GEELON SO

Ph.D. Student, University of California, San Diego
Department of Computer Science and Engineering
9500 Gilman Drive, Mail Code 0404 · La Jolla, CA 92093

Phone: (773) 621-9265
Email: geelon@ucsd.edu
Website: https://geelon.github.io

RESEARCH

Machine learning theory, sequential decision making, stochastic analysis, optimization, geometry

EDUCATION

University of California, San Diego   La Jolla, CA
Ph.D. Student, Computer Science   *Sep 2019—present*
- Advisors: *Sanjoy Dasgupta, Yian Ma*
- Relevant courses: *Probability Theory, Stochastic Analysis, Differential Geometry, Computational Statistics, Unsupervised Learning, Continual Learning, Computational Neurobiology, Lattice Algorithms, Cybersecurity*

Columbia University   New York, NY
M.S. Computer Science   *May 2019*
- Advisor: *Daniel Hsu*
- Thesis: *Active learning with noise*
- Relevant courses: *Learning theory, Unsupervised Learning, Graph Theory, Information Theory, Algorithms through Geometric Lens, Deep Learning, Computer Networks, Privacy-Preserving Technologies*

The University of Chicago   Chicago, IL
B.S. Mathematics with Honors   *Jun 2017*
- *Advisor: Stuart Kurtz*
- Relevant courses: *Algebraic Geometry, Commutative Algebra, Algebraic Number Theory, Topology, Honors Algebra, Honors Analysis (real & functional analysis, measure theory), Markov Chains, Quantum Mechanics*

EXPERIENCE

Seekr (news search engine company)   Carlsbad, CA
Machine learning research and engineering intern   *Jun 2022—Jun 2023*
- Spearheaded the overhaul of the trending news labeling process; developed a human-in-the-loop system to enable rapid experimentation with diverse rules- and ML-methods for keyword extraction and labeling
- Engineered a data exploration tool to help developers gain intuition for the underlying news/text data; created modular system, streamlining new and iterative refinement of label extraction methods
- Built a user-friendly GUI for human evaluation of label quality
- The incorporation of the redesign resulted in significantly more expressive and meaningful topic labels

Home Partners of America (REIT company)   Chicago, IL
Data science and engineering intern   *Jun 2019—Aug 2019*
- Identified, planned and executed a high-impact project given limited resource and time constraints
- Streamlined ETL for the analytics team by providing API to automatically join, transform and aggregate data, while performing data quality checks with statistical guarantees; eliminated need to understand how the history of business decisions affects the correct way to join tables
- Designed backend to API to be easily adaptable to future changes in data collection/database

AumiPhyte Health (healthcare startup)   New York, NY
Machine learning consultant   *Feb 2018—Mar 2019*
  - Designed user-centric system and tools to analyze and process medical texts, presenting recommendations in a whitepaper that detailed motivation, implementation, rationale, and limitations
  - Implemented rules- and machine-learning based text processing methods using python and spaCy to generate label data for small set of medical texts; estimated 0.9 accuracy with 0.9 confidence
  - Developed GUI utilizing active learning techniques to help user rapidly label data

Options for Youth/University of Chicago (non-profit consulting course)   Chicago, IL
Non-profit consultant   *Mar 2016—Jun 2016*
  - Conducted quantitative/qualitative impact assessment for a local nonprofit by (i) researching and building a model to estimate the return on investment and (ii) interviewing stakeholders and recording impact; provided actionable best practices for using impact assessment to target donors

UChicago Math REU (research experience for undergrads)   Chicago, IL
Undergraduate researcher   *Summers 2014, 2015, 2016*
  - Studied foundations of mathematics, intuitionistic type theory and category theory; advised by Stuart Kurtz. Studied dynamical systems and quantum computing; advised by Clark Butler, Tori Akin, and Peter May

Chicago Summer School in Mathematics (analysis and topology)   Chicago, IL
Invited student   *Summers 2015, 2016*

TEACHING

The Institute for Emerging CORE Methods in Data Science   La Jolla, CA
Lecturer for the Foundations in Data Science high school summer program   *Jul 2023—Aug 2023*
  - Designed and presented 5 lectures/homework for a module on linear algebra; director: *Rajiv Ghandi*

University of California, San Diego   La Jolla, CA
Teaching assistant
  - Machine Learning, *Yian Ma*, *Fall 2022*; Probability and Statistics, *Sanjoy Dasgupta*, *Fall 2020*

Columbia University   New York, NY
Teaching assistant
  - Awarded fellowship; presented 15 hours of lectures on unsupervised learning techniques; designed 9 homework problems; taught over 100 hours during office hours and individual meetings
  - Unsupervised Learning, *Nakul Verma*, *Summer 2018*; Machine Learning, *Nakul Verma*, *Summer 2018*; Graph Theory, *Tim Sun*, *Spring 2018*; Geographic Information Systems, *Michael Parrott*, *Fall 2017*

SKILLS

Python, Pytorch, C, Haskell, Bash, Linux, GIS, SQL, technical writing, LaTeX

AWARDS

2020 UC San Diego Changemaker Challenge 1st Place
In the UC San Diego COVID-19 Contact Tracing Challenge, sponsored by XYO

2019 Andrew P. Kosoresow Memorial Award for Excellence in Teaching and Service
Awarded for outstanding contributions to teaching in the Department of Computer Science at Columbia University and exemplary service to the Department and its mission

SERVICE

Conference Reviewing
Reviewer for AISTATS *2022*, *2023*, *2024*; NeurIPS *2023*

UC San Diego Diversity Fellowship Committee　　　　　　　　　　　La Jolla, CA
Reviewer　　　　　　　　　　　　　　　　　　　　　　　　　　*Jan 2020, Jan 2021*

CSE/HDSI Ph.D. Visit Day　　　　　　　　　　　　　　　　　　　La Jolla, CA
Coordinator for AI/ML group　　　　　　　　　　　　　　　　　*Mar 2020, Mar 2021*

Google/UC San Diego ExploreCSR Mentorship Program　　　　　　San Diego, CA
Volunteer mentor　　　　　　　　　　　　　　　　　　　　　*Oct 2019—Jun 2020*
- Designed/taught a computational thinking course for underserved students in computer science

Friends of Washington Park　　　　　　　　　　　　　　　　　　Chicago, IL
Volunteer mentor　　　　　　　　　　　　　　　　　　　　　*Jan 2014—Jul 2017*
- Tutored 5th–8th grader students in an after-school program in local neighborhood of Hyde Park

PUBLICATIONS

Authors in alphabetical order unless otherwise noted.

Abhishek Roy*, Geelon So*, and Yi-An Ma. *Optimization on Pareto sets: On a theory of multi-objective optimization*. arXiv, under review, 2023. *equal contribution.

Sanjoy Dasgupta and Geelon So. *Online nearest neighbor classification*. arXiv, under review, 2023.

Sanjoy Dasgupta, Gaurav Mahajan, and Geelon So. *Convergence of online k-means*. International Conference on Artificial Intelligence and Statistics, 2022.

Geelon So. *Active learning with noise*. Master's thesis, Columbia University, 2019.

Geelon So. *Quantum computing: efficient prime factorization*. REU report, UChicago, 2015.

Geelon So. *Dynamical systems: symbolic dynamics*. REU report, UChicago, 2014.

TALKS

Paper talks and posters
- FODSI CCSI Student Posters (MIT)　　　　　　　　　　　　　　　*Jun 2023*
- EnCORE Student Social (UCSD)　　　　　　　　　　　　　　　　*Mar 2023*
- AISTATS Poster Session (Virtual)　　　　　　　　　　　　　　　*Mar 2022*

Reading seminars
- Forecasting and calibration (UCSD)　　　　　　　　　　　　　　*2023*
- Algorithmic game theory (UCSD)　　　　　　　　　　　　　　*2022—2023*
- Weak supervision for text data (Seekr)　　　　　　　　　　　　　*2022*
- Unsupervised learning (UCSD)　　　　　　　　　　　　　　　　*2022*
- Sampling/MCMC (UCSD)　　　　　　　　　　　　　　　　　*2021—2022*
- Neural networks (UCSD)　　　　　　　　　　　　　　　　　　*2020*
- Learning theory (UCSD)　　　　　　　　　　　　　　　　　　*2019*
- Sums-of-squares optimization (Columbia)　　　　　　　　　　　　*2018*
- Generalization theory (Columbia)　　　　　　　　　　　　　　　*2018*
- Type theory (UChicago)　　　　　　　　　　　　　　　　　*2016—2017*

Reading seminar presentations

- Universal prediction of individual sequences — *Nov 2023*
- Introduction to neural nets — *Jul 2023*
- Convergent message passing — *Jul 2023*
- Gradient-based multiobjective optimization — *Apr 2023*
- Introduction to mean-field games — *Feb 2023*
- The double descent phenomenon — *Nov 2022*
- Multicriteria decision making — *Oct 2022*
- Learning with multi-modal data: canonical correlation analysis — *Sep 2022*
- Gradient-based learning in games — *Apr 2022*
- Linear system identification with reverse experience replay — *Apr 2022*
- Introduction to algorithmic game theory — *Mar 2022*
- Independent component analysis — *Feb 2022*
- Equilibrium computation: motivation and problems — *Feb 2022*
- Scalable sampling for discrete distributions — *Nov 2021*
- Graphical games — *Nov 2021*
- Active learning for maximum likelihood estimation — *Oct 2021*
- Stochastic calculus on manifolds: part 1, part 2 — *Aug 2021*
- Linear system identification without mixing — *Jun 2021*
- Sequential kernel herding — *Jun 2021*
- Log-sobolev inequalities and concentration — *Apr 2021*
- Learning language games through interaction — *Apr 2021*
- Global non-convex optimization with discretized diffusion — *Apr 2021*
- Model of conserved macroscopic dynamics predicts future motor commands — *Feb 2021*
- A theory of universal learning — *Nov 2020*
- Oja's rule for streaming PCA — *Sep 2020*
- Proving the lottery ticket hypothesis — *Aug 2020*
- Approximate guarantees for dictionary learning — *Jun 2020*
- k-SVD for dictionary learning — *May 2020*
- Proximal methods for hierarchical sparse coding — *May 2020*
- Transformers are universal approximators — *Apr 2020*
- Using SVD to learn HMMs — *Feb 2020*
- Conditional mutual information and generalization — *Feb 2020*
- Generalization and adaptive data analysis — *Jan 2020*
- Generalization and differential privacy — *Nov 2019*
- Invariant risk minimization — *Nov 2019*
- Complexity: beyond space and time — *Aug 2019*
- Zero-knowledge proofs from MPCs — *Apr 2019*
- Geometry of gradient descent and lower bounds — *Feb 2019*
- Homomorphic encryption — *Feb 2019*
- Approximate nearest-neighbor search — *Dec 2018*
- Introduction to tensor decompositions — *Dec 2018*
- Sums-of-squares for robust estimation — *Nov 2018*
- Spectral graph theory, earlier version — *Oct 2018*
- Sums-of-squares for maxcut — *Sep 2018*
- Topological data analysis — *Jul 2018*
- Tensor decomposition for parametric estimation — *Jul 2018*
- PAC-Bayes for neural networks — *Apr 2018*
- Graph robustness and percolation theory — *Mar 2018*