Tayler A. Blake

Columbus, OH
☎ *(740) 607-9508*
✉ *tayler.a.blake@gmail.com*
http://taylerablake.github.io

Education

August 2018 **Ph.D. Statistics**, *The Ohio State University*, Columbus, Ohio.
Dissertation: *Nonparametric Covariance Estimation for Longitudinal Data*
Advisor: Yoonkyung Lee

January 2010 **M.S. Statistics**, *The Ohio State University*, Columbus, Ohio.

May 2007 **B.A. Mathematics, Computer Science**, *Capital University*, Columbus, Ohio.

Experience

Senior Data Scientist

Information Control Company, *Columbus, OH.*

January 2017-present

Advanced Analytics consultant serving on teams in a Senior Data Scientist role; leads data science teams working alongside delivery team members and data engineers. Responsible for planning, designing, and executing analytical plans to address business questions for a variety of clients.

Internally at ICC, contributes to Advanced Analytics Development Seminars and heads the Advanced Analytics Journal Club, which meets biweekly to read and discuss the most current literature on developments in the field of data science, statistics, and machine learning.

Machine Learning Specialist

Pillar Technology, *Columbus, OH.*

February 2016-December 2016

Machine learning specialist on a team responsible for designing and developing adaptive software for an IoT product to be embedded in a line of luxury vehicles. Responsible for setting expectations given the available resources such as data, available in-device storage space, and computational constraints with both internal team and with the client.

Challenged to design and implement predictive algorithms in the absence of data a priori, making the balance between bias and stability of statistical estimates and predictions a crucial consideration, in addition to how imperfect statistical performance may impact user experience. Automatic variable and model selection are key techniques in ensuring that algorithms adaptively learn as more time is spent using the product. Follows works from end to end, from writing exploratory analysis to production-level implementation alongside developers using Agile best practices, with a strong emphasis on test-driven development in both Java and Scala.

Spearheaded components of project planning pertaining to data and the corresponding infrastructure necessary for data collection, storage, and modeling at scale. Lead efforts to establish client trust in data modeling and algorithms, a new operating space for Pillar.

Data Scientist

Crosschx, Inc., *Columbus, OH.*

November 2015-February 2016

Hired as the first data science employee in the company; sought to develop a data pipeline to separate high level analysis from raw production data, emphasizing scalability, reproducibility, and data integrity while simultaneously extracting as much reliable knowledge as possible from the existing infrastructure.

Utilized Apache Spark to aggregate, summarize, and visualize data of moderately large dimension, with special focus on data quality assurance and anomaly detection. Built notebooks with Apache Zeppelin to initiate cross-team consumption of reproducible reporting of core product KPIs, with focus on providing insights using clean, trustworthy data.

Built robust, parsimonious models to forecast patient wait times using generalized additive models in tangent with penalization techniques for variable selection and imposi smoothness on specific model components. Used variable importance measures as initial screen for identifying variables and engineered features having strong predictive relationships with the response variable.

Data Scientist

Crosschx, Inc., *Columbus, OH.*

November 2015-February 2016

Hired as the first data science employee in the company; sought to develop a data pipeline to separate high level analysis from raw production data, emphasizing scalability, reproducibility, and data integrity while simultaneously extracting as much reliable knowledge as possible from the existing infrastructure.

Utilized Apache Spark to aggregate, summarize, and visualize data of moderately large dimension, with special focus on data quality assurance and anomaly detection. Built notebooks with Apache Zeppelin to initiate cross-team consumption of reproducible reporting of core product KPIs, with focus on providing insights using clean, trustworthy data.

Built robust, parsimonious models to forecast patient wait times using generalized additive models in tangent with penalization techniques for variable selection and imposi smoothness on specific model components. Used variable importance measures as initial screen for identifying variables and engineered features having strong predictive relationships with the response variable.

Data Scientist

Store Development, Starbucks Coffee Company, *Seattle, WA.*

May 2014 - present

Data scientist on a team serving the company in market planning and strategy. Utilized a variety of statistical and machine learning methods, both supervised and unsupervised such as penalized regression and classification, generalized linear models, ensemble methods including boosting, bagging, and random forests, with applications of the latter to both classification and clustering.

Estimated causal impact of a variety of interventions, such as competitor store openings, pricing changes, or new product launches on store performance using Bayesian structural time series models and summarizes comparisons between this methodology and prior approaches, including the traditional difference-in-differences estimators.

Cleaned and analyzed heterogeneous data from several different sources having both potential spatial and temporal components which pose challenges in aggregation and summarization.

Adjunct Statistics Instructor

Department of Mathematics, Columbus State Community College, *Columbus, Ohio.*

August 2013 - January 2014

Lecturer for an introductory statistics course for undergraduate students, Statistics 1350. Non-instruction responsibilities included curriculum and assessment development, including lecture presentations and online learning tools and learning assessments.

Graduate Research Assistant

Comprehensive Cancer Center, The Ohio State University, *Columbus, Ohio.*

October 2011 - August 2012

My responsibilities included analysis of large microarray data sets, in particular utilizing data mining and dimension reduction techniques to find genetic markers in leukemic patients, sharing results and collaborating with medical professionals to both direct further laboratory investigation as well as further statistical investigation.

Graduate Research Assistant

Department of Statistics, The Ohio State University, *Columbus, Ohio.*

June 2011 - October 2011

Responsibilities included reformulation of undergraduate introductory statistics curriculum in preparation for the university conversion from quarters to semesters, primarily writing course materials including notes and presentation slides as well as developing online learning tools.

Graduate Research Assistant

Nationwide Center for Advanced Customer Insights, The Ohio State University, Nationwide Insurance, *Columbus, Ohio.*

June 2010 - June 2011

My responsibilities included work on projects modeling agency behavior using high dimensional demographic and marketing data, specifically by modeling survival times using Cox proportional hazards models with both static and time-varying coefficients. Modeling was done with an emphasis on building parsimonious, interpretable models. I was responsible for presenting results in a corporate setting to high level company executives with motive to encourage and motivate business decisions and action.

Graduate Teaching Assistant

Department of Statistics, The Ohio State University, *Columbus, Ohio.*

August 2007 - May 2014

Lecturer for multiple undergraduate statistics courses, including Statistics 145, 427, and 3470.

Lab instructor for multiple undergraduate statistics courses, including Statistics 133, 145, 245, and 2480.

Research Interests

My research interests are centered in methods for Statistical Learning, particularly nonparametric function estimation techniques and reproducing kernel Hilbert space methods. The focus of my dissertation work is nonparametric covariance estimation for longitudinal data. Our proposed framework permits unconstrained covariance estimation and accommodates imbalanced longitudinal data without prerequisite data imputation. The estimation procedure relies on a specific decomposition of the covariance matrix which re-renders the problem as the estimation of a particular regression model. This recasting grants access to the bounty of penalized regression techniques for use in covariance estimation.

Computing Skills

Statistical/Analysis Software: R, ArcGIS, Alteryx, Apache Spark

Programming Languages: MySQL, Git, elementary experience in Scala, Java, Python, HTML

Proficient in Markdown, LaTeX, MS Office, Project management and issue-tracking software including Jira, Confluence

Presentations

August 2017 Joint Statistical Meetings *Nonparametric Covariance Estimation for Longitudinal Data via Tensor Product Smoothing*

October 2016 Wards Auto User Experience Conference *The Machines Are Coming: Will Algorithms Replace Designers in the UX World?*

August 2012 Joint Statistical Meetings *Nonparametric Covariance Estimation for Functional Data with Shrinkage Toward Stationary Models*

October 2011 Ohio State Data Mining and Statistical Learning Discussion Group

Discussion of "A nonparametric view of network models and Newman-Girvan and other modularities," Peter J. Bickel and Aiyou Chen

October 2010 Ohio State Data Mining and Statistical Learning Discussion Group

Discussion of "The 'Independent Components' of Natural Scenes are Edge Filters," Anthony J. Bell and Terrance J. Sejnowski

Hobbies and Interests

Competitive distance running, cooking, photography, practicing yoga

References

Available upon Request