Eddie Yue

Phone: (773) 543-6233, Email: edyue@uchicago.edu

Personal GitHub:
- https://github.com/eddieyue

EDUCATION:

University of Chicago, Chicago, IL   09/2013—06/2015
Master of Science in Statistics - Computing and Optimization Track
- GPA:3.6, 6 PhD level courses, 5 CS courses out of the 12 classes completed so far

Imperial College London, United Kingdom   07/2010—07/2013
Bachelor of Science in Mathematics and Statistics
- GPA: 3.9/4.0, Dean's List, First Class Honor

PROGRAMMING / STATISTICAL SKILLS:
- Programming: Python, R, Matlab, SQL, Pig, Hive, Spark and Cassandra
- Data Science: Machine Learning, Pattern Recognition, Hadoop, Mapreduce, Data Munging
- Statistical Techniques: A/B Testing, Experiment Design, Predictive Modeling, Bayesian Modeling, Nonlinear Optimization, Generalized Linear Models

ANALYTICAL PROJECTS:

Projects on Amazon Web Services:
- Predict crime rate one week in advance and generate a heatmap on Google Maps by implementing a Poisson Process based on 7 years of crime data in Chicago.
- Analyze over 1,000,000 tweets around New Year's Day 2013 by fitting L2-logistic regression with stochastic gradient descent to find tweets that contain wishes.
- Come up with the parser to extract & clean text, fit a set of topics of 12,800 random Wikipedia articles, train Latent Dirichlet Allocation on PNAS dataset to find prominent words in each topic.
- Process over 10 gigabytes of genome data, construct a similarity matrix of genome for DNA distance computation, reconstruct the Cladogram by grouping the genomes.

Project on Google Cloud:
- Build a web-interface query application for severe weather for all US counties, store all of the historical data as a HDFS master data set, process real-time data though Spark, store the precomputed data in Cassandra to speed up web queries.

Project on Udacity.com:
- Analyze how weather affects New York City subway ridership, fit and find the best linear model with the hourly ridership as the response and with temperature, time and location as predictors.

RESEARCH EXPERIENCE:

Imperial College, London, UK   06/2012—07/2012
Researcher, cluster analysis for image restoration
- Implement cluster analysis algorithms: hierarchical clustering, k-mean clustering and bi-clustering.
- Compare the effectiveness of these three algorithms in restoring a distorted image. Find bi-clustering to be the most effective and k-mean clustering the least.

Southeast University, Nanjing, China   07/2012—09/2012
Research Assistant, data collection and sampling
- Collect and group multilayer data of over 1.6 million families' daily commute in Huzhou.
- Sample data from every town, county and community to cover all demographic characteristics.