EE375/STAT375: Project rules and ideas

Andrea Montanari

Rules

1. Students will form groups of two students each. (If the number of students enrolled is odd, one group of one or three students will be allowed).

2. They will choose a topic and a set of (at least 2) papers to read. The group composition and chosen topic need to be communicated to Andrea by April 23.

3. They will write a report on the topic, max 6 pages in NeurIPS format.

4. They will prepare a 15 minutes videos to present their report. Videos and reports will be posted on Canvas, and made available to other students.

5. We will devote the last 4 lectures of the class (May 25, 27, June 1, 3) to discuss these projects.

An important remark. If you list a paper among the ones you read, it will be assumed that you read all of it (including proofs and appendices).

Topics

This is a list of topics and references. Most topics are broad enough to be split across multiple groups.

Three remarks: (*i*) The reference lists given here are far from edhaustive. You should use them as pointers to explore the literature. (*ii*) Some of these topics have been (partially) discussed in the lectures; Whenever this is the case, you should briefly recall what was discussed in class and go further. (*iii*) You are welcome to suggest topics that are not in this list. In this case you are invited to find relevant references.

1. Low-dimensional behaviors [BRT19, RZ19].

2. Complexity bounds [GRS18, AGNZ18, SH20].

3. Convex neural nets [Bac17].

4. Stability [BE02, RRT+16, HRS16]

5. Overparametrized linear regression [HMRT19, BLLT20, TB20].

6. Kernel ridge regression [LR20, LRZ19, DL20, MMM21a].

7. Mean field theory of two-layers neural networks [CB18, MMN18, MMM19, CB20, JMM+20, Chi21].

8. Mean field theory of multilayer networks [AOY19, NP20].

9. Implicit/algorithmic bias [GWB+17, LMZ18, GLSS18, SESS19, BM19].

10. Convolutional networks [LWY+19, LZA20, NXB+19, MMM21b].

11. Boosting and interpolation [ZRTH04, RSSZ07, LS20].

12. Optimization in the linear/lazy regime [COB19, ZCZG18, AZLS19, AZLL19, OS20, LZB20a].
13. Generalization in the linear/lazy regime [GMMM19b, MM19, MZ20, MMM21a].
14. Gaps between linear and nonlinear regimes [YS19, GMMM19a, WLLM19, GMMM20, MKAS21].
15. Beyond the linear regime [BL19, AZL19, WGL+20, AZL20, LZB20b].
16. Learning and long tail [Fel20, FZ20].
17. Robustness [MMS+17, SST+18, RXY+19, BLN20].

,

References

[AGNZ18] Sanjeev Arora, Rong Ge, Behnam Neyshabur, and Yi Zhang, *Stronger generalization bounds for deep nets via a compression approach*, Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research, vol. 80, PMLR, 10–15 Jul 2018, pp. 254–263.

[AOY19] Dyego Araújo, Roberto I Oliveira, and Daniel Yukimura, *A mean-field limit for certain deep neural networks*, arXiv preprint arXiv:1906.00193 (2019).

[AZL19] Zeyuan Allen-Zhu and Yuanzhi Li, *What can resnet learn efficiently, going beyond kernels?*, arXiv preprint arXiv:1905.10337 (2019).

[AZL20] ______, *Backward feature correction: How deep learning performs deep learning*, arXiv preprint arXiv:2001.04413 (2020).

[AZLL19] Zeyuan Allen-Zhu, Yuanzhi Li, and Yingyu Liang, *Learning and generalization in overparameterized neural networks, going beyond two layers*, Advances in Neural Information Processing Systems, 2019, pp. 6155–6166.

[AZLS19] Zeyuan Allen-Zhu, Yuanzhi Li, and Zhao Song, *A convergence theory for deep learning via over-parameterization*, International Conference on Machine Learning, PMLR, 2019, pp. 242–252.

[Bac17] Francis Bach, *Breaking the curse of dimensionality with convex neural networks*, The Journal of Machine Learning Research **18** (2017), no. 1, 629–681.

[BE02] Olivier Bousquet and André Elisseeff, *Stability and generalization*, Journal of machine learning research **2** (2002), no. Mar, 499–526.

[BL19] Yu Bai and Jason D Lee, *Beyond linearization: On quadratic and higher-order approximation of wide neural networks*, arXiv:1910.01619 (2019).

[BLLT20] Peter L Bartlett, Philip M Long, Gábor Lugosi, and Alexander Tsigler, *Benign overfitting in linear regression*, Proceedings of the National Academy of Sciences (2020).

[BLN20] Sébastien Bubeck, Yuanzhi Li, and Dheeraj Nagaraj, *A law of robustness for two-layers neural networks*, arXiv preprint arXiv:2009.14444 (2020).

[BM19] Alberto Bietti and Julien Mairal, *On the inductive bias of neural tangent kernels*, arXiv preprint arXiv:1905.12173 (2019).

[BRT19] Mikhail Belkin, Alexander Rakhlin, and Alexandre B Tsybakov, *Does data interpolation contradict statistical optimality?*, The 22nd International Conference on Artificial Intelligence and Statistics, PMLR, 2019, pp. 1611–1619.

[CB18] Lenaic Chizat and Francis Bach, *On the global convergence of gradient descent for over-parameterized models using optimal transport*, Advances in neural information processing systems, 2018, pp. 3036–3046.

[CB20] ———, *Implicit bias of gradient descent for wide two-layer neural networks trained with the logistic loss*, Conference on Learning Theory, PMLR, 2020, pp. 1305–1338.

[Chi21] Lenaic Chizat, *Sparse optimization on measures with over-parameterized gradient descent*, Mathematical Programming (2021), 1–46.

[COB19] Lenaic Chizat, Edouard Oyallon, and Francis Bach, *On lazy training in differentiable programming*, Advances in Neural Information Processing Systems, 2019, pp. 2933–2943.

[DL20] Xialiang Dou and Tengyuan Liang, *Training neural networks as learning data-adaptive kernels: Provable representation and approximation benefits*, Journal of the American Statistical Association (2020), 1–14.

[Fel20] Vitaly Feldman, *Does learning require memorization? a short tale about a long tail*, Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, 2020, pp. 954–959.

[FZ20] Vitaly Feldman and Chiyuan Zhang, *What neural networks memorize and why: Discovering the long tail via influence estimation*, arXiv preprint arXiv:2008.03703 (2020).

[GLSS18] Suriya Gunasekar, Jason D Lee, Daniel Soudry, and Nati Srebro, *Implicit bias of gradient descent on linear convolutional networks*, Advances in Neural Information Processing Systems, 2018, pp. 9461–9471.

[GMMM19a] Behrooz Ghorbani, Song Mei, Theodor Misiakiewicz, and Andrea Montanari, *Limitations of lazy training of two-layers neural network*, Advances in Neural Information Processing Systems, 2019, pp. 9108–9118.

[GMMM19b] ———, *Linearized two-layers neural networks in high dimension*, Annals of Statistics (2019).

[GMMM20] ———, *When do neural networks outperform kernel methods?*, Advances in Neural Information Processing Systems **33** (2020).

[GRS18] Noah Golowich, Alexander Rakhlin, and Ohad Shamir, *Size-independent sample complexity of neural networks*, Proceedings of the 31st Conference On Learning Theory, Proceedings of Machine Learning Research, vol. 75, PMLR, 06–09 Jul 2018, pp. 297–299.

[GWB+17] Suriya Gunasekar, Blake E Woodworth, Srinadh Bhojanapalli, Behnam Neyshabur, and Nati Srebro, *Implicit regularization in matrix factorization*, Advances in Neural Information Processing Systems, 2017, pp. 6151–6159.

[HMRT19] Trevor Hastie, Andrea Montanari, Saharon Rosset, and Ryan J Tibshirani, *Surprises in high-dimensional ridgeless least squares interpolation*, arXiv:1903.08560 (2019).

[HRS16] Moritz Hardt, Ben Recht, and Yoram Singer, *Train faster, generalize better: Stability of stochastic gradient descent*, International Conference on Machine Learning, PMLR, 2016, pp. 1225–1234.

[JMM+20] Adel Javanmard, Marco Mondelli, Andrea Montanari, et al., *Analysis of a two-layer neural network via displacement convexity*, Annals of Statistics **48** (2020), no. 6, 3619–3642.

[LMZ18] Yuanzhi Li, Tengyu Ma, and Hongyang Zhang, *Algorithmic regularization in over-parameterized matrix sensing and neural networks with quadratic activations*, Proceedings of the 31st Conference on Learning Theory, Proceedings of Machine Learning Research, vol. 75, PMLR, 06–09 Jul 2018, pp. 2–47.

[LR20] Tengyuan Liang and Alexander Rakhlin, *Just interpolate: Kernel "ridgeless" regression can generalize*, Annals of Statistics **48** (2020), no. 3, 1329–1347.

[LRZ19] Tengyuan Liang, Alexander Rakhlin, and Xiyu Zhai, *On the risk of minimum-norm interpolants and restricted lower isometry of kernels*, arXiv preprint arXiv:1908.10292 (2019).

[LS20] Tengyuan Liang and Pragya Sur, *A precise high-dimensional asymptotic theory for boosting and minimum-l1-norm interpolated classifiers*.

[LWY+19] Zhiyuan Li, Ruosong Wang, Dingli Yu, Simon S Du, Wei Hu, Ruslan Salakhutdinov, and Sanjeev Arora, *Enhanced convolutional neural tangent kernels*, arXiv preprint arXiv:1911.00809 (2019).

[LZA20] Zhiyuan Li, Yi Zhang, and Sanjeev Arora, *Why are convolutional nets more sample-efficient than fully-connected nets?*, arXiv preprint arXiv:2010.08515 (2020).

[LZB20a] Chaoyue Liu, Libin Zhu, and Mikhail Belkin, *On the linearity of large non-linear models: when and why the tangent kernel is constant*, arXiv preprint arXiv:2010.01092 (2020).

[LZB20b] ———, *Toward a theory of optimization for over-parameterized systems of non-linear equations: the lessons of deep learning*, arXiv preprint arXiv:2003.00307 (2020).

[MKAS21] Eran Malach, Pritish Kamath, Emmanuel Abbe, and Nathan Srebro, *Quantifying the benefit of using differentiable learning over tangent kernels*, arXiv:2103.01210 (2021).

[MM19] Song Mei and Andrea Montanari, *The generalization error of random features regression: Precise asymptotics and double descent curve*, Communications in Pure and Applied Mathematics (2019), arXiv preprint arXiv:1908.05355.

[MMM19] Song Mei, Theodor Misiakiewicz, and Andrea Montanari, *Mean-field theory of two-layers neural networks: dimension-free bounds and kernel limit*, Proceedings of the Thirty-Second Conference on Learning Theory, Proceedings of Machine Learning Research, vol. 99, PMLR, 25–28 Jun 2019, pp. 2388–2464.

[MMM21a] ———, *Generalization error of random features and kernel methods: hypercontractivity and kernel matrix concentration*, arXiv:2101.10588 (2021).

[MMM21b] ———, *Learning with invariances in random features and kernel models*, arXiv preprint arXiv:2102.13219 (2021).

[MMN18] Song Mei, Andrea Montanari, and Phan-Minh Nguyen, *A mean field view of the landscape of two-layer neural networks*, Proceedings of the National Academy of Sciences **115** (2018), no. 33, E7665–E7671.

[MMS+17] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu, *Towards deep learning models resistant to adversarial attacks*, arXiv preprint arXiv:1706.06083 (2017).

[MZ20] Andrea Montanari and Yiqiao Zhong, *The interpolation phase transition in neural networks: Memorization and generalization under lazy training*, arXiv:2007.12826 (2020).

[NP20] Phan-Minh Nguyen and Huy Tuan Pham, *A rigorous framework for the mean field limit of multilayer neural networks*, arXiv preprint arXiv:2001.11443 (2020).

[NXB+19] Roman Novak, Lechao Xiao, Yasaman Bahri, Jaehoon Lee, Greg Yang, Daniel A. Abolafia, Jeffrey Pennington, and Jascha Sohl-dickstein, *Bayesian deep convolutional networks with many channels are gaussian processes*, International Conference on Learning Representations, 2019.

[OS20] Samet Oymak and Mahdi Soltanolkotabi, *Toward moderate overparameterization: Global convergence guarantees for training shallow neural networks*, IEEE Journal on Selected Areas in Information Theory **1** (2020), no. 1, 84–105.

[RRT+16] Maxim Raginsky, Alexander Rakhlin, Matthew Tsao, Yihong Wu, and Aolin Xu, *Information-theoretic analysis of stability and bias of learning algorithms*, 2016 IEEE Information Theory Workshop (ITW), IEEE, 2016, pp. 26–30.

[RSSZ07] Saharon Rosset, Grzegorz Swirszcz, Nathan Srebro, and Ji Zhu, *l 1 regularization in infinite dimensional feature spaces*, International Conference on Computational Learning Theory, Springer, 2007, pp. 544–558.

[RXY+19] Aditi Raghunathan, Sang Michael Xie, Fanny Yang, John C Duchi, and Percy Liang, *Adversarial training can hurt generalization*, arXiv preprint arXiv:1906.06032 (2019).

[RZ19] Alexander Rakhlin and Xiyu Zhai, *Consistency of interpolation with laplace kernels is a high-dimensional phenomenon*, Proceedings of the Thirty-Second Conference on Learning Theory, Proceedings of Machine Learning Research, vol. 99, PMLR, 25–28 Jun 2019, pp. 2595–2623.

[SESS19] Pedro Savarese, Itay Evron, Daniel Soudry, and Nathan Srebro, *How do infinite width bounded norm networks look in function space?*, Proceedings of the Thirty-Second Conference on Learning Theory, Proceedings of Machine Learning Research, vol. 99, PMLR, 25–28 Jun 2019, pp. 2667–2690.

[SH20] Johannes Schmidt-Hieber, *Nonparametric regression using deep neural networks with relu activation function*, Annals of Statistics **48** (2020), no. 4, 1875–1897.

[SST+18] Ludwig Schmidt, Shibani Santurkar, Dimitris Tsipras, Kunal Talwar, and Aleksander Madry, *Adversarially robust generalization requires more data*, arXiv preprint arXiv:1804.11285 (2018).

[TB20] Alexander Tsigler and Peter L Bartlett, *Benign overfitting in ridge regression*, arXiv preprint arXiv:2009.14286 (2020).

[WGL+20] Blake Woodworth, Suriya Gunasekar, Jason D Lee, Edward Moroshko, Pedro Savarese, Itay Golan, Daniel Soudry, and Nathan Srebro, *Kernel and rich regimes in overparametrized models*, Conference on Learning Theory, PMLR, 2020, pp. 3635–3673.

[WLLM19] Colin Wei, Jason D Lee, Qiang Liu, and Tengyu Ma, *Regularization matters: Generalization and optimization of neural nets vs their induced kernel*, Advances in Neural Information Processing Systems, 2019, pp. 9709–9721.

[YS19] Gilad Yehudai and Ohad Shamir, *On the power and limitations of random features for understanding neural networks*, Advances in Neural Information Processing Systems, 2019, pp. 6598–6608.

[ZCZG18] Difan Zou, Yuan Cao, Dongruo Zhou, and Quanquan Gu, *Stochastic gradient descent optimizes over-parameterized deep relu networks*, arXiv:1811.08888 (2018).

[ZRTH04] Ji Zhu, Saharon Rosset, Robert Tibshirani, and Trevor J Hastie, *1-norm support vector machines*, Advances in Neural Information Processing Systems, 2004, pp. 49–56.