Zachary Novack

zacharynovack.github.io
novackze@gmail.com
(561) 866-0646

ACADEMIC INTERESTS

Machine learning (ML) for public policy, computational social science, fundamentals of deep learning, explainable ML, fairness in ML, music information retrieval

EDUCATION BACKGROUND

B.S. in Statistics & Machine Learning — Expected May 2022
Carnegie Mellon University, Pittsburgh, PA

- Minor in Sonic Arts (music technology)
- 3.91/4.0 GPA
- *Selected Coursework:*
  - In Progress: Convex Optimization, Algorithms & Data Structures, Philosophy of ML
  - Completed: ML w/Large Datasets, Real Analysis, Numerical Linear Algebra, Probability & Statistics, Introduction to ML, Statistical Computing, Linear Algebra, Multimedia Signal Processing

RESEARCH EXPERIENCE

Research Assistant — Spring 2021 - Present
ACMI Lab, Carnegie Mellon University, Pittsburgh, PA

- Investigated what properties of Stochastic Gradient Descent (SGD) noise contribute to improved generalization performance over full-batch methods. Empirically discovered that while gradient covariance is not a necessary condition for optimal generalization in large neural networks, reconstruction of both the shape and scale of the empirical noise distribution is a key factor in generalization.
- Maintained large codebase in Pytorch to run suite of experiments with different deep architectures and image recognition tasks
- Prepared paper for submission to **ICML 2022**

Research Assistant — Summer 2020 - Present
Laboratory for Social Minds, Carnegie Mellon University, Pittsburgh, PA

- Implemented Latent Dirichlet Allocation (LDA) to investigate ideological network evolution on the fringe web forums /pol/ (4chan) and The Red Pill (Reddit)
- Designed a Bayesian autoregressive model to analyze addiction effects on social media websites, to be submitted to **Nature Human Behavior**
- Explored how structural differences in online communities may influence the cognitive entropy of a given website's topic distribution, to be submitted to **Entropy**

Undergraduate Researcher — Summer 2020 - Spring 2021
Dietrich College, Carnegie Mellon University, Pittsburgh, PA

- Constructed filtering algorithm to parse sparse text documents for specific topic occurrences
- Modified existing sentiment analysis implementation to account for valence-shifters in congressional speeches

- Implemented multiple behavioral game theoretic models in matlab to simulate strategic choice patterns in asymmetric two-player games

WORK EXPERIENCE

AI/ML Intern Summer 2020 - Spring 2021
Unisys Corporation, Blue Bell, PA

- Designed python implementation of categorical distance metrics to interface with scikit-learn clustering algorithms
- Deployed time-series models (ARIMA, LSTM, Facebook Prophet) to predict computer resource utilization
- Developed model retraining infrastructure to automatically track distribution shift in time-series models

Studio Intern Summer 2019
Joy Records, Tel Aviv, Israel

- Analyzed commercial streaming data to construct customized playlists for clients
- Assisted in website developement for Hebrew-to-English translations

Percussion Arranger Fall 2018 - Spring 2019
Tomball High School Indoor Percussion, Tomball, TX

- Arranged musical production for large percussion ensemble in order to compete in the Winter Gaurd International (WGI) national circuit

TEACHING EXPERIENCE

Teaching Assistant
Carnegie Mellon University, Pittsburgh, PA

- *10-301/601: Introduction to Machine Learning* *Fall 2021*
  - Spearheaded team maintaining autograder implementation for coding assignments
  - Led recitation and designed homework questions for class of 500+ students
  - Topics Covered: Decision Trees, Linear & Logistic Regression, Regularization, Dense and Convolutional Neural Networks, PAC Learning, Generative Models, MAP Estimation, Bayesian Networks, Hidden Markov Models, Markov Decision Processes, Clustering, Ensemble Methods
- *85-340: Research Methods for Social Psychology* *Fall 2021*
  - Fully created and taught course module introducing R for psychology students, including computer science fundamentals and applications for experiment design and data analysis
  - Topics Covered: Basic types, functions, vectorized programming, workflow in dplyr, basic statistical analysis, one-way and two-way ANOVA
- *36-225: Introduction to Probability Theory* *Summer 2021*
  - Topics Covered: Basic probability, random variables, univariate/multivariate probability distributions, moment-generating functions, central limit theorem
- *36-226: Introduction to Statistical Inference* *Spring 2021*
  - Topics Covered: Maximum likelihood estimation, method of moments, large & small sample hypothesis testing, properties of point estimators, confidence intervals, order statistics, Type I & Type II errors, ANOVA
- *88-300: Programming for Social Scientists* *Summer 2020 - Spring 2021*

- Topics Covered: Basic data analysis, workflow in dplyr, basic text analysis, linear regression

PUBLIC WORKS

Poster Presentations

- **Zachary Novack**, Eden Hu, and Mason Lin, *Tracking Political Sentiment on Cold War China in Congressional Speeches*, Carnegie Mellon University Statistics and Data Science Research Showcase, May 2021

Blog Posts

- **Zachary Novack**, *Armchair Statistics: Benford's Law and other Misconceptions in the Age of Data*, Carnegie Mellon University Triple Helix, April 2021

PROJECTS

RoboPierre Spring 2020

Adaptive Impressionist Music via Generative Modeling

- Developed interactive web app to randomly generate polyphonic music trained on impressionistic composers
- Implemented using Google Magenta's Polyphony RNN and custom stochastic voice leading algorithm

ThereMyn Spring 2019

Motion-Controlled Monophonic Synthesizer

- Used infrared distance monitor to drive audio signal creation
- Created front-end GUI to translate audio signals into a usable motion-controlled synthesizer

ACCOLADES

Honors Programs

- *Phi Beta Kappa*, October 2021 - Present
- *Andrew Carnegie Society Scholar*, September 2021 - Present
- *Quantitative Social Science Scholar*, August 2018 - Present
- *Dean's List: High Honors*, December 2018 - Present

Awards

- *Small Undergraduate Research Grant (SURG)* for "Statistical Inference of Online Radicalization in Extremist Communities", Carnegie Mellon University, June 2021
- *Dietrich Senior Honors Research Fellowship* for "Autoregressive Models of Online Addiction", Dietrich College, Carnegie Mellon University, May 2021
- *First Place: Statistics & Data Science Research Showcase*, for "Tracking Political Sentiment on Cold War China in Congressional Speeches", Carnegie Mellon University, May 2021
- *Summer Undergraduate Research Fellowship (SURF)*, for "Empirical Test of the Dual Accumulator Model", Carnegie Mellon University, June 2020
- *Second Place: 15-112 Term Project Showcase* for "ThereMyn: Motion-Controlled Monophonic Synthesizer", School of Computer Science, Carnegie Mellon University, April 2019

Scholarships

- *Paul Mellon Memorial Presidential Scholarship* (merit-based), August 2018 - Present

SKILLS

Programming Languages and Packages

- Python (Pytorch, Tensorflow, Scikit-Learn, PySpark, CVXPY), R (dplyr, tscount, zoo), C, Matlab, SQL (postgres, MySQL), Stan, Git, Shell, Max/MSP/Jitter

Other Skills

- AWS (S3, EC2, EMR), Microsoft Azure, Docker, Agile, Jira, Grafana, Ableton Live

EXTERNAL ACTIVITIES

Professional Event Coordinator Spring 2021 - Present
American Statistical Association, Carnegie Mellon University, Pittsburgh, PA

- Coordinated multi-part speaker series featuring both faculty and external researchers
- Facilitated peer-mentorship program within the Statistics environment for future course planning

Staff Writer Fall 2020 - Present
The Triple Helix, Carnegie Mellon University, Pittsburgh, PA

- Wrote journal articles on wide-scale statistical literacy and societal impacts of misreporting experimental results

Performer and Composer Spring 2019 - Spring 2020
Exploded Ensemble, Carnegie Mellon University, Pittsburgh, PA

- Designed large-scale Max/MSP programs for multimedia interactive performances
- Composed electro-acoustic pieces for mixed instrumentation ensembles