Yunzhen Feng

felix.feng681@gmail.com · (+86) 150-1082-6920 · Homepage: https://fengyzpku.github.io/

Education

Peking University (PKU), Beijing, China Aug'17 - June'21 (expected)

- *B.S.* in Applied Mathematics.
  - Majoring in Data Science and Big Data Technology.
  - Ranking: 1/18. Major GPA: 3.76/4.0. Overall GPA: 3.66/4.0.
  - Supported by Elite Undergraduate Training Program of Applied Mathematics (top 15%).
  - Advisor: Prof. Bin Dong

Research Interests

I am interested in understanding the theories behind current data-driven machine learning and, in return, using them to guide the designs of reliable models and algorithms. I like to adopt new perspectives from related areas, such as statistical physics, game theory, optimal transport, and differential equation.

Publications and Manuscripts

1. **Yunzhen Feng**, Yue M. Lu. A Precise High-dimensional Analysis of Laplacian Semi-Supervised Learning. *, in preparation for ICML 2021.*
2. **Yunzhen Feng***, Chenkai Yu*, Adam Wierman, Bin Dong. Pure Nash Equilibria in Adversarial Robustness. *, in preparation.*
3. Chizhou Liu, **Yunzhen Feng**, Ranran Wang, Bin Dong. Enhancing Certified Robustness of Smoothed Classifiers via Weighted Model Ensembling. *arXiv:2005.09363, submitted to ICLR 2021.*
4. **Yunzhen Feng***, Runtian Zhai*, He Di, Liwei Wang, Bin Dong. Transferred Discrepancy: Quantifying the Difference Between Representations. *arXiv:2007.12446.*

Research

A Precise High-dimensional Analysis of Laplacian Semi-Supervised Learning (SSL)

with Prof. Yue M. Lu, SEAS, Harvard University Aug'20 - Present

- Consider a high-dimensional mixture of two Gaussians in the noisy regime and the semi-supervised learning setting where only partial labelings are known.
- Provide a rigorous analysis of the generalization error of Laplacian regularized convex classifiers in the high-dimensional limit when the number of labeled samples, the number of unlabeled samples, and their dimension go to infinity with a fixed ratio.
  - Generalize previous works to data-dependent regularization and general data covariances.
  - Discuss how the ratios of labeled data and unlabeled data affect the generalization error.
  - Observe a transition between unlabeled data's harms and benefits to the classification.

Pure Nash Equilibria in Adversarial Robustness

with Prof. Bin Dong, BICMR, Peking University July'20 - Present

- Analyze the adversarial robustness in a game-theoretic perspective and prove the existence of pure Nash equilibria for binary classification in the one-dimensional case.
  - The only assumption is the continuity of data distribution's density.
  - Locate the optimal solution to the inf-sup problem in a compact set that has regularity in the classification boundary. The solution is therefore attainable.
  - Construct a pure Nash equilibrium with each solution to the inf-sup problem, inspired by optimal transport. If adversarial training converges to one solution in the inf-sup problem in the one-dimensional case, the trained classifier is an optimal defense for any attacks.
- Construct a data-dependent preprocessing on inputs from synthetic datasets. Standard training on the processed data can have non-trivial robustness.

Enhancing Certified Robustness with Weighted Ensembling
with Prof. Bin Dong, BICMR, Peking University Jan'20 - June'20

- Proposed a weighted ensembling scheme (SWEEN) to improve certified robustness for randomized smoothing with extensive experiments.
  - Ensembling several small diverse models outperformed an individual model with more parameters.
  - Improved the SOTA with 31% less training time and 36% fewer FLOPs.
  - Applied an adaptive prediction algorithm to save computational cost by at least 45%.
- Provided theoretical guarantees on the scheme:
  - Generality: SWEEN can achieve optimal certified robustness,
  - Optimization: SWEEN can be provably trained to achieve near-optimal risk with a surrogate loss.

Quantifying the Difference between Representations
with Dr. Di He, Microsoft Research Asia Jan'20 - June'20

- Proposed Transferred Discrepancy (TD), a new metric to measure the difference between two networks' representation based on their downstream performances.
- Analyzed TD's connection with Canonical Correlation Analysis in linear probing setting.
- Used TD to investigate the effect of various deep learning factors, and discovered that the factors improving the performance always lead to more similar representations under different random initializations.

Other Experiences

Competition on Kaggle: Long-term Traffic Flow Prediction Dec'19

- Optimized 3D-convolution-based and WaveNet-based networks for temporal sequential data.
- Ranked 11 / 75.

Research Experiences for Undergraduates Program
with Prof. Bin Dong, BICMR, Peking University Mar'19 - Jan'20

- Proposed a generalization bound exploiting the connection between deep ResNets and differential equations.
- Analyzed how training labels and label noises affected adversarial robustness with Neural Tangent Kernel.

Specialized Skills

- **Programming**: Python (PyTorch), MATLAB, LaTeX, C.
- **Language**: English, TOEFL(108 with S26), GRE(159 + 170 + 3.5).

Selected Honors and Awards

Bronze Medal, S.-T.Yau College Student Mathematics Contests 2020
Meritorious Award, Mathematical Contest in Modeling (top 9%) 2019
The elite undergraduate training program of Applied Math (top 15%) 2019
Gold Medal, China Mathematics Olympiad (top 100, same level as USAMO) 2016

Selected Courses

Statistical Learning 3.85 Deep Learning: Advanced Topics 3.81 Introduction to Data Science 3.88
Mathematical Theory of Neural Networks Models (Princeton, Prof. Weinan E) 3.97
Topology 3.85 Partial Differential Equations 3.73 Ordinary Differential Equations 3.85
Randomized algorithms 3.97 Data Structure and Algorithm 3.83 Introduction to Computation 4.0

Activities and Interests

- 3.0 (NTRP standard) Tennis Player.
- Rock-climber and outdoor hiker trained by committees from Chinese Mountaineering Association.
- One-year volunteer at Arthur M. Sackler Museum of Art and Archaeology, Peking University.