Thesis Defense

MACHINE LEARNING DEPARTMENT

New Advances in Sparse Learning, Deep Networks, and Adversarial Learning: Theory and Applications

Abstract:

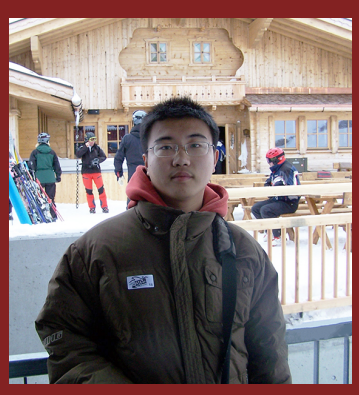
Sparse learning, deep networks, and adversarial learning are new paradigms and have received significant attention in recent years due to their wide applications to various big data problems in computer vision, natural language processing, statistics, and theoretical computer science. The paradigms include learning with sparsity, learning with low-rank approximations, and learning with deep neural networks, corresponding to the assumptions that data lie with only a few non-zero coordinates, lie on low-rank subspaces, and lie on low-dimensional manifolds, respectively. The focus of the study is to develop algorithms which are sample-efficient, are easier to optimize, and are more robust to adversarial corruptions

Despite a large amount of work on these new paradigms, many fundamental questions remain unresolved. From the statistical aspect, understanding the tight sample complexity of big data problems is an important research question. Intuitively, the intrinsic dimension of structured data should be much smaller than their ambient dimension. Because the true sample complexity should be comparable to the intrinsic dimension rather than the ambient dimension, this implies the possibility of sub-linear sample complexity w.r.t. the ambient dimension. In this thesis, we design principled, practical and scalable algorithms for big data problems with near-optimal sample complexity. These include models of matrix completion, robust PCA, margin-based active learning, property testing, compressed sensing.

From the computational aspects, direct formulations of these new paradigms are non-convex and NP-hard to optimize in general. Therefore, one of the long-standing questions is designing computationally efficient algorithms by taking into account the structure of the data. In this thesis, we develop new paradigms toward global optimality of non-convex optimization in polynomial time. In particular, we design algorithms and understand landscape (e.g., duality gap) for the problems of (1-bit) compressed sensing, deep neural network, GAN, matrix factorization.

From the robustness aspects, models such as deep networks are vulnerable to adversarial examples. Although the problem has been widely studied empirically, much remains unknown concerning the theory underlying designing defense methods. There are two types of adversarial example: training-time adversarial example, such as data positioning, and inference-time adversarial example. We discuss both of adversarial examples in this thesis, for the problems of (1-bit) compressed sensing and robust PCA, and deep classification models, respectively.

Beyond theoretical contributions, our work also has significant practical impact. For example, inspired by our theoretical analysis, we design a new defense method TRADES against inference-time adversarial examples. Our proposed algorithm is the winner of the NeurIPS 2018 Adversarial Vision Challenge in which we won the 1st place out of 1,995 submissions, surpassing the runner-up approach by 11.41% in terms of mean L2 perturbation distance.



Speaker:

Hongyang Zhang

Thesis Committee:
Nina Balcan (Co-Chair)
David Woodruff (Co-Chair)
Ruslan Salakhutdinov
Avrim Blum (Toyota Technical Institute at Chicago)

Apr. 25, 2019
10:00am
GHC 9115

Link to draft document:
https://www.dropbox.com/s/xqno5fufdfhvvwx/cmuthesis_template.pdf?dl=0