Jiayun Dong

jiayun.dong@gmail.com

EDUCATION

Duke University

Ph.D. in Economics
Aug 2015 - May 2020

M.A. in Economics
Aug 2013 - May 2015

Fudan University, China

B.A., Finance
Aug 2009 - Jun 2013

SKILLS

Data Processing, Visualization & Programming

- R: dplyr, ggplot
- SQL
- Python
- Matlab
- Java

Machine Learning

- Decision trees
- Logistic regression
- kNN
- Support vector machine
- Neural networks
- Hierarchical clustering
- t-SNE

COURSEWORK

Statistics / Computer Science

- Machine Learning (A+)
- Bayesian Statistics (A-)
- Numerical Analysis (A)
- Data Structure & Algorithms(A)

Mathematics

- Real Analysis (A-)
- Probability (A)

Economics

- Econometrics I & II (A / A-)
- Demand Estimation (A-)

04.20.2020

WORK

Data Science Intern at Airbnb Experiences (June - Aug 2019)

- Improved host finding strategy by identifying Airbnb Experience categories that better match guests' interest in each regional market, and made recommendations to Operation teams.
- Newly acquired Experiences matching my recommendations were 9% more likely to be booked. The recommendations were based on collaborative filtering algorithm implemented by kNN.
- Defined a metric to measure the similarity in booking behavior among travel corridors (e.g. from UK to Paris vs. from US to Paris). Wrote Python script in data pipeline for automatic update.
- Created a self-service dashboard for regional managers to understand guest behavior in their markets.
- Hosted workshops for America, Europe and APAC Operation teams to introduce my dashboard and host acquisition recommendations.

PROJECTS

Racial Bias in the COMPAS Model (Working paper)

- Conducted variable importance analysis and found no evidence of racial bias in the COMPAS model used by the US court system, which countered the claim made by the ProPublica News Organization.
- Developed a classification algorithm to identify a set of well-performing (i.e., with loss below a threshold) logistic regression models: Sample a number of such models and use principle component analysis to identify the boundary of this set. Used a greedy search algorithm to identify a set of well-performing decision tree models.

Demand Estimation via Support Vector Machine

- Used SVM (with polynomial kernel) to predict consumer demand using an e-commerce data set. Testing error reduced by 20% by making a behavioral assumption that consumers paid attention only to the close substitutes of what they bought (compared to the discrete choice model commonly used in demand estimation in economics).
- Found the substitution structure by the implementing hierarchical agglomerative clustering algorithm on product features.

Analyzing & Visualizing Covid-19 Data (Personal project, Github)

- Visualized Covid-19 data to understand its development.
- Analyzed test positive rate and death rate, and found inadequate testing is one of the reasons of the recent slowdown in case growth in the US.
- Identified states of higher risks by jointly analyzing two metrics, test positive rates and number of tests per million population.