Himadri Mandal

mandalhimadri06@gmail.com • quirtt.github.io • quirtt
in quirtt

Education

Indian Statistical Institute, Kolkata **August 2023 — present**
Statistics Undergraduate, 2nd year — 88.7% cumulative (1st year)

Programming experience

Python, Bash, R, Next.js + TailwindCSS, LaTeX, Linux (Arch Linux on i3)

Skills

Leadership, Deductive Reasoning, Debate, Direct Communication, Typing (100+ wpm)

Selected fellowships and awards

ISI K. Semester 1 Outstanding Performance: awarded January 2024, received ₹1500
ISI Kolkata B.Stat. Entrance: awarded August 2023, ranked 11th
Indian Olympiad Qualifier for Mathematics, KV: awarded February 2021, rank 4 in my region
Atlas Fellowship India Finalist: awarded September 2022, received 1000$, top 200 in a rationality fellowship
IISc Enumeration Finalist: awarded October 2022
CMI Tessellate Finalist: awarded October 2021

Selected Courses

CaMLAB: Cambridge AI Safety Hub *April 8 — April 21*
Course to build ML engineering fundamentals for AI Safety research. Includes basics of PyTorch, training and tuning GPTs and ResNets, interpreting models with TransformerLens, and an introduction to RL, RLHF.

Deep Learning: ISI Kolkata *January 2024 — March 2024*
A winter course on deep learning covering Autoencoders, CNNs, GANs, GNNs, Diffusion models, RNNs, Attention mechanics, Transformers, etc.

Measure Theory: Maths Club, ISI Kolkata *December 2023 — February 2024*
Introductory course on Measure Theory which helped me understand all the details in our Probability courses.

Projects

Research

Circuit Phenomenology Using Sparse Autoencoders: w/ David Udell *June 2024 — July 2024*
Preprint. Sparse autoencoders enable interpretable representations of model activations, aiding mechanistic interpretability by uncovering causal circuits. We went through the literature, independently implemented circuit discovery for GPT-2-small and ended up finding big errors in the implementation of the algorithm in the latest paper by David Bau. We solved those issues, and experiemented with newer ideas, improving circuit discovery in GPT-2-small.

Theoretical

Universal Source Coding:
Report. The broad setup is the following: there's data coming in from some source. If the source distribution is known, then Huffman Encoding gives the optimal encoding scheme. This project tries to figure out a good encoding scheme (in multiple contexts!) that guarantees performance against all source distributions!

Last updated December 3, 2024

Independence Is Almost Dependence:
Blog. My independently discovered proof to a theorem: given two independent random variables X, Y you can come up with two new random variables U, V which have the same marginals and ϵ-close joints but are deterministically dependent.

Cold Reflections:
Link. My blog on Mathematics, Statistics, Philosophy and everything else that interests me. I add some of my work there.

Empirical/Programming

Ponderings on OthelloGPT:
Blog. Mechanistic Interpretability project. OthelloGPT is a GPT model trained on Othello games to predict all the possible legal moves. I look into how the model computes how a certain cell is blank.

ORIGAMI:
Repo. Implements arXiv:2303.17062, AISTATS 2023. A paper on dimensionality reduction of the support to improve computational efficiency in downstream decision making.

Shannon's Mind Reading Game:
Blog. The magician sends you a deck of cards. You riffle shuffle it three times. You pick the card from the top and put it back in the deck, somewhere. You send the deck back to the magician. Can he find the chosen card?

Selected work experience

Software

Website Lead: MTRP, Integration *November 2023 — January 2024*
Repo. Preview. Deployed a website for the university's fest's annual mathematics competition after learning Next.JS and TailwindCSS, all of it took a week. Maintained it for efficiency and bugfixes.

Volunteer and outreach

Owner: awas *October 2020 — October 2022*
Served as the **organizer and mentor** for daily math problem solving sessions, philosopical debates, and programming discussions for over two years on **Discord**. Mentored smart math enthusiasts, from all over India, learn hard math, and guided a few of them to get into the Indian training camp of IMO.

Last updated December 3, 2024