandkumar, D. Foster, D. Hsu, S. M. Kakade, and Y.-K. Liu. A spectral algorithm for latent dirichlet allocation. In Advances in Neural Information Processing Systems 25, pp. 926-934, 2012.
- A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky. Tensor decompositions for learning latent variable models. CoRR abs/1210.7559 (2012).
- S. Arora, R. Ge, A. Moitra, and S. Sachdeva. Provable ICA with unknown gaussian noise, with implications for gaussian mixtures and autoencoders. In Proc. of NIPS, pp. 2384-2392, 2012.
- M. Belkin, L. Rademacher, and J. Voss. Blind signal separation in the presence of Gaussian noise. In Proc. of COLT, 2013.
- J. Cardoso. Source separation using higher order moments. In International Conference on Acoustics, Speech, and Signal Processing, 1989.
- P. Comon, and C. Jutten. Eds. Handbook of Blind Source Separation. Academic Press, 2010.
- A. M. Frieze, M. Jerrum, and R. Kannan. Learning linear transformations. In Proc. of FOCS, pp. 359-368, 1996.
- N. Goyal, S. Vempala, and Y. Xiao. Fourier PCA and robust tensor decomposition. In Proc. of STOC, pp. 584-593, 2014
- A. Hyvarinen, J. Karhunen, and E. Oja. Independent Component Analysis. Wiley, 2001.
- P. Q. Nguyen, and O. Regev. Learning a parallelepiped: Cryptanalysis of GGH and NTRU signatures. J. Cryptology 22, 2, pp. 139-160, 2009.
- S. Vempala and Y. Xiao Max vs Min: Independent Component Analysis with nearly Linear Sample Complexity CoRR abs/1412.2954 (2014)
- Y. Xiao. Fourier pca package. GitHuB (2014). http://github.com/yingusxiaous/libFPCA
- A. Yeredor. Blind source separation via the second characteristic function. Signal Processing 80, 5, pp. 897-902, 2000
- B. Barak, J. A. Kelner, and D. Steurer, Dictionary learning and tensor decomposition via the sum-of-squares method, STOC'15 - Proceedings of the 2015 ACM Symposium on Theory of Computing, ACM, New York, pp. 143-151, 2015.
- P. K. Kothari and D. Steurer, Outlier-robust moment-estimation via sum-of-squares, CoRR abs/1711.11581 (2017).

3.6 PAC learning

- D. Angluin, and P. Laird. Learning from noisy examples. Machine Learning, 2(4), 343-370,1988.
- A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Occam's razor. Information processing letters, 24(6), 377-380, 1987.
- A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. Journal of the ACM (JACM) 36.4, pp. 929-965, 1989.
- M. Collins, R. E. Schapire, and Y. Singer. Logistic regression, AdaBoost and Bregman distances. Machine Learning, 48(1-3), 253-285, 2002
- C. Cortes and V. Vapnik. Support-vector networks. Machine learning, 20(3), 273-297, 1995.
- A. Ehrenfeucht, D. Haussler, M. Kearns, and L. Valiant. A general lower bound on the number of examples needed for learning. Information and Computation, 82(3), 247-261, 1989
- Y. Freund, and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, 55(1), 119-139, 1997.
- P. Gopalan, A. T. Kalai, A. R. Klivans. Agnostically Learning Decision Trees In the Proceedings of the 40th ACM Symposium on Theory of Computing (STOC), 2008.
- D. Haussler. Learning conjunctive concepts in structural domains. Machine learning 4, no. 1, pp. 7-40. 1989.
- D. Haussler. Probably approximately correct learning. University of California, Santa Cruz, Computer Research Laboratory, 1990.

- D. Haussler, M. Kearns, N. Littlestone, and M. K. Warmuth. Equivalence of models for polynomial learnability. Information and Computation, 95(2), pp 129-161, 1991.
- A. T. Kalai, A. R. Klivans, Y. Mansour, and R. A. Servedio. Agnostically Learning Halfspaces. In Proceedings of the Foundations of Computer Science (FOCS), 2006.
- D. M. Kane. The Average Sensitivity of an Intersection of Half Spaces. In Proceedings of STOC, pp. 437-440, 2014.
- A. T. Kalai, Y. Mansour, and E. Verbin. On agnostic boosting and parity learning. In Proceedings of the fortieth annual ACM symposium on Theory of computing, pp. 629-638, 2008.
- M. Kearns and M. Li. Learning in the presence of malicious errors. SIAM Journal on Computing, 22(4), 807-837, 1993
- M. J. Kearns and R. E. Schapire. Efficient distribution-free learning of probabilistic concepts. In Foundations of Computer Science, 1990. Proceedings., 31st Annual Symposium on, pp. 382-391, 1990.
- M. J. Kearns, R. E. Schapire, and L. M. Sellie. Toward efficient agnostic learning. Mach. Learn., 17(2-3), 1994.
- A. R. Klivans, P. M. Long, & R. A. Servedio. Learning halfspaces with malicious noise. The Journal of Machine Learning Research, 10, 2715-2740, 2009.
- A. R. Klivans and R. A. Servedio. Toward attribute efficient learning of decision lists and parities. The Journal of Machine Learning Research, 7, 587-602, 2006.
- A. R. Klivans and A. A. Sherstov. Cryptographic hardness for learning intersections of halfspaces. In Proc. of FOCS, pp 553-562, 2006.
- N. Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. Machine learning, 2(4), 285-318, 1988.
- L. Pitt and L. G. Valiant Computational limitations on learning from examples. Journal of the ACM (JACM), 35(4), pp. 965-984, 1988.
- J. R. Quinlan. Induction of decision trees. Machine learning, 1(1), 81-106, 1986.
- J. R. Quinlan and R. L. Rivest. Inferring decision trees using the minimum description length principle. Information and computation, 80(3), 227-248, 1989.
- F. Rosenblatt. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan, 1962.
- R. E. Schapire. The strength of weak learnability Machine learning, 5(2), 197-227, 1990.
- L.G. Valiant. Deductive learning. Philosophical Transactions of the Royal Society of London A, 312:441-446, 1984.
- L.G. Valiant A Theory of the Learnable Communications of the ACM 27, no. 11, pp. 1134-1142, 1984.
- L.G. Valiant. Learning disjunctions of conjunctions. In Proc. of 9th IJCAI, pp. 560-566, 1985.
- V. N. Vapnik and A. Y. Chervonenkis. On the uniform convergence of relative frequencies of events to thier probabilities. Theory of Probability and its Applications, 16(2):264-280, 1971.
- S. Vempala. A random-sampling-based algorithm for learning intersections of halfspaces. Journal of the ACM (JACM), 57(6), 32, 2010.
- S. Vempala. Learning convex concepts from Gaussian distributions with PCA. FOCS 2010.
- A. Daniely. Complexity theoretic limitations on learning halfspaces. In Proc. of 48th STOC, pp. 105-117, 2016.

3.7 Online learning

- J. Abernethy, E. Hazan, and A. Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In Proceedings of COLT, 2008.
- S. Arora, E. Hazan, and S. Kale. The Multiplicative Weights Update Method: a Meta-Algorithm and Applications. Theory of Computing, 8(1), 121-164, 2012.
- P. Auer. Using confidence bounds for exploitation-exploration trade-offs. Journal of Machine Learning Research, vol. 3, pp. 397-422, 2002.
- P. Auer, N. Cesa-Bianchi, and P. Fischer Finite-time analysis of the multiarmed bandit problem. Machine Learning Journal, vol. 47, no. 2-3, pp. 235-256, 2002.

- P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire. The non-stochastic multi-armed bandit problem. SIAM Journal on Computing, vol. 32, no. 1, pp. 48-77, 2002.
- N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, R. E. Schapire, and M. K. Warmuth. How to use expert advice. Journal of the ACM (JACM), 44(3), 427-485, 1997.
- A. Flaxman, A. Kalai, and B. McMahan Online convex optimization in the bandit setting: Gradient descent without a gradient In Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2005.
- Y. Freund and R. E. Schapire. Game theory, on-line prediction and boosting. In Proceedings of the ninth annual conference on Computational learning theory, pp. 325-332, 1996.
- C. Gentile. The robustness of the p-norm algorithms. Machine Learning, 53(3), 265-299, 2003.
- E. Hazan, A. Agarwal, and S. Kale. Logarithmic regret algorithms for online convex optimization. Machine Learning, 69(2-3), 169-192, 2007.
- A. Kalai and S. Vempala. Efficient algorithms for online decision problems. Journal of Computer and System Sciences, 71(3):291-307, 2005.
- N. Littlestone and M. K. Warmuth. The weighted majority algorithm. Information and computation, 108(2), 212-261, 1994.
- M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In Proc. of ICML, pp. 928-936, 2003.
- D. J. Foster, A. Rakhlin, and K. Sridharan. Adaptive online learning. arXiv preprint arXiv:1508.05170, 2015.
- D. J. Foster, A. Rakhlin, and K. Sridharan. Zigzag: A new approach to adaptive online learning. In Conference on Learning Theory, pages 876-924. PMLR, 2017.
- D. J. Foster, S. Kale, M. Mohri, and K. Sridharan. Parameter-free online learning via model selection. arXiv preprint arXiv:1801.00101, 2017.
- D. J. Foster, A. Rakhlin, and K. Sridharan. Online learning: Sufficient statistics and the burkholder method. In Conference On Learning Theory, pages 3028-3064. PMLR, 2018.
- D. Foster and M. Simchowitz. Logarithmic regret for adversarial online control. In International Conference on Machine Learning, pages 3211-3221. PMLR, 2020.
- M. Simchowitz and D. Foster. Naive exploration is optimal for online LQR. In International Conference on Machine Learning, pages 8937-8948. PMLR, 2020.

3.8 Statistical query learning

- J. A. Aslam and S. E. Decatur. General bounds on statistical query learning and PAC learning with noise via hypothesis boosting. Information and Computation, 141(2), 85-118, 1998.
- P. Awasthi, M. F. Balcan, and P. M. Long. The power of localization for efficiently learning linear separators with noise. In Proceedings of STOC, pp. 449-458, 2014.
- A. Blum, A. Frieze, R. Kannan, and S. Vempala. A polynomial-time algorithm for learning noisy linear threshold functions. Algorithmica, 22(1-2), 35-52, 1998.
- A. Blum, M. Furst, J. Jackson, M. Kearns, Y. Mansour, and S. Rudich. Weakly learning DNF and characterizing statistical query learning using Fourier analysis. In Proceedings of STOC, pages 253-262, 1994.
- A. Blum, A. Kalai, and H. Wasserman. Noise-tolerant learning, the parity problem, and the statistical query model. Journal of the ACM (JACM), 50(4), 506-519, 2003.
- V. Feldman, E. Grigorescu, L. Reyzin, S. Vempala, and Y. Xiao. Statistical algorithms and a lower bound for detecting planted cliques. In Proceedings of the forty-fifth annual ACM symposium on Theory of computing, pp. 655-664, 2013.
- V. Feldman and V. Kanade. Computational Bounds on Statistical Query Learning. In Proc. of COLT, pp. 16-1, 2012.
- V. Feldman, W. Perkins, and S. Vempala. On the complexity of random satisfiability problems with planted solutions. In Proc. of STOC, 2015.
- M. Kearns. Efficient noise-tolerant learning from statistical queries. Journal of the ACM, 45(6):983-1006, 1998.
- S. Vempala and J. Wilmes. Gradient descent for one-hidden-layer neural networks: Polynomial con-

vergence and sq lower bounds. In Conference on Learning Theory, pp. 3115-3117. PMLR, 2019.

3.9 Brain learning

- V. Feldman and L. G. Valiant. Experience-induced neural circuits that achieve high capacity. Neural computation, 21(10), 2715-2754, 2009
- C. H. Papadimitriou and S. Vempala. Unsupervised Learning through Prediction in a Model of Cortex. CoRR abs/1412.7955, 2014.
- L. G. Valiant. Circuits of the mind. Oxford University Press, 1995.
- L. G. Valiant. A neuroidal architecture for cognitive computation. J. ACM, 47(5):854-882, 2000.
- L. G. Valiant. Memorization and association on a realistic neural model. Neural Computation, 17(3):527-555, 2005
- L. G. Valiant. A quantitative theory of neural computation. Biological Cybernetics, 95(3):205-211, 2006.
- Leslie G. Valiant. The hippocampus as a stable memory allocator for cortex. Neural Computation, 24(11):2873-2899, 2012.
- C. H. Papadimitriou, S. S. Vempala, D. Mitropolsky, M. Collins, and W. Maass. Brain computation by assemblies of neurons. Proceedings of the National Academy of Sciences, 117(25), pp.14464-14472, 2020

3.10 Deep learning

- S. Arora, A. Bhaskara, R. Ge, and T. Ma. Provable bounds for learning some deep representations. arXiv preprint arXiv:1310.6343, 2013.
- Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle. Greedy layer-wise training of deep networks. In Proc. of NIPS, 2007.
- G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. Neural Comput., 18(7):1527-1554, 2006.
- C. Poultney, S. Chopra, and Y. Lecun. Efficient learning of sparse representations with an energy-based model. In Proc. of NIPS 2006, 2006.
- L. Song, S. Vempala, J. Wilmes, and B. Xie. On the complexity of learning neural networks. arXiv preprint arXiv:1707.04615, 2017.
- A. Jacot, F. Gabriel, C. Hongler. Neural tangent kernel: Convergence and generalization in neural networks. arXiv preprint arXiv:1806.07572, 2018.
- S. Vempala and J. Wilmes. Gradient descent for one-hidden-layer neural networks: Polynomial convergence and sq lower bounds. In Conference on Learning Theory, pp. 3115-3117. PMLR, 2019.

3.11 Robust learning

- I. Diakonikolas, G. Kamath, D. M. Kane, J. Li, A. Moitra, and A. Stewart, Robust estimators in high dimensions without the computational intractability, Proc. 57th IEEE Symposium on Foundations of Computer Science (FOCS), pp. 655-664, 2016.
- K. A. Lai, A. B. Rao, and S. Vempala, Agnostic estimation of mean and covariance, Proc. 57th IEEE Symposium on Foundations of Computer Science (FOCS), 2016, pp. 665-674.
- S. Balakrishnan, S. S. Du, J. Li, and A. Singh, Computationally efficient robust sparse estimation in high dimensions, Proc. 30th Annual Conference on Learning Theory, pp. 169-212, 2017.
- M. Charikar, J. Steinhardt, and G. Valiant, Learning from untrusted data, Proc. 49th Annual ACM Symposium on Theory of Computing, pp. 47-60, 2017.
- I. Diakonikolas, G. Kamath, D. M. Kane, J. Li, A. Moitra, and A. Stewart, Being robust (in high dimensions) can be practical, Proc. 34th International Conference on Machine Learning (ICML), pp. 999-1008, 2017.
- I. Diakonikolas, D. M. Kane, and A. Stewart, Statistical query lower bounds for robust estimation of high-dimensional Gaussians and Gaussian mixtures, Proc. 58th IEEE Symposium on Foundations of Computer Science (FOCS), pp. 73-84, 2017.

- P. K. Kothari and D. Steurer, Outlier-robust moment-estimation via sum-of-squares, CoRR abs/1711.11581 (2017).
- Y. Cheng, I. Diakonikolas, D. M. Kane, and A. Stewart, Robust learning of fixed-structure Bayesian networks, Proc. 33rd Annual Conference on Neural Information Processing Systems (NeurIPS), pp. 10304-10316, 2018.
- I. Diakonikolas, D. M. Kane, and A. Stewart, List-decodable robust mean estimation and learning mixtures of spherical Gaussians, Proc. 50th Annual ACM Symposium on Theory of Computing (STOC), pp. 1047-1060, 2018.
- S. B. Hopkins and J. Li, Mixture models, robustness, and sum of squares proofs, Proc. 50th Annual ACM Symposium on Theory of Computing (STOC), pp. 1021-1034, 2018.
- I. Diakonikolas, G. Kamath, D. M. Kane, J. Li, J. Steinhardt, and A. Stewart, Sever: A robust meta-algorithm for stochastic optimization, CoRR abs/1803.02815 (2018), Conference version in ICML 2019.
- I. Diakonikolas and D. M. Kane, Recent advances in algorithmic high-dimensional robust statistics, CoRR abs/1911.05911, 2019.
- I. Diakonikolas, S. Karmalkar, D. Kane, E. Price, and A. Stewart, Outlier-robust high-dimensional sparse estimation via iterative filtering, Advances in Neural Information Processing Systems 33, NeurIPS 2019, 2019.
- I. Diakonikolas, T. Gouleakis, C. Tzamos. Distribution-independent PAC learning of halfspaces with massart noise. arXiv preprint arXiv:1906.10075, 2019.
- A. Bakshi and P. Kothari, List-decodable subspace recovery via sum-of-squares, arXiv preprint arXiv:2002.05139, 2020.
- A. Bakshi and P. Kothari, Outlier-robust clustering of non-spherical mixtures, CoRR abs/2005.02970, 2020.
- I. Diakonikolas, S. B. Hopkins, D. Kane, and S. Karmalkar, Robustly learning any clusterable mixture of gaussians, CoRR abs/2005.06417, 2020.
- A. Bakshi and A. Prasad, Robust linear regression: Optimal rates in polynomial time, arXiv preprint arXiv:2007.01394, 2020.
- Y. Cheng, I. Diakonikolas, R. Ge, and M. Soltanolkotabi, High-dimensional robust mean estimation via gradient descent, CoRR abs/2005.01378, 2020.
- D. M. Kane, Robust learning of mixtures of gaussians, CoRR abs/2007.05912 (2020).
- A. Bakshi, I. Diakonikolas, H. Jia, D. M. Kane, P. K. Kothari, and S. S. Vempala. Robustly learning mixtures of k arbitrary gaussians. arXiv preprint arXiv:2012.02119, 2020.
- A. Liu and A. Moitra. Settling the robust learnability of mixtures of gaussians. Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, 2021.