DAVID LAING

LinkedIn: https://www.linkedin.com/in/laingdk

Website: https://laingdk.github.io/

davidkendalllaing@gmail.com

1-778-834-1745

Summary

I used to teach writing, and I was set on a life working with words. But as I gained experience, I realized that I could find more interesting and urgent problems by working with numbers. So, over the past year I've been deepening my knowledge of statistics and computer science in the Master of Data Science program at UBC.

Education

Master of Data Science, 2016 – 2017
University of British Columbia
GPA: 4.0

Honours Bachelor of Arts & Science, 2010 – 2014
McMaster University
GPA: 3.7

Technical experience

Machine learning
- Classified handwritten digits using convolutional neural networks, with Keras (Tensorflow backend) and EC2 on Amazon Web Services.
- Built and tested recommender systems using collaborative filtering and hybrid models, using sklearn.
- Used forward selection, backward selection, and recursive feature elimination to identify relevant variables from among hundreds, using sklearn and bestglm (R).
- Used regularization, model averaging, and Bayesian methods to prevent overfitting.
- Implemented k-means, k-means++, and k-medoids for clustering unlabelled datasets, using R.
- Implemented the expectation-maximization algorithm for created mixed models, using R.
- Used cross-validation to select among multiple models, using sklearn.
- Compressed and reconstructed images of faces via principal-component analysis, using R and sklearn.
- Used factor analysis to discover latent variables driving variation in observed data.
- Used Latent Dirichlet Allocation (LDA) for topic modelling, using Python.
- Created Markov chain models of words and natural language.

Statistics
- Fit and interpreted linear models, mixed effects models, generalized linear models, generalized additive models, LOESS models, splines, and robust models.
- Used multiple imputation to deal with missing data.
- Used Markov-Chain Monte Carlo (MCMC) for Bayesian models of baseball data, using rjags (R).
- Performed classical hypothesis tests, exact tests, and permutation tests on many datasets.
- Applied Bonferroni and Benjamini-Hochberg corrections to account for multiple comparisons.
- Decomposed, analysed, and forecasted time series data.

Programming
- Built and automated the testing of R and Python packages, with continuous integration, using TravisCI, pytest (Python), and testthat (R).
- Used dynamic programming to write a seam-carving algorithm for image dimension reduction.
- Implemented standard searches of lists and graphs (insertion, binary, merge, depth-first, breadth-first).
- Analyzed time and space complexity of simple algorithms.
- Wrote and used objects and classes in Python and R.

Data visualization
- Created interactive Shiny apps to explore encodings and subsets of US crime data, as well as demographics of US colleges.
- Created static visualizations of network data (arcplots, network graphs, adjacency matrices), such as Twitter hashtag networks and character networks in Jane Austen novels.
- Created static visualizations of spatial data (choropleth maps, dot maps), such as crime rates in Vancouver and earthquakes across the west coast of North America.

Web and cloud computing
- Scraped data from static web-pages.
- Conducted analysis of data queried from the Twitter API, using tweepy (Python) and twitteR (R).
- Analysed historical word frequencies in the Google n-grams dataset, using Simple Storage Service (S3) and Elastic Map Reduce (EMR) on Amazon Web Services.

Databases
- Queried from SQLite databases and a Google BigQuery database.
- Designed database schemas in SQLite.
- Queried from XML databases, using Xpath.
- Applied k-anonymity and l-diversity to protect against privacy threats.

Data wrangling
- Used dplyr to manipulate dataframes, using window functions, grouping functions, and mutations.
- Converted character representations of date-times into POSIX date-times.
- Used standard joins (inner join, left join, full join, etc) to combine datasets with common variables.
- Used regular expressions to clean messy text data.
- Converted JSON and XML data to R lists to enable targeted queries.

Selected Professional History

(current) Data Science Capstone, Summer 2017
UBC Centre for Teaching, Learning, and Technology

- In online learning, students generate reams of data. To help instructors use this data effectively, my teammates and I are building an interactive dashboard for UBC's edX platform.

Bryan Turner Intern in Business Ethics, Fall 2016
Ethical Systems

- Academic research in business ethics is useful and important, but often inaccessible. So, I wrote and edited web documents to help business leaders make the most of the latest science.

Communications Lecturer, 2015 – 2016
Niagara College Canada

- I planned and delivered over 350 hours worth of lessons and workshops on writing and argumentation. I taught 150-170 students per semester.

Physics Research Assistant, Summer 2012
McMaster Physics Department

- The department wanted to administer undergraduate assignments through LON-CAPA, an open-source e-learning platform. So, I coded 2,000 physics problems for this platform, in Perl.

Interests

Reading
- Science, classics, sci-fi, fantasy

Music
- Hip-hop, funk, folk, alternative

Exercise
- Squash, running, windsurfing

Podcasts
- Philosophy, psychology