

ISYE8803 Advanced Machine Learning

SPRING 2018

Instructor: Tuo Zhao

Time and Location: MW 3:00-4:15pm, Instructional Center 213

Contact: tourzhao@gatech.edu

Office Hours: T 2-3pm, Groseclose 344

TAs: Zhehui Chen

References

- Lecture Notes
- Foundations of Machine Learning, Mohri, Rostamizadeh and Talwalker
- Peter Bartlett. CS281B Lecture Slides: <https://bcourses.berkeley.edu/courses/1409209>

Honor code

For any questions involving Academic Honor Code issues, please consult me, the class teaching assistant or <http://www.honor.gatech.edu>.

Presentation

Each student will be required to give a 35-minute presentation.

Covered Topics

- Statistical learning theory
- Sparse Representation Learning
- Low Rank Representation Learning
- Theory of deep learning
- Nonconvex optimization

Reading List

- (DL) Spectrally-normalized margin bounds for neural networks
- (DL) Nearly-tight VC-dimension and pseudodimension bounds for piecewise linear neural networks
- (DL) A PAC-Bayesian Approach to Spectrally-Normalized Margin Bounds for Neural Networks
- (DL) Benefits of depth in neural networks
- (SRL) A Unified Framework for High-Dimensional Analysis of M-Estimators with Decomposable Regularizers (Haoming Jiang)
- (SRL) A Proximal-Gradient Homotopy Method for the Sparse Least-Squares Problem
- Randomized sketches for kernels: Fast and optimal non-parametric regression
- (DL) Deep Learning without Poor Local Minima (Yuanshuo Zhao)
- (NO) No Spurious Local Minima in Nonconvex Low Rank Problems: A Unified Geometric Analysis
- (DL) Generalization and Equilibrium in Generative Adversarial Nets
- (NO) Streaming PCA: Matching Matrix Bernstein and Near-Optimal Finite Sample Guarantees for Oja's Algorithm (Minshuo Chen)
- (DL) Spurious Local Minima are Common in Two-Layer ReLU Neural Networks
- (NO) Efficient Algorithms for Large-scale Generalized Eigenvector Computation and Canonical Correlation Analysis
- (NO, LRR) Symmetry, Saddle Points and Global Geometry of Nonconvex Matrix Factorization (Xingguo Li)
- (NO, LRR) Online Multiview Learning: Dropping Convexity for Better Efficiency (Zhehui Chen)
- (DL) Deep Hyperspherical Learning (Weiyang Liu)
- (LRR) Rates of Convergence of Spectral Methods for Graphon Estimation
- (NO, LRR) Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees
- (NO, LRR) Complexity Theoretic Lower Bounds for Sparse Principal Component Detection
- (SRL) Statistical Limits of Convex Relaxations
- (NO, LRR) Weak Convergence Analysis of Nonconvex Stochastic Optimization with Polyak's Momentum (Tianyi Liu)
- (DL) Gradient Descent Learns One-hidden-layer CNN: Don't be Afraid of Spurious Local Minima.

- (NO) First-order Methods Almost Always Avoid Saddle Points
- (DL) When is a Convolutional Filter Easy To Learn?
- A Markov Chain Theory Approach to Characterizing the Minimax Optimality of Stochastic Gradient Descent
- (DL) Generalization and Equilibrium in Generative Adversarial Nets (GANs)
- (NO) On the Optimization Landscape of Tensor Decompositions
- (DL) Gradient Descent Learns Linear Dynamical Systems
- (DL, NO) Weight Sharing is Crucial to Successful Optimization
- (DL) Composite Functional Gradient Learning of Generative Adversarial Models