Qingru Zhang

qrzhang98@gmail.com • qingruzhang.github.io • QingruZhang

Research Interests

Optimization and generalization of deep learning , Machine Learning

Education

Shanghai Jiao Tong University **Shanghai, China**
B.Eng. in Cyber Security *Sep. 2016 - Jun. 2020 (expected)*
- Zhiyuan Honors Program of Engineering (an elite program for top 5% students in SJTU)
- Cumulative GPA: Overall: 89.05/100 | Major: 89.2/100 (10/103)

Selected Courses: Discrete Mathematics (100, 1/277), Mathematical Analysis (H) I (94, 8/205), Mathematical Analysis (H) II (93, 8/225), Probability and Statistics (95, 202/1259), Information theory and Coding (94, 5/108), Data Mining (95, 1/55), Course Design on Application Software (95, 1/103), Electronics System Design Based on Android (100, 1/93)

Research Internships

University of Toronto **Ontario, Canada**
Research Intern, Vector Institute, Machine Learning Group of UofT *Jul. 2019 - Oct. 2019*
- **Research Focus**: Stochastic Variance Reduction Method
- **Advisor**: Prof. Jimmy Ba

Shanghai Jiao Tong University **Shanghai, China**
Research Assistant, Apex Data and Knowledge Management Lab *Jun. 2018 - Jun. 2019*
- **Research Focus**: Optimization and Neural Network
- **Advisor**: Prof. Weinan Zhang and Prof. Yong Yu

Publications and Manuscripts

A Non-asymptotic comparison of SVRG and SGD: tradeoffs between compute and speed.
- **Qingru Zhang**, Yuhuai Wu, Fartash Faghri, Tianzong Zhang, Jimmy Ba.
- *Submitted to International Conference on Learning Representations*, **ICLR 2020**

AdaShift: Decorrelation and Convergence of Adaptive Learning Rate Methods.
- Zhiming Zhou*, **Qingru Zhang***, Guansong Lu, Hongwei Wang, Weinan Zhang, Yong Yu.
- *Published on International Conference on Learning Representations*, **ICLR 2019**

* denotes equal contribution (co-first author).

Research Projects

A Non-asymptotic analysis on SVRG. **Jul. 2019 - Oct. 2019**
Advisor: Prof. Jimmy Ba *Machine Learning Group, University of Toronto*
- The traditional asymptotic analysis in SVRG provides limited insight into its ineffectiveness in training deep learning models under a fixed number of epochs.
- Provided a non-asymptotic analysis of SVRG under a noisy least squares regression problem and connected our analysis to neural networks via Neural Tangent Kernel (NTK).
- Independently derived the learning dynamic and exact expected loss for both SVRG and SGD.
- Compared their exact expected loss for two settings: with and without label noise.
- When with label noise, the learning dynamics of our regression model closely matched with that of under-parameterized neural networks on MNIST and CIFAR-10.
- There is a trade-off between the computational cost and the convergence speed in under-parametrized neural networks. SVRG outperforms SGD after a few epochs in this regime.
- When without label noise, the learning dynamics of our regression model closely matched with experiments of over-parameterized neural networks on MNIST and CIFAR-10.
- SGD is shown to always outperform SVRG in the over-parameterized regime.

AdaShift: resolve the non-convergence of Adam via temporal shifting **Jul. 2018 - Nov. 2018**
Advisor: Prof. Weinan Zhang and Prof. Yong Yu *Shanghai Jiao Tong University*

- Adam optimizer shows non-convergence behaviors under some situations.
- Theoretically analyzed the role of exponential decay rates on these situations.
- Drew the reasons of non-convergence lie in the biased step size caused by the tight correlation.
- Proposed a revised optimizer, AdaShift, based on temporal shifting.
- Experimented and compared AdaShift with Adam and AMSGrad on various benchmark tasks.
- AdaShift outperforms AMSGrad in term of convergence speed and generalization on these tasks.
- AdaShift can address the non-convergence issue of Adam, while still maintaining a competitive performance with Adam in terms of both training speed and generalization.

A Survey on the generalization theory of neural network **Feb. 2019 - May. 2019**
Advisor: Prof. Weinan Zhang *Shanghai Jiao Tong University*

- Conducted a survey on the generalization theory and its related work.
- Studied the proof of various generalization bounds and obtained a deep understanding on PAC-Bayesian framework, Rademacher complexity and VC-dimension.
- Planned to extend PAC-Bayesian generalization bounds by considering the difference between structured and unstructured data.

Honors and Awards

- **National Cyber Security Scholarship** (RMB 30,000) (Top 2% in School of Cyber Security), China Internet Development Foundation. 2019.
- **ICLR Travel Award**, ICLR2019. 2019.
- **Zhiyuan College Honors Scholarship** (Top 5%), Shanghai Jiao Tong University. 2017 & 2018.
- **Pan Wenyuan Scholarship** (10/337), Shanghai Jiao Tong University. 2017.
- **Meritorious Winner of The Mathematical Contest in Modeling** (Top 10% out of 10670), Consortium for Mathematics and Its Application. 2018& 2019.
- **Second Prize** in China Undergraduate Mathematical Contest in Modeling, Shanghai Division. 2017.
- **SJTU Merit Students**, Shanghai Jiao Tong University. 2017.

Extracurriculars

Experience of Summer Support Education

- As a volunteer teacher, attended a support educational program of SJTU in July 2017. Taught math, art and physical classes to the primary students of a village in Henan province, China.

Monitor of Class, School of Electrical Engineering

- As the class leader, take charge of daily issues of my class.

Interests & Skills

- **Hobbies**: Basketball, Tennis, Badminton, Table Tennis, Hiking, Movie.
- **Skills**: Python (Tensorflow, Pytorch) , C++ , Matlab , JavaScript, Verilog , LaTeX.