Subhabrata Majumdar

Address:
1112 8th Street SE Apt 7
Minneapolis, MN 55414

Phone: (612) 806-9996
e-mail: zoom.subha@gmail.com
Web: http://stat.umn.edu/~majum010

Summary

- PhD in Statistics from University of Minnesota;
- Area of research: dimension reduction and variable selection methods. Expert level knowledge in these areas;
- Strong conceptual background in statistics and machine learning techniques;
- Technical skills: R, Matlab, Stata (strong), C/C++ (proficient), Python, SQL, HTML (basic);
- Experienced in working as part of a team in industry, as well as interdisciplinary academic setup.

Experience

- **University of Florida Informatics Institute**, Gainesville, FL; Postdoctoral researcher, starting July 2017;
  Shall work with Prof. George Michailidis on developing statistics/ ML methods for the modelling of high-dimensional omics (e.g. genomics, proteomics, metabolomics) data;

- **IBM Research**, Yorktown Heights, NY; Research intern, May 2016 – Aug 2016;
  As an IBM Social Good fellow, collaborated with scientists in the data science group and the Cary Institute of Ecosystem Studies to mine biological and ecological data and devise cognitive algorithms to determine potential new primate carriers of the Zika virus. As part of this, also built a R-shiny app to visualize the results in an interactive fashion.

- **University of Minnesota - Twin Cities**, Minneapolis, MN; Research assistant, Jan 2015 – May 2016;
  Worked as a research assistant at the Department of Computer Science and Engineering, in the $10 million interdisciplinary project on developing data-driven approaches to understand climate change.

- **Santander Consumer USA**, Dallas, TX; Data Science intern, May 2015 – Aug 2015;
  Worked with the Statistical Analysis team on implementing new machine learning methods in Loss Forecast Score prediction. Achieved 3% improvement in Area Under Curve metric over current model in practice by implementing a gradient boosting model.

- **Data Science for Social Good fellowship**, University of Chicago. June 2014 – Aug 2014;
  Worked in a team that collaborated with the Chicago Department of Public Health in building a predictive model for lead poisoning prevention in the city of Chicago using large amounts of data on previous testing and house inspections. Our random forest-based model achieved 70% accuracy in out-of-sample prediction.

- **National Marrow Donor Program**, Minneapolis, MN; Statistician Intern, June 2013 – Aug 2013;
  Collaborated with scientists at the Bioinformatics division and Prof. Snigdhansu Chatterjee of the School of Statistics at my university in designing a spatial algorithm for data-driven marrow donor recruitment for Leukemia patients with rare alleles.

Awards

Best Student Paper Award, International Indian Statistical Association Conference, Corvallis, OR, 2016;
Martin Award in Statistics, University of Minnesota School of Statistics, 2016 – 2017;
Interdisciplinary Doctoral Fellowship, University of Minnesota Graduate School, 2016 – 2017;
University of Minnesota School of Statistics travel award, 2013 – 2017;
5th International Workshop on Climate Informatics travel award, 2015;
Debesh-Kamal Scholarship for Higher Studies Abroad, Ramakrishna Mission Institute of Culture, Kolkata, India, 2012;
KVPY national fellowship, Department of Science and Technology, Govt. of India, 2008 – 2012;
National scholar, National Council of Educational Research and Training, Govt. of India, 2005 – 2008.

Qualifications

PhD Statistics, University of Minnesota - Twin Cities, 2017;
Advisor: Snigdhansu Chatterjee

M.S. Statistics, Indian Statistical Institute, 2012. (Specialization: Actuarial science and Genetics)
Masters thesis: Adjusting for Treatment Effects in Studies of Quantitative Traits

B.S. Statistics, Indian Statistical Institute, 2010.

Research interests

Theory: Robust variable selection and dimension reduction using multivariate quantiles and statistical depth functions. High-dimensional statistical methods.

Application: Statistical computing, machine learning algorithms and high-dimensional models in chemometrics, public health and human genomics.

Publications

- **Majumdar, S.** and Basak, S. C. Exploring intrinsic dimensionality of chemical spaces for robust QSAR model development: A comparison of several statistical approaches. *Current Computer-Aided Drug Design*, **2016**, *12*, 294-301;
- (Book chapter) Basak, S.C. and **Majumdar, S.** Current landscape of hierarchical QSAR modeling and its applications: Some comments on the importance of mathematical descriptors as well as rigorous statistical methods of model building and validation. In: *Advances in Mathematical Chemistry and Applications: vol. 1*, **2016**, 251-281;
- **Majumdar, S.**, Dietz, L. and Chatterjee, S. Identifying Driving Factors Behind Indian Monsoon Precipitation using Model Selection based on Data Depth. *Proceedings of the Fifth International Workshop on Climate Informatics: CI 2015. J. G. Dy, J. Emile-Geay, V. Lakshmanan, Y. Liu (Eds.)*, ISBN: 978-0-9973548-0-5, **2015**;
- Basak, S.C. and **Majumdar, S.** Prediction of Mutagenicity of Chemicals from Their Calculated Molecular Descriptors: A Case Study with Structurally Homogeneous versus Diverse Datasets. *Current Computer-Aided Drug Design*, **2015**, *11*, 117–123;
- With 9 authors. Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. *Proceedings of the 21st ACM SIGKDD conference on Knowledge Discovery and Data Mining, Sydney, Australia*, **2015**, 2039–2047;
- Mukherjee, U., **Majumdar, S.** and Chatterjee, S. Fast and Robust Supervised Learning in High Dimensions Using the Geometry of the Data. In: *Advances in Data Mining: Applications and Theoretical Aspects*, ser. *Lecture Notes in Computer Science*, **2015**, *9165*, 109–123;
- **Majumdar, S.**, Basak, S.C. and Grunwald, G.D. Adapting Interrelated Two-Way Clustering Method for Quantitative Structure-Activity Relationship (QSAR) Modeling of Mutagenicity/ Non-Mutagenicity of a Diverse Set of Chemicals. *Current Computer-Aided Drug Design*, **2013**, *9*, 463–471.

Conference presentations

- (Invited) American Statistical Association Twin Cities Chapter Spring Meeting, Minneapolis, MN (April 2017);
- ENAR Spring Meeting, Washington, DC (March 2017);
- Data Science for Social Good conference, Chicago, IL (Aug 2016);
- (Invited) International Indian Statistical Association conference, Corvallis, OR (Aug 2016);
- Joint Statistical Meetings: Chicago, IL (July 2016); Seattle, WA (Aug 2015); Boston, MA (Aug 2014);
- 9th International Triennial Calcutta Symposium, Kolkata, India, Dec 2015;
- (Invited) 5th International Workshop on Climate Informatics, Boulder, CO, Sep 2015;
- 20th ACM SIGKDD conference on Knowledge Discovery and Data Mining, New York City, NY: Aug 2014.