NIKHYL BRYON ARAGAM

Pittsburgh, PA • naragam@cs.cmu.edu • www.cs.cmu.edu/~naragam/

CURRENT POSITION

Project Scientist, Machine Learning Department, Carnegie Mellon University — Sep 2016–present

EDUCATION

Doctor of Philosophy in Statistics, University of California, Los Angeles — Sep 2015

Thesis: Structure Learning of Linear Bayesian Networks in High-Dimensions

Advisor: Qing Zhou

Master of Arts in Applied Mathematics, University of California, Los Angeles — Dec 2010

Bachelor of Science with Honors in Mathematics, University of Tennessee, Knoxville — May 2008

Thesis: Volume comparison in Lorentz manifolds with integral bounds

RESEARCH INTERESTS

- High-dimensional statistics and machine learning
- Unsupervised and semi-supervised learning
- Network inference and graphical models
- Computational biology and precision medicine
- Computational statistics and distributed learning

RECOGNITION AND AWARDS

Young Academics Scholarship — Jul 2017
UseR!2017 Conference

Most Promising Theoretical Statistician — Jun 2015
UCLA

Dissertation Year Fellowship — Apr 2014
UCLA

Teaching Assistant Coordinator — Apr 2014
UCLA

National Science Foundation Graduate Research Fellowship — Jun 2010

John H. Barrett Prize for outstanding undergraduate mathematics major — May 2008
University of Tennessee

Chancellor's Honors Program Undergraduate Research Grant — Dec 2007
University of Tennessee

First Place, Allen Medal Mathematics Competition — May 2005
University of Tennessee

PUBLICATIONS AND PREPRINTS

[1] Concave penalized estimation of sparse Gaussian Bayesian networks.
B. Aragam and Q. Zhou.
The Journal of Machine Learning Research, 16:2273–2328, 2015.

[2] Learning large-scale Bayesian networks with the sparsebn package.
B. Aragam, J. Gu, and Q. Zhou.
Journal of Statistical Software, to appear. arXiv:1703.04025, 2017.

[3] Learning high-dimensional DAGs: Provable statistical guarantees and scalable approximation.
B. Aragam, J. Gu, A. A. Amini, and Q. Zhou.
NIPS Workshop on Advances in Modeling and Learning Interactions from Complex Data. 2017.

[4] Precision Lasso: Accounting for correlations and linear dependencies in high-dimensional genomic data.
H. Wang, B. Lengerich, **B. Aragam**, and E. P. Xing.
Bioinformatics, to appear. 2018.

[5] Personalized regression enables sample-specific pan-cancer analysis.
B. Lengerich, **B. Aragam**, and E. P. Xing.
Bioinformatics, Volume 34, Issue 13, 1 July 2018, Pages i178–i186.
Also appears in the *26th Conference on Intelligent Systems for Molecular Biology.*

[6] Variable selection in heterogeneous datasets: A truncated-rank sparse linear mixed model with applications to genome-wide association studies.
H. Wang, **B. Aragam**, and E. P. Xing.
Methods, Volume 145, 2–9. 2018.

[7] DAGs with NO TEARS: Continuous optimization for structure learning.
X. Zheng, **B. Aragam**, P. Ravikumar, and E. P. Xing.
NIPS 2018 **(spotlight)**, to appear. arXiv:1803.01422, 2018.

[8] Sample complexity of nonparametric semi-supervised learning.
C. Dan, L. Liu, **B. Aragam**, P. Ravikumar, and E. P. Xing.
NIPS 2018, to appear. arXiv:1809.03073, 2018.

[9] Learning directed acyclic graphs with penalized neighbourhood regression.
B. Aragam, A. A. Amini, and Q. Zhou.
Submitted. arXiv:1511.08963, 2015.

[10] Partial correlation graphs and the neighborhood lattice.
A. A. Amini, **B. Aragam**, and Q. Zhou.
Submitted. arXiv:1711.00991, 2017.

[11] Identifiability of nonparametric mixture models and Bayes optimal clustering.
B. Aragam, C. Dan, P. Ravikumar, and E. P. Xing.
Under review. arXiv:1802.04397, 2018.

[12] Fault tolerance in iterative-convergent machine learning.
A. Qiao, **B. Aragam**, B. Zhang, and E. P. Xing.
Under review. arXiv:1810.07354, 2018.

[13] Counterfactual utility and counterfactual curves for explanation.
D. Inouye, L. Liu, J. Kim, **B. Aragam**, and P. Ravikumar.
Under review. 2018.

PRESENTATIONS

**Invited* †*Sponsored (e.g. travel award)*

[1] *Personalized regression. Data science seminar, Hulu, November 2018, Santa Monica, CA.

[2] *,†Identifiability of nonparametric mixture models, clustering, and semi-supervised learning. Statistics seminar, UCLA, November 2018, Los Angeles, CA.

[3] *,†Identifiability of nonparametric mixture models, clustering, and semi-supervised learning. Statistics seminar, Oregon State University, October 2018, Corvallis, OR.

[4] *Identifiability of nonparametric mixture models, clustering, and semi-supervised learning. Machine learning lunch, Carnegie Mellon University, October 2018, Pittsburgh, PA.

[5] Identifiability of nonparametric mixture models and Bayes optimal clustering. Statistics and machine learning seminar, Carnegie Mellon University, February 2018, Pittsburgh, PA.

[6] *Learning high-dimensional Bayesian networks on your laptop with the sparsebn package. Center for Causal Discovery, University of Pittsburgh, February 2018, Pittsburgh, PA.

[7] †Collaborative Development in R: A Case Study with the sparsebn package. User!2017, July 2017, Brussels, Belgium.

[8] Learning high-dimensional structural equation models. Statistics and machine learning seminar, Carnegie Mellon University, February 2017, Pittsburgh, PA.

[9] *Reconstructing Biological Networks with Structural Equations: A New Machine Learning Approach to Better Understanding Drug Abuse. Center for Drug Abuse Research, Carnegie Mellon University, December 2016, Pittsburgh, PA.

[10] *Learning Graphical Models with Nonconvex Empirical Risk Minimization: Algorithms and theory. Machine learning lunch, Carnegie Mellon University, October 2016, Pittsburgh, PA.

[11] *Estimating high-dimensional Bayesian networks: Algorithms and theory. Symposium on *Sparse Modeling and Its Applications*, UCLA, January 2016, Los Angeles, CA.

[12] †Concave penalized estimation of sparse Gaussian Bayesian networks: Algorithms and theory. AMS Fall Western Sectional Meeting, October 24-25, 2015, Fullerton, CA.

[13] †Concave penalized estimation of sparse Gaussian Bayesian networks. Joint Statistical Meetings, August 2-7, 2014, Boston, MA.

[14] †Volume comparison in Lorentz manifolds with integral bounds. Joint Mathematics Meetings, January 6-9, 2008, San Diego, CA.

SOFTWARE

Precision Lasso (GitHub link): A variant of the Lasso designed to adapt to and account for correlations and dependencies in high-dimensional data.

Personalized regression (GitHub link): Python code for learning sample-specific, personalized regression models.

`sparsebn` package for R (CRAN link): Comprehensive package for learning large-scale Bayesian networks based on sparse regularization.

`ccdr` package for R (link): Software for learning Gaussian BNs with thousands of variables.

TEACHING

Instructor

Data Analysis Project Preparation (Machine Learning 10-821, Fall 2017)

Overall evaluation: 4.5/5

Introduction to Statistical Reasoning (Statistics 10, Fall 2015, Winter 2016, and Spring 2016)

Overall evaluation: 8/9

Teaching College Statistics (Statistics 495A, Winter 2015)

Overall evaluation: 8/9

Teaching Assistant

Introduction to Probability (Statistics 100A, Spring 2014)

Overall evaluation: 8/9

Introduction to Design and Analysis of Experiments (Statistics 101B, Winter 2014)

Overall evaluation: 7.75/9

Introduction to Computational Statistics with R (Statistics 102A, Fall 2013)

Overall evaluation: 9/9

Introduction to Computing for Social Sciences and Humanities (Program in Computing 3, Spring 2010)

Overall evaluation: 9/9

Introduction to C++ Programming (Program in Computing 10A, Fall 2009 and Winter 2010)

Overall evaluation: 9/9 (Winter 2010)
Overall evaluation: 8/9 (Fall 2009)

Guest Lecturer

Introduction to Machine Learning (Machine Learning 10-701, Spring 2017)
Probabilistic Graphical Models (Machine Learning 10-708, Spring 2017)
Theoretical Statistics (Statistics 200C, Winter 2016)

SERVICE

High-Dimensional Statistics and Machine Learning Reading Group Fall 2015
Organizer

Teaching Assistant Coordinator Fall 2014
Instructor and organizer

Graduate Student Orientation Fall 2014
Organizer