Yu Bai

Website: https://yubai.org E-mail: yub@stanford.edu

EDUCATION

***Ph.D. in Statistics*, Stanford University**, Stanford, CA 2014 - 2019
Specialization in Statistical Learning Theory and Non-convex Optimization.
Advisor: Prof. John Duchi

***B.S. in Mathematics*, Peking University**, Beijing, China Sep 2010 - July 2014
GPA: 3.9 / 4, *Honored Graduate*

RESEARCH INTERESTS

- Optimization and generalization theories for deep neural networks;
- Statistical reinforcement learning
- Generative models
- Efficient training and inference for neural networks (e.g. quantization, compression)

PUBLICATIONS

(1) **Provably Efficient Q-Learning with Low Switching Cost.**
Yu Bai, Tengyang Xie, Nan Jiang, Yu-Xiang Wang, 2019.
Submitted. arXiv preprint arXiv:1905.12849.

(2) **ProxQuant: Quantized Neural Networks via Proximal Operators**
Yu Bai, Edo Liberty, Yu-Xiang Wang, 2018.
ICLR 2019. arXiv preprint arXiv:1810.00861.

(3) **Subgradient Descent Learns Orthogonal Dictionaries.**
Yu Bai, Qijia Jiang, Ju Sun, 2018.
ICLR 2019. arXiv preprint arXiv:1810.10702.

(4) **Approximability of Discriminators Implies Diversity in GANs.**
Yu Bai, Tengyu Ma, Andrej Risteski, 2018.
ICLR 2019. arXiv preprint arXiv:1806.10586.

(5) **Model Fidelity, Randomization, and Adaptivity in Linear Experimental Design.**
Yu Bai, John Duchi, 2018+. *In preparation.*

(6) **Analysis of Sequential Quadratic Programming through the Lens of Riemannian Optimization.**
Yu Bai, Song Mei, 2018. arXiv preprint arXiv:1805.08756.

(7) **Proximal Algorithms for Constrained Composite Optimization, with Applications to Solving Low-rank SDPs**
Yu Bai, Song Mei, John Duchi, 2019.

(8) **TAPAS: Two-pass Approximate Adaptive Sampling for Softmax.**
Yu Bai, Sally Goldman, Li Zhang, 2017. arXiv preprint arXiv:1707.03073.

(9) **The Landscape of Empirical Risk for Non-convex Losses.**
Song Mei, Yu Bai, Andrea Montanari.
The Annals of Statistics 46 (6A), 2747-2774, 2018. arXiv preprint arXiv:1607.06534.

INTERNSHIPS

Research Intern, *Amazon AI* Palo Alto, CA
Host: Edo Liberty & Yu-Xiang Wang June 2018 - Sep 2018

Proposed ProxQuant, a prox-gradient method with quantization-inducing regularizers for training quantized neural networks. The training adds a simple prox-operator step in between existing full-precision training. On ResNets and LSTMs, ProxQuant beats state-of-the-art methods on binary quantization and achieves comparable performance on multi-bit quantization. Further theoretical and empirical evidence suggests that the optimization stability of ProxQuant is better than the commonly used straight-through gradient method.

Research Intern, *Google Research* Mountain View, CA
Host: Li Zhang June 2016 - Sep 2016

Proposed adaptive sampling strategies for softmax in feedforward neural networks for extreme classification. The adaptive sampling works better than non-adaptive strategies on simulated datasets and achieves new state-of-the-art accuracy on a large-scale Youtube benchmark dataset. The algorithm was made available in Tensorflow (`tf.contrib.nn.rank_sampled_softmax_loss`).

PRESENTATIONS

Subgradient Descent Learns Orthogonal Dictionaries
ICLR, May 2019.

ProxQuant: Quantizing Neural Networks via Proximal Operators
ICLR, May 2019.
Bytedance AI Lab, December 2018.
Amazon AI, September 2018.

On the Generalization and Approximation in GANs
ICLR, May 2019. Google Brain, November 2018.
Salesforce Research, November 2018.
Stanford ML Seminar, October 2018.

Optimization Landscape of Some Non-convex Learning Problems
Stanford Theory Seminar, April 2018.
Stanford ML Seminar, April 2017.

REVIEWING EXPERIENCE

Conference reviewing: NIPS (top 30% reviewer), ICML, ICLR, AISTATS, IEEE-ISIT.
Journal reviewing: JMLR, IEEE-TSP, SICON (SIAM Journal on Control and Optimization).

SELECTED COURSEWORK

Reinforcement Learning (CS234).
Convolutional Neural Networks for Visual Recognition (CS231N).
Theories of Deep Learning (Stats385).
Numerical Linear Algebra (CME302).
Inference, Estimation, and Information Processing (EE378B).
Machine Learning Theory (CS229T).
Convex Optimization (EE364A).
Information Theory and Statistics (Stats311/EE377).
Theory of Statistics (Stats300A/B/C).
Theory of Probability (Stats310A/B/C).

TEACHING EXPERIENCE

As Instructor:

Guest Lecturer, Theory of Statistics (Stats300B).
Session Instructor, Theory of Probability (Stats310A).

As Teaching Assistant (selected):

Statistical Learning Theory (CS229T), as head TA.
Modern Markov Chains (Stats 318).
Theory of Probability (Stats310A/B/C).
Theory of Statistics (Stats300A/B).
Statistical Inference (Stats200).
Introduction to Stochastic Processes (Stats217).