

Ryan Theisen

ryantheisen.com ryanctheisen@gmail.com

San Francisco, CA

EDUCATION

University of California, Berkeley, Berkeley, CA 2018-2023

Ph.D. in Statistics. Thesis: *Beyond Worst-Case Generalization in Modern Machine Learning*,
Advised by Michael W. Mahoney

Arizona State University, Tempe, AZ 2013-2018

M.A. in Mathematics. Thesis: *Convergence Results for Two Models of Interaction*, Advised by
Sebastien Motsch

B.S. in Mathematics and Economics. Phi Beta Kappa, Summa Cum Laude, Outstanding
Graduate in both Mathematics and Economics

EXPERIENCE

Machine Learning Scientist, Harmonic Discovery Inc., San Francisco 2021-Present

Third employee and only machine learning scientist at Harmonic, leading the development of
three core machine learning tools central to HD's kinase drug discovery platform:

- Kanopy, model(s) that predict the kinome-wide bioactivity and selectivity of small molecules,
- Banyan, a generative modeling platform that uses molecular language models and tree search to design molecules with desired properties,
- Terra, a kinase/ligand co-folding model that samples ligand-conditioned conformational states of kinase binding pockets.

Graduate Student Researcher, University of California, Berkeley 2018-2023

Worked on fundamental research in machine learning under Professor Michael W Mahoney,
focusing on understanding generalization in over-parameterized machine learning.

Deep Learning Research Intern, Salesforce Research, Palo Alto Summer 2019, 2020

Worked with several Salesforce deep learning researchers on projects including: obtaining
improved norm-based generalization bounds for deep neural networks, and using normalizing
flows to estimate the Bayes error of high-dimensional machine learning datasets.

PEER-REVIEWED PUBLICATIONS

Hyunsuk Kim, Liam Hodgkinson, **Ryan Theisen**, and Michael W. Mahoney. How many
classifiers do we need? In *Advances in Neural Information Processing Systems*, 2024

Ryan Theisen, Tianduanyi Wang, Balaguru Ravikumar, Rayees Rahman, and Anna Cichońska.
Leveraging multiple data types for improved compound-kinase bioactivity prediction. *Nature
Communications*, 2024

Ryan Theisen, Hyunsuk Kim, Yaoqing Yang, Liam Hodgkinson, and Michael W. Mahoney. When are ensembles really effective? In *Advances in Neural Information Processing Systems*, 2023

Yaoqing Yang, **Ryan Theisen**, Liam Hodgkinson, Joseph E. Gonzalez, Kannan Ramchandran, Charles H. Martin, and Michael W. Mahoney. Test accuracy vs. generalization gap: Model selection in NLP without accessing training or testing data. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023

Yaoqing Yang, Liam Hodgkinson, **Ryan Theisen**, Joe Zou, Joseph E Gonzalez, Kannan Ramchandran, and Michael W Mahoney. Taxonomizing local versus global structure in neural network loss landscapes. In *Advances in Neural Information Processing Systems*, 2021

Ryan Theisen, Huan Wang, Lav R Varshney, Caiming Xiong, and Richard Socher. Evaluating state-of-the-art classification models against Bayes optimality. In *Advances in Neural Information Processing Systems*, 2021

Ryan Theisen, Jason Klusowski, and Michael Mahoney. Good classifiers are abundant in the interpolating regime. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, 2021

Fei Cao, Sebastien Motsch, Alexander Reamy, and **Ryan Theisen**. Asymptotic flocking for the three-zone model. *Mathematical Biosciences and Engineering*, 2020

Dylan Weber, **Ryan Theisen**, and Sebastien Motsch. Deterministic versus stochastic consensus dynamics on graphs. *Journal of Statistical Physics*, 2019

PRE-PRINTS AND WORKSHOP PAPERS

Annalise Schweickart, Juhi Somani, **Ryan Theisen**, Navriti Sahni, Anna Cichonska, and Rayees Rahman. Delineating drug class and target-specific adverse events of kinase inhibitors. Technical Report <https://doi.org/10.1101/2024.08.07.24311573>, 2024

Ryan Park, **Ryan Theisen**, Navriti Sahni, Marcel Patek, Anna Cichońska, and Rayees Rahman. Preference optimization for molecular language models. In *Workshop on Generative AI and Biology, NeurIPS*, 2023

Ryan Theisen, Jason M. Klusowski, Huan Wang, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. Global capacity measures for deep ReLU networks via path sampling. Technical Report arXiv:1910.10245, 2019

TEACHING

University of California, Berkeley

Graduate Student Instructor, Stat 157: Deep Learning

Graduate Student Instructor, Stat 154: Modern Statistical Prediction and Machine Learning

Graduate Student Instructor, Stat 89A: Linear Algebra for Data Science

Arizona State University

Teaching Assistant, ECN 312: Intermediate Microeconomics

Teaching Assistant, ECN 212: Principles of Microeconomics

SERVICE

Volunteer, Kino Border Initiative, Nogales, Mexico.

2012-Present

AWARDS AND HONORS

Dean's Medal, Mathematics, ASU, 2017 Awarded as the outstanding graduating senior in the School of Mathematical and Statistical Sciences.

Dean's Medal, Economics, ASU, 2017 Awarded as the outstanding graduating senior in the Department of Economics.

Moeur Award, ASU, 2017 Granted University-wide academic distinction at graduation.

J.P. Morgan Chase Scholar, ASU, 2016 Awarded fellowship for top economics undergraduates.

Fulbright Summer Scholarship, University of Bristol, UK, 2014 UK Summer Institute for Young American Student Leaders.

SKILLS AND INTERESTS

Programming: Python, JavaScript, SQL, L^AT_EX.

Languages: English, Spanish (elementary proficiency).

Research Interests: Theoretical Machine Learning, Deep Learning, Statistics, Optimization.

Personal Interests: Photography, Guitar, Arsenal FC, Green Bay Packers, Hiking, Travel.