SHUNING SHANG

shuning@princeton.edu ◇ https://nooraovo.github.io/

EDUCATION

Princeton University

Ph.D in Computer Science *Aug, 2025 – Present*

- Advisor: Sanjeev Arora

Zhejiang University

B.Eng in Computer Science (Honors Program "Mixed Class") *Aug, 2021 – Jul, 2025*

- **GPA: 3.99/4.0, Ranking: 1^{st}/304**

University of Michigan, MI, U.S.

Research Intern *May, 2024 – Oct, 2024*

University of California, Berkeley, CA, U.S.

Exchange Student *Jan, 2024 – May, 2024*

RESEARCH INTEREST

My long-term goal is to build safe and scalable AI systems that can drive significant scientific progress. My recent research interests focus on reinforcement learning theory and reasoning for large language models. I also worked on deep learning theory and optimization.

PUBLICATIONS & MANUSCRIPTS

1 **Shuning Shang***, Hubert Strauss*, Stanley Wei, Sanjeev Arora, Noam Razin. When Errors Can Be Beneficial: A Categorization of Imperfect Rewards for Policy Gradient.
Submitted to *ICLR.* [arXiv]

2 Roey Magen*, **Shuning Shang***, Zhiwei Xu, Spencer Frei, Wei Hu, Gal Vardi. Benign Overfitting in Single-Head Attention. *The Thirty-ninth Annual Conference on Neural Information Processing Systems(Nuerips 2025).* [arXiv]

3 **Shuning Shang**, Xuran Meng, Yuan Cao, Difan Zou. Initialization Matters: On the Benign Overfitting of Two-Layer ReLU CNN with Fully Trainable Layers. Submitted to *Journal of Machine Learning Research(JMLR).* Under review. [arXiv]

RESEARCH EXPERIENCE

The University of Michigan MI, US
Department of Computer Science and Engineering *May, 2024 – May, 2025*
Advisor: Prof. Wei Hu
Research Topic: Benign Overfitting in Single-Head Attention Model

- Analyzed the training dynamics of attention mechanisms on learning signal-noise data models using max-margin interpolators, from both empirical and theoretical perspectives.
- Investigated benign overfitting phenomenon in attention models by extending the "Optimal Token Selection" mechanism to non-asymptotic scenarios, enhancing its practical applicability.
- Established a precise signal-to-noise ratio (SNR) threshold to distinguish between benign and harmful overfitting, and estimated the corresponding test errors.

The University of Hong Kong Hong Kong, P.R.China
Department of Computer Science *Jul, 2023 – Nov, 2024*
Advisor: Prof. Difan Zou
Research Topic: Two-layer ReLU CNN with Fully Trainable Weights

- Extended two-layer CNN models to a more practical setting where both layers are trainable, revealing that the initialization scale of the output layer is crucial to training dynamics.
- Applied two-stage analysis method to better track the dynamic of the training process and discovered a phase transition occurs when the output layer initialization scale is small.
- Characterized benign overfitting behavior in two-layer CNNs and established nearly matching upper and lower bounds on the test error.

SCHOLARSHIPS & AWARDS

- **2022,2023 National Scholarship** (Highest scholarship awarded by the Chinese government, $<$ **0.1%**)
- **2022,2023 The First Prize** Scholarship of Zhejiang University (Highest scholarship awarded by Zhejiang University, $<$ **3%**)
- **2022 Top Ten** First-year Students in Engineering Department (**10/731**)

SKILLS

Programming	Python, PyTorch, C/C++, MATLAB, LaTeX, HTML, etc.
Language	TOEFL iBT 108/120 (Reading 26, Listening 28, Speaking 25, Writing 29)
Volunteer	Taught science and music in Summer 2022 for underprivileged primary school students