STAT260 Mean Field Asymptotics in Statistical Learning

Final project

Song Mei *Due: May 9th 2021*

Project logistics

The final project can either be a literature review or an original research.

1. **Literature review**. We will provide a list of related papers not covered in the lectures, and the literature review should involve in-depth summaries and exposition of one of these papers.

2. **Original research**. It can be either theoretic or experimental (ideally a mix of the two), with approval from the instructor. If you choose this option, you can do it either individually or in groups of two. You are encouraged to combine your current research with your final project.

There are 3 milestones / deliverables to help you through the process.

1. **Proposal (due Feb. 28)**. Submit a short report (no more than 1 page) to the instructor stating the papers you plan to survey or the research problems that you plan to work on. Describe why they are important or interesting, and provide some appropriate references. If you elect to do original research, please do not propose an overly ambitious project that cannot be completed by the end of the semester, and do not be too lured by generality. Focus on the simplest scenarios that can capture the issues you'd like to address.

2. **In-class presentation (Apr. 26 and Apr. 28)**. Prepare an oral presentation with slides (the exact time will depend on the number of projects in the class). Focus on high-level ideas, and leave most technical details to your report.

3. **A written report (due May 9)**. You are expected to submit a final project report up to 8 pages in NeurIPS format with unlimited appendix summarizing your findings / contributions. You must turn in an electronic copy to the instructor or submit via gradescope.

Project topics

The following are a few references collected under certain topics. You are free to dig up more references of relevance for the project you end up pursuing.

Deep learning theory

- 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

- Oussama Dhifallah and Yue M Lu. A precise performance analysis of learning with random features. *arXiv preprint arXiv:2008.11904*, 2020

- Hong Hu and Yue M Lu. Universality laws for high-dimensional learning with random features. *arXiv preprint arXiv:2009.07669*, 2020

- Greg Yang and Edward J Hu. Feature learning in infinite-width neural networks. *arXiv preprint arXiv:2011.14522*, 2020

- Cong Fang, Jason D Lee, Pengkun Yang, and Tong Zhang. Modeling from features: a mean-field framework for over-parameterized deep neural networks. *arXiv preprint arXiv:2007.01452*, 2020
- Yiping Lu, Chao Ma, Yulong Lu, Jianfeng Lu, and Lexing Ying. A mean-field analysis of deep resnet and beyond: Towards provable optimization via overparameterization from depth. *arXiv preprint arXiv:2003.05508*, 2020
- Zeyuan Allen-Zhu and Yuanzhi Li. Towards understanding ensemble, knowledge distillation and self-distillation in deep learning. *arXiv preprint arXiv:2012.09816*, 2020

Computational and statistical gaps

- Matthew Brennan, Guy Bresler, Samuel B Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low-degree tests are almost equivalent. *arXiv preprint arXiv:2009.06107*, 2020
- Matthew Brennan and Guy Bresler. Reducibility and statistical-computational gaps from secret leakage. *arXiv preprint arXiv:2005.08099*, 2020
- Michael Celentano and Andrea Montanari. Fundamental barriers to high-dimensional regression with convex penalties. *arXiv preprint arXiv:1903.10603*, 2019
- Michael Celentano, Andrea Montanari, and Yuchen Wu. The estimation error of general first order methods. *arXiv preprint arXiv:2002.12903*, 2020
- Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *arXiv preprint arXiv:2001.00904*, 2020
- Gérard Ben Arous, Alexander S Wein, and Ilias Zadik. Free energy wells and overlap gap property in sparse pca. In *Conference on Learning Theory*, pages 479–482. PMLR, 2020
- Mrinalkanti Ghosh, Fernando Granha Jeronimo, Chris Jones, Aaron Potechin, and Goutham Rajendran. Sum-of-squares lower bounds for sherrington-kirkpatrick via planted affine planes. *arXiv preprint arXiv:2009.01874*, 2020
- Alessandro Rudi, Ulysse Marteau-Ferey, and Francis Bach. Finding global minima via kernel approximations. *arXiv preprint arXiv:2012.11978*, 2020
- Ashwin Pananjady and Richard J Samworth. Isotonic regression with unknown permutations: Statistics, computation, and adaptation. *arXiv preprint arXiv:2009.02609*, 2020

High dimensional estimation

- Adel Javanmard, Andrea Montanari, and Federico Ricci-Tersenghi. Phase transitions in semidefinite relaxations. *Proceedings of the National Academy of Sciences*, 113(16):E2218–E2223, 2016
- Hua Wang, Yachong Yang, Zhiqi Bu, and Weijie Su. The complete lasso tradeoff diagram. *Advances in Neural Information Processing Systems*, 33, 2020
- Michael Celentano, Andrea Montanari, and Yuting Wei. The lasso with general gaussian designs with applications to hypothesis testing. *arXiv preprint arXiv:2007.13716*, 2020
- Marco Mondelli and Ramji Venkataramanan. Approximate message passing with spectral initialization for generalized linear models. *arXiv preprint arXiv:2010.03460*, 2020
- Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, and Yuling Yan. Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *arXiv preprint arXiv:1902.07698*, 2019

- Christos Thrampoulidis, Samet Oymak, and Mahdi Soltanolkotabi. Theoretical insights into multiclass classification: A high-dimensional asymptotic view. *Advances in Neural Information Processing Systems*, 33, 2020
- Wei-Kuo Chen and Wai-Kit Lam. Universality of approximate message passing algorithms. *arXiv preprint arXiv:2003.10431*, 2020

Multiple hypothesis testing

- Asaf Weinstein, Rina Barber, and Emmanuel Candes. A power and prediction analysis for knockoffs with lasso statistics. *arXiv preprint arXiv:1712.06465*, 2017
- Wenshuo Wang and Lucas Janson. A power analysis of the conditional randomization test and knockoffs. *arXiv preprint arXiv:2010.02304*, 2020
- Asaf Weinstein, Weijie J Su, Małgorzata Bogdan, Rina F Barber, and Emmanuel J Candès. A power analysis for knockoffs with the lasso coefficient-difference statistic. *arXiv preprint arXiv:2007.15346*, 2020
- Zhiqi Bu, Jason M Klusowski, Cynthia Rush, and Weijie J Su. Algorithmic analysis and statistical estimation of slope via approximate message passing. *IEEE Transactions on Information Theory*, 2020
- Jingbo Liu and Philippe Rigollet. Power analysis of knockoff filters for correlated designs. In *Advances in Neural Information Processing Systems*, pages 15446–15455, 2019

Statistical learning

- Xiangyu Chang, Yingcong Li, Samet Oymak, and Christos Thrampoulidis. Provable benefits of overparameterization in model compression: From double descent to pruning neural networks. *arXiv preprint arXiv:2012.08749*, 2020
- Hongyang R Zhang, Fan Yang, Sen Wu, Weijie J Su, and Christopher Ré. Sharp bias-variance tradeoffs of hard parameter sharing in high-dimensional linear regression. *arXiv preprint arXiv:2010.11750*, 2020
- Yu Bai, Minshuo Chen, Pan Zhou, Tuo Zhao, Jason D Lee, Sham Kakade, Huan Wang, and Caiming Xiong. How important is the train-validation split in meta-learning? *arXiv preprint arXiv:2010.05843*, 2020
- Luca Saglietti and Lenka Zdeborová. Solvable model for inheriting the regularization through knowledge distillation. *arXiv preprint arXiv:2012.00194*, 2020
- Adel Javanmard, Mahdi Soltanolkotabi, and Hamed Hassani. Precise tradeoffs in adversarial training for linear regression. *arXiv preprint arXiv:2002.10477*, 2020
- Adel Javanmard and Mahdi Soltanolkotabi. Precise statistical analysis of classification accuracies for adversarial training. *arXiv preprint arXiv:2010.11213*, 2020
- Hossein Taheri, Ramtin Pedarsani, and Christos Thrampoulidis. Asymptotic behavior of adversarial training in binary classification. *arXiv preprint arXiv:2010.13275*, 2020

References

[1] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *arXiv preprint arXiv:2001.00904*, 2020.

[2] Zeyuan Allen-Zhu and Yuanzhi Li. Towards understanding ensemble, knowledge distillation and self-distillation in deep learning. *arXiv preprint arXiv:2012.09816*, 2020.

[3] Gérard Ben Arous, Alexander S Wein, and Ilias Zadik. Free energy wells and overlap gap property in sparse pca. In *Conference on Learning Theory*, pages 479–482. PMLR, 2020.

[4] Yu Bai, Minshuo Chen, Pan Zhou, Tuo Zhao, Jason D Lee, Sham Kakade, Huan Wang, and Caiming Xiong. How important is the train-validation split in meta-learning? *arXiv preprint arXiv:2010.05843*, 2020.

[5] Matthew Brennan and Guy Bresler. Reducibility and statistical-computational gaps from secret leakage. *arXiv preprint arXiv:2005.08099*, 2020.

[6] Matthew Brennan, Guy Bresler, Samuel B Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low-degree tests are almost equivalent. *arXiv preprint arXiv:2009.06107*, 2020.

[7] Zhiqi Bu, Jason M Klusowski, Cynthia Rush, and Weijie J Su. Algorithmic analysis and statistical estimation of slope via approximate message passing. *IEEE Transactions on Information Theory*, 2020.

[8] Michael Celentano and Andrea Montanari. Fundamental barriers to high-dimensional regression with convex penalties. *arXiv preprint arXiv:1903.10603*, 2019.

[9] Michael Celentano, Andrea Montanari, and Yuting Wei. The lasso with general gaussian designs with applications to hypothesis testing. *arXiv preprint arXiv:2007.13716*, 2020.

[10] Michael Celentano, Andrea Montanari, and Yuchen Wu. The estimation error of general first order methods. *arXiv preprint arXiv:2002.12903*, 2020.

[11] Xiangyu Chang, Yingcong Li, Samet Oymak, and Christos Thrampoulidis. Provable benefits of overparameterization in model compression: From double descent to pruning neural networks. *arXiv preprint arXiv:2012.08749*, 2020.

[12] Wei-Kuo Chen and Wai-Kit Lam. Universality of approximate message passing algorithms. *arXiv preprint arXiv:2003.10431*, 2020.

[13] Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, and Yuling Yan. Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *arXiv preprint arXiv:1902.07698*, 2019.

[14] Oussama Dhifallah and Yue M Lu. A precise performance analysis of learning with random features. *arXiv preprint arXiv:2008.11904*, 2020.

[15] Cong Fang, Jason D Lee, Pengkun Yang, and Tong Zhang. Modeling from features: a mean-field framework for over-parameterized deep neural networks. *arXiv preprint arXiv:2007.01452*, 2020.

[16] Mrinalkanti Ghosh, Fernando Granha Jeronimo, Chris Jones, Aaron Potechin, and Goutham Rajendran. Sum-of-squares lower bounds for sherrington-kirkpatrick via planted affine planes. *arXiv preprint arXiv:2009.01874*, 2020.

[17] Hong Hu and Yue M Lu. Universality laws for high-dimensional learning with random features. *arXiv preprint arXiv:2009.07669*, 2020.

[18] Adel Javanmard, Andrea Montanari, and Federico Ricci-Tersenghi. Phase transitions in semidefinite relaxations. *Proceedings of the National Academy of Sciences*, 113(16):E2218–E2223, 2016.

[19] Adel Javanmard and Mahdi Soltanolkotabi. Precise statistical analysis of classification accuracies for adversarial training. *arXiv preprint arXiv:2010.11213*, 2020.

[20] Adel Javanmard, Mahdi Soltanolkotabi, and Hamed Hassani. Precise tradeoffs in adversarial training for linear regression. *arXiv preprint arXiv:2002.10477*, 2020.

[21] Jingbo Liu and Philippe Rigollet. Power analysis of knockoff filters for correlated designs. In *Advances in Neural Information Processing Systems*, pages 15446–15455, 2019.

[22] Yiping Lu, Chao Ma, Yulong Lu, Jianfeng Lu, and Lexing Ying. A mean-field analysis of deep resnet and beyond: Towards provable optimization via overparameterization from depth. *arXiv preprint arXiv:2003.05508*, 2020.

[23] Marco Mondelli and Ramji Venkataramanan. Approximate message passing with spectral initialization for generalized linear models. *arXiv preprint arXiv:2010.03460*, 2020.

[24] 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.

[25] Ashwin Pananjady and Richard J Samworth. Isotonic regression with unknown permutations: Statistics, computation, and adaptation. *arXiv preprint arXiv:2009.02609*, 2020.

[26] Alessandro Rudi, Ulysse Marteau-Ferey, and Francis Bach. Finding global minima via kernel approximations. *arXiv preprint arXiv:2012.11978*, 2020.

[27] Luca Saglietti and Lenka Zdeborová. Solvable model for inheriting the regularization through knowledge distillation. *arXiv preprint arXiv:2012.00194*, 2020.

[28] Hossein Taheri, Ramtin Pedarsani, and Christos Thrampoulidis. Asymptotic behavior of adversarial training in binary classification. *arXiv preprint arXiv:2010.13275*, 2020.

[29] Christos Thrampoulidis, Samet Oymak, and Mahdi Soltanolkotabi. Theoretical insights into multiclass classification: A high-dimensional asymptotic view. *Advances in Neural Information Processing Systems*, 33, 2020.

[30] Hua Wang, Yachong Yang, Zhiqi Bu, and Weijie Su. The complete lasso tradeoff diagram. *Advances in Neural Information Processing Systems*, 33, 2020.

[31] Wenshuo Wang and Lucas Janson. A power analysis of the conditional randomization test and knockoffs. *arXiv preprint arXiv:2010.02304*, 2020.

[32] Asaf Weinstein, Rina Barber, and Emmanuel Candes. A power and prediction analysis for knockoffs with lasso statistics. *arXiv preprint arXiv:1712.06465*, 2017.

[33] Asaf Weinstein, Weijie J Su, Małgorzata Bogdan, Rina F Barber, and Emmanuel J Candès. A power analysis for knockoffs with the lasso coefficient-difference statistic. *arXiv preprint arXiv:2007.15346*, 2020.

[34] Greg Yang and Edward J Hu. Feature learning in infinite-width neural networks. *arXiv preprint arXiv:2011.14522*, 2020.

[35] Hongyang R Zhang, Fan Yang, Sen Wu, Weijie J Su, and Christopher Ré. Sharp bias-variance tradeoffs of hard parameter sharing in high-dimensional linear regression. *arXiv preprint arXiv:2010.11750*, 2020.