Ye **Wang**

Statistician · Machine Learning Practitioner

Department of Statistical Science, Duke University, Durham, NC, US

| eric.ye.wang@duke.edu | ericyewang.github.io | ericyewang | ericyewang

Education

Duke University — *Durham, NC, US*

Ph.D. in Statistics — *Aug. 2014 - PRESENT*

- Advisor: Dr. David Dunson.
- Coursework: Spatial Statistics, Nonparametric Methods, Generalized Linear Model.

Duke University — *Durham, NC, US*

M.S. in Statistical and Economical Modeling — *Aug. 2012 - May 2014*

- Coursework (Stat): Bayesian Statistics, Inference, Time Series, Linear Model.
- Coursework (Econ): Microeconomics, Macroeconomics, Game Theory, Finance.

HIT (Harbin Institute of Technology) — *Harbin, Heilongjiang, CN*

B.S. in Computational Mathematics — *Aug. 2008 - May 2012*

- Related Coursework: Numerical Analysis, C/C++, Data Structure, Database System.

Experience

IBM TJ Watson Research Center — *Yorktown Heights, NY, US*

Summer research intern — *May 2015 - Aug. 2015*

- Mentor: Dr. Youngdeok Hwang.
- We propose a hierarchical clustering model for large scale monitoring network data. To fit the model, we design a scalable EM algorithm that can be easily paralleled and implemented online. We also propose a new information criterion for model selection.
- Paper in preparation.

Department of Statistics, Duke University — *Durham, NC, US*

Teaching Assistant (Multiple Times) — *Aug. 2012 - PRESENT*

- Teaching labs.
- Lab website (Link).

Research & Publications

Electrostatic Gaussian Process (electroGP)

First Author

- Ye Wang, David Dunson, Probabilistic curve learning: *Coulomb repulsion and the electrostatic Gaussian process*, in *NIPS*, volume 28, pages 1729-1737, 2015.
- There is a lack of probabilistic methods that allow learning of the manifold along with the generative distribution of the observed data. The best attempt is the Gaussian process latent variable model, but identifiability issues lead to poor performance. We solve these issues by proposing a novel Columb repulsive process for locations of points on the manifold, inspired by physical models of electrostatic interactions among particles. Combining this process with a GP prior for the mapping function yields a novel electrostatic GP (electroGP) process.

Geometric Density Estimation (GEODE)

First Author

- Ye Wang, Antonio Canale, David Dunson, *Scalable geometric density estimation*, submitted to *AISTAT* 2016.
- It is standard to assume a low-dimensional structure in estimating a high-dimensional density. However, popular methods, such as probabilistic principal component analysis, scale poorly computationally. We introduce a novel empirical Bayes method that we term geometric density estimation (GEODE) and show that, with mild assumptions, the subspace spanned by the principal axes of the data is the MAP linear subspace under the proposed model. With these axes pre-computed using fast singular value decomposition, GEODE easily scales to high dimensional problems while providing uncertainty characterization. The model is also capable of imputing missing data and learning the true intrinsic dimension.

Honors & Awards

2015 **Student Travel Award**, Annual Conference on Neural Information Processing Systems (NIPS)
2014 **First Year Statistical Science Fellowship**, Department of Statistics, Duke University
2011 **Third Prize**, Matlab Application Programming Contest, GPUSolution Technology
2010 **National scholarship**, Ministry of Education of the People's Republic of China

Presentation

Department of Statistics, Duke University *Durham, NC, US*

Student Seminar Speaker *Sept. 2015*

- Title: *Probabilistic curve learning: Coulomb repulsion and the electrostatic Gaussian process.*

Skills

Programming R, Matlab, C/C++, LaTeX, SQL
Web HTML, CSS
Languages Mandarin, English