Rajiv Khanna

rajivak@berkeley.edu
+1 512 947 6729

SKILLS

Programming: Python (Use extensively), Matlab, R, C (Have used)
Quantitative: Discrete/continuous optimization, Bayesian Statistics, Approximate Inference.

EDUCATION

University of Texas at Austin, 2018
PhD. (4.0/4.0) Electrical and Computer Engineering

- Thesis on sparsity constrained optimization in classical and Bayesian settings.

Indian Institute of Technology, 2008
MS. (9.67/10) Computer Science and Engineering

- Graduated top of class, awarded "Most Outstanding Student" of the graduating batch.

National Institute of Technology Jallandhar, 2006
BS. (83/100) Computer Science and Engineering

- Graduated in top three of class

AWARDS

- Best Paper Award in the NeurIPS 2020 (top 3 out of over 9400 paper submissions). NeurIPS is the biggest top tier machine learning conference of the year
- Simons-Berkeley Fellowship at UC Berkeley. Awarded to promising early career machine learning researchers based on their recent publications
- Phillips Scholarship at IIT Bombay

EXPERIENCE

Postdoc at Department of Statistics at UC Berkeley **(Current)**

- Beyond worst case analysis of the column subset selection problem; novel results showing that previously known lower bounds only occur in rare corner cases. Won Best paper award at NeurIPS 2020.
- Deep learning: Proposed a metric we call Boundary Thickness that is a generalization of the classical margin; theoretically showed a data augmentation technique called mixup optimizes for the proposed metric, and empirically showed that this metric correlates very well with generalization.
- Adversarial neural network training: Showed that adversarially trained neural networks on the ImageNet dataset transfer better to other datasets as opposed to non-adversarially trained ones. Also showed the former contains human-interpretable information and retains shapes instead of textures.
- Bayesian coresets: Research on use of optimization based algorithms for identifying a small representative sample of an otherwise large dataset. Provided new theoretical guarantees and state-of-the-art empirical performance.

- Ongoing work on obtaining algorithmic generalization bounds, and on speeding up the evaluation of influence functions for interpretability.

Simons Institute for Theory of Computing at UC Berkeley Fall 2018
Research Fellow

- Engaged in a semester long focused group on Foundations of Data Science as a Research Fellow to explore theoretical foundations of data science with several eminent researchers from all over the world. Research Fellows are selected based on past relevant research and an interview.

UT Austin
PhD Student

- Research on intersection of continuous and discrete optimization. Showed that a greedy algorithm for feature selection will give constant factor approximation, if certain conditions are satisfied. This is because feature selection satisfies a generalization of submodularity, that we call weak submodularity. Also obtained auxiliary results for other variants of the classic greedy algorithm. Published series of papers in various top tier machine learning conferences and Annals of Statistics (journal)
- Designed a new sparsity-introducing Bayesian prior by the use of information projection. Obtained Theoretical guarantees for the prior, and showed state-of-the art empirical quantitative performance on fMRI Brain scan applications. Research was in collaboration with a neuroscientist who also validated the qualitative aspects of the obtained results. Published series of papers in various top tier machine learning conferences
- New analysis and theoretical guarantees for Frank-Wolfe and Iterative Hard Thresholding optimization algorithms. Published series of papers in various top tier machine learning conferences
- Research on interpretability in machine learning to address "which training data points are responsible for making given test predictions". New theoretical guarantees for the algorithm Sequential Bayesian Quadrature. Published series of papers in various top tier machine learning conferences.

ETH Zurich Summer 2015
Visiting Researcher

- Explored a unified analysis of greedy atomic optimization methods (Frank-Wolfe algorithm and Matching Pursuit). Promising initial theoretical and empirical results. Mentored a junior PhD student who picked up my project after I left. Published the results in top tier ML conferences.

Microsoft Research Summer 2014
Research Intern

- Worked on a text mining problem that involves representation of words as vectors which can then be used in various ways such as features for a prediction problem.

LinkedIn Summer 2013
Software Engineering Intern

- Explored inclusion of online features within limitations of the existing Ad serving infrastructure. Proposals showed lifts, slated for bucket testing.

Yahoo! Labs Jul 2008 – Jul 2012

Research Engineer

- Web scale click prediction, recommendation systems, Modeling skewed data, Information corroboration. My role was of a liaison between science and engineering teams. I worked with computer scientists (CS/Statistics PhD holders) to develop web scale prototypes for a problem. Then, I collaborated with engineering teams to push the technology into production. Often it was an iterative process that required a tradeoff between the best model and a feasible one depending on engineering constraints. Technologies include Hadoop (Java), C, R, Matlab, Perl, Python. Click prediction models were bucket-tested and pushed into production, in addition to a publication at a top tier conference. Similarly, recommendation systems were pushed into the Yahoo! Frontpage. Information corroboration models were pushed into Yahoo! Local pages for populating various properties (e.g. Hours of Operations) of local businesses.
- Academic service: Program Committee/ Reviewer for various machine learning conferences: AISTATS 2020, CVPR 2020, ICML 2020, AAAI 2020, ICLR 2020, NIPS 2019, AAAI 2019, CVPR 2019, ICML 2019, ICCV 2019, AISTATS 2019, NIPS 2018, ICML 2018, NIPS 2017, ICML 2017, NIPS 2016, WWW 2017, Workshop on Advances in Approx. Bayesian Inference 2015 / 2016 / 2017 / 2018.

PUBLICATIONS

- Bayesian Coresets: An Optimization Perspective. (submitted to AISTATS 2021)
- Adversarially-trained deep nets transfer better. (ICLR 2021)
- Improved guarantees and a multiple-descent curve for the CSSP and the Nystrom method. (NeurIPS 2020)
- Boundary thickness and robustness in learning models. (NeurIPS 2020)
- Learning Sparse Distributions using Iterative Hard Thresholding. NIPS 2019
- Interpreting Black Box Predictions using Fisher Kernels. AISTATS 2019
- Boosting Black Box Variational Inference. NIPS 2018
- Restricted Strong Convexity implies Weak Submodularity. Accepted to Annals of Statistics 2018 (a shorter version appeared in NIPS 2016 Workshop on Learning in High Dimensions with Structure).
- Provable Accelerated Iterative Hard Thresholding. AISTATS 2018
- Boosting Variational Inference: An Optimization Perspective. AISTATS 2018 (a shorter version appeared at NIPS 2017 workshop on Approx. Inference)
- Co-regularized Monotone Retargeting for Semi-supervised LeTOR. SDM 2018
- On Approximation Guarantees for Greedy Low Rank Optimization. ICML 2017
- Scalable Greedy Support Selection via Weak Submodularity. AISTATS 2017
- Information Projection and Approximate Inference for Structured Sparse Variables. AISTATS 2017
- A Unified Analysis of Frank Wolfe and Matching Pursuit. AISTATS 2017
- Pursuits in Structured Non-Convex Matrix Factorizations. Arxiv Report.
- A Deflation Method for Structured Probabilistic PCA. SDM 2017
- Examples are not Enough, Learn to Criticize! Criticism for Interpretability. NIPS 2016
- Towards a Better Understanding of Predict and Count Models. Arxiv report.
- Sparse Submodular Probabilistic PCA. AISTATS 2015
- A Deflation Method for Probabilistic PCA. NIPS 2015 Workshop on Advances in Approximate Bayesian Inference.
- On Prior Distributions and Approximate Inference for Structured Variables. NIPS 2014
- Parallel Matrix Factorization for Binary Response. IEEE BigData 2013

- Estimating Rates of Rare Events with Multiple Hierarchies through Scalable Log-linear Models. KDD 2010
- Translating Relevance Scores to Probabilities for Contextual Advertising. CIKM 2009
- Structured Learning for Non-Smooth Ranking Losses. KDD 2008