Yunzhen Feng's Statement of Purpose

Machine learning has attracted a lot of attention around the world with its increasing impacts. While its applications have shown great success in media, communication, medical research, and etc., completing the theory underneath is still an intriguing and ongoing work. I aimed to understand the theories behind current data-driven machine learning, and guide the designs of reliable models and algorithms.

My motivation into machine learning theory came from an internship experience at China Academy of Railway Sciences, where I used neural networks to classify trains and capture their serial numbers from images. Surprisingly, a big model trained with only a few training samples performed well on test data. However, the performance degraded drastically when facing noise and adversaries online in real scenarios. These problems related to generalization and robustness stayed in my mind and became the center of my undergraduate researches. Working with Prof. Bin Dong at PKU and Prof. Yue M. Lu at Harvard, I undertook four projects: the generalization of deep networks and semi-supervised learning, Nash Equilibria in adversarial robustness, and enhancement of certified robustness. These experiences helped shape my research interests and cultivate my ability to independently formulate and solve problems.

"Why deep networks generalize well" was the first question that lingered in my mind. I learned that deep ResNet could be characterized as a continuous dynamical system by viewing the layer index as time. For this continuum of ResNet, I established a depth-independent generalization bound using the Rademacher complexity. The difference between discrete ResNet and its continuum can also be added into the bound. Although the result successfully bounds the generalization gap, showing why networks generalize, it is too weak to reflect the benefits of increasing the depth in practice. One step further, such result cannot be compared to guide selecting models in practice. While searching for beauties in theory, I hope the theory can help the practice in return. Thus, I turned to provide an analysis that fits the performances well with methodology from statistical physics.

With this idea, I analyzed the generalization behavior in semi-supervised learning. In this scenario, labeled data are expensive, and unlabeled data are collected to improve trained models' performance. I am wondering: with a limited budget, how many labeled data and unlabeled data should be collected to maximize the performance? Many previous works either replaced the generalization with accuracy on unlabeled data or assumed unlabeled data is free. I approached the problem by calculating the true generalization in the high dimensional limit concerning labeled and unlabeled data ratios. In detail, I analyzed the training of linear classifiers with Laplacian regularization on Gaussian mixtures. The Convex Gaussian Min-Max theorem was generalized to address two technical difficulties: the data-dependent regularization and the general data covariance. The result was exciting: it matched the experimental performances and illustrated how to improve the generalization by changing the ratio of unlabeled and labeled data, the loss function, and the regularization. This paper will be submitted to ICML 2021.

Besides generalization, another mystery in deep learning is adversarial robustness. It is in nature a game between a defender selecting models and an attacker crafting the input. Numerous

publications played this game by proposing attacks and defenses against previous works, but little improvement was made over the past few years. What robustness can be eventually achieved? A fundamental question is whether there exists a Nash equilibrium. If so, there exists an optimal model and we should seek to approximate it; If not, we shall increase the cost of attacks or rethink the notion of robustness. In general cases, the difficulty was that the attacker and the defender could choose strategies in two continuous function spaces. I managed to locate the defenses in a compact set to reach the optimal solution in the inf-sup problem. Each solution could be constructed into a pure Nash equilibrium rigorously in the one-dimensional case. I realized that the equilibrium's existence implied a distribution-dependent preprocessing on the training samples, on which standard training may yield good robustness. This observation was confirmed by experiments and was well connected with Prof. Madrys "robust feature" notion.

Theories provide in-depth understandings, and a combination of theories and practices can further impact. Observing the similarity between ensemble and random features, I have proved that weighted ensembling could achieve near-optimal risk (the generality) and the weights can be efficiently optimized since it is a convex problem. I realized that ensembling could help replace optimization, but how can this understanding help practice? Following previous works that applied randomized smoothing to provide certified adversarial robustness, I realized that the primary challenge was how to train a proper model for smoothing. The randomness in the framework made direct optimization impractical, and that was where ensembling helped. With an algorithm to save computational cost, I employed ensembling and improved the SOTA with 31% less training time and 36% fewer parameters. The generality and optimization guarantees were also proven for certified robustness. This work was submitted to ICLR 2021.

The current literature on deep learning is flourishing with interesting numerical observations and experiments that call for explanations. However, as the number of publications and preprints increases, some works are just post-mortem analysis. I hope I can unravel the theories behind interesting practices, and in return, use theories to improve practices. Right now, most low-hanging fruits have been taken, and it is time to step further to key problems in machine learning theory. How to understand current optimization on finding better solutions? How previous wisdom can overcome the curse of dimensionality? How we can build reliable models with good robustness and generalization? How should we seek interpretations for deep networks? I am devoted towards these questions.