Yunzhen Feng's Statement of Purpose

Thanks to the advances in computing power and new designs, machine learning achieves great success and is being deployed in high-stakes application areas (e.g., industry, government, and health care). Due to the high dimensions and complicated models, there are stronger needs for completing the underneath theories. How accurate will a learned model be when it generalizes? Is it robust to adversarial attacks? Around these two questions, I undertook four undergraduate research aiming to understand the theories behind and to design reliable models and algorithms. These experiences helped shape my research interests and cultivate my ability to independently formulate and solve problems.

Why overparameterized networks generalize well? I decompose the overparameterization into depth and width. I learned that deep ResNet could be characterized as a continuous dynamical system by viewing the layer index as time. Based on this perspective, I established a generalization bound using the Rademacher complexity. Although the result successfully bounds the generalization gap, showing why deep networks generalize, the bound is too weak: as the depth increase, the bound converges to a large value while the generalization in practice keeps diminishing. The culprits are the complexity of the continuum model and the uniform-convergence nature of the bound. Stuck with how to improve, I turned to the statistical physics perspective, hoping that the analysis fitting the performance well could help guide the practice.

Statistical physics can efficiently address the network's width in the high dimensional limit. Advised by Prof. Yue M. Lu at Harvard, I analyzed the generalization in semi-supervised learning. In this scenario, labeled data are expensive, and unlabeled data are collected to improve trained models' performance. I wonder: 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 (unlimited). I wanted to tightly analyze how the true generalization depends on the number of unlabeled and labeled data, and tools like the Convex Gaussian Min-Max theorem helped in the high dimensional limit. In detail, I precisely characterized the training of linear classifiers with Laplacian regularization on Gaussian mixtures and addressed two technical difficulties: the data-dependent regularization and the general data covariance. The result was exciting: it matched the experimental performances when the width dimension is larger than 200; it 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, whether the network is robust to adversarial examples is another vital problem for applications. I approached it from two perspectives.

On the one hand, we want the network to be robust to any imperceptible noises in a small ball. Randomized smoothing can provide the certified radius by enforcing Lipschitzness. However, due to the randomness, it is hard to train the base classifier for smoothing. I recalled my previous understandings of the ensembling's benefits on optimization. With enough diverse models, the weighted ensemble can achieve near-optimal risk, and the weights can be efficiently optimized due to convexity. Therefore, I combined ensembling with randomized smoothing and

theoretically proved the optimization guarantees for certified robustness. With an algorithm to save computational cost, I improved the SOTA with 31% less training time and 36% fewer parameters. These results are excellent, but I am not surprised: as theories provide in-depth understandings, the combination of theories and practices can definitely further impact.

On the other hand, adversarial robustness is, in nature, a game between a defender selecting models and an attacker crafting the input. Numerous publications participated in this game by proposing attacks and defenses against previous works, but little improvement was made over the past two years. Will such a process end to an optimal defense, and what robustness can be eventually achieved? This question is equivalent to whether there exists a Nash equilibrium. If so, we can combine game-theoretic algorithms with adversarial training to approximate the equilibrium; If not, we may turn to increase the attacker's costs as current practices in cyber-security. Advised by Prof. Bin Dong, I approached it without assumptions on the attacks and defenses. In the general case, traditional wisdom cannot be used due to the two continuous strategy spaces that contain functions. Using functional analysis and optimal transport, I located that strategy in compact sets and rigorously constructed Nash equilibria in the one-dimensional case. This result provided another explanation for the "robust feature" notion in related works.

The current literature on deep learning is flourishing with interesting numerical observations and experiments that call for explanations. I want to go beyond post-mortem analysis: unraveling the theories behind interesting practices, and in return, using theories to improve practices. Right now, as most low-hanging fruits have been taken, it is time to step further to vital problems. How to understand current optimization on finding generalizable 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 think these are the vital obstacles ahead and I would like to devote my next several years on one of them.