Huy Nguyen

Last update: February 05, 2025

CONTACT INFORMATION

`Phone:` (+1) 571 356-7068
`University Email:` huynm@utexas.edu
`Personal Email:` huynguyen.hmn@gmail.com
`Homepage:` https://huynm99.github.io/

EDUCATION

The University of Texas at Austin, Austin, TX, USA. 2022-2027
Ph.D Candidate in Statistics at the Department of Statistics and Data Sciences.

- Advisors: Professor Nhat Ho and Professor Alessandro Rinaldo.
- GPA: 4.0/4.0

Ho Chi Minh University of Science, Ho Chi Minh City, Vietnam. 2017–2020
Bachelor of Science in Mathematics and Computer Science, Honor Program.

- Advisor: Professor Dang Duc Trong.
- GPA: 9.75/10.0 - *Summa Cum Laude*.

RESEARCH EXPERIENCE

Microsoft AI, Redmond, WA, USA. Summer 2024
Research Intern.

- Research topics: Applications of Mixture of Experts in Large Language Models.
- Propose a method for selecting crucial attention heads in the multi-head mechanism based on the routing strategy in mixture of experts to improve the efficiency of the Large Language Models.

The University of Texas at Austin, Austin, TX, USA. 2024-Present
Graduate Research Assistant.

- Research topics: Towards Understanding Mixture of Experts: From Theory to Practice.
- Supervisors: Professor Nhat Ho and Professor Alessandro Rinaldo.

VinAI, Hanoi, Vietnam. 2020–2022
AI Research Resident.

- Research topics: Optimal Transport theory and its applications in Domain Adaptation.
- Skill gained: Did research on Optimal Transport (Sinkhorn algorithms, Barycenter computation, etc) and applied them to study Data Shift and Label Shift problems in Domain Adaptation.

RESEARCH INTERESTS

My research focuses on theoretical foundations for the Mixture-of-Experts (MoE) models where I study the effects of different gating functions (namely the softmax gate, the top-K sparse gate, the dense-to-sparse gate and the sigmoid gate, etc) on the convergence behavior of expert estimation. Based on insights from these results, I aim to design novel gating functions and characterize expert networks which help improve the efficiency and scalability of the MoE applications, including Large Language Models, Multimodal Learning and Parameter-efficient Fine-tuning (including prompt-based tuning and adapter-based tuning). Additionally, I am also interested in Optimal Transport theory.

PUBLICATIONS

20. **Huy Nguyen**, Pedram Akbarian*, Trang Pham*, Trang Nguyen*, Shujian Zhang, Nhat Ho. Statistical Advantages of Perturbing Cosine Router in Mixture of Experts. *In International Conference on Learning Representations, 2025.*
19. Minh Le*, Chau Nguyen*, **Huy Nguyen***, Quyen Tran, Trung Le, Nhat Ho. Revisiting Prefix-tuning: Statistical Benefits of Reparameterization among Prompts. *In International Conference on Learning Representations, 2025.*
18. **Huy Nguyen**, Nhat Ho**, Alessandro Rinaldo**. Sigmoid Gating is More Sample Efficient than Softmax Gating in Mixture of Experts . *Advances in Neural Information Processing Systems, 2024.*

17. Xing Han, **Huy Nguyen***, Carl Harris*, Nhat Ho, Suchi Saria. FuseMoE: Mixture-of-Experts Transformers for Fleximodal Fusion . *Advances in Neural Information Processing Systems, 2024.*

16. Minh Le, An Nguyen*, **Huy Nguyen***, Trang Nguyen*, Trang Pham*, Linh Van Ngo, Nhat Ho. Mixture of Experts Meets Prompt-Based Continual Learning . *Advances in Neural Information Processing Systems, 2024.*

15. **Huy Nguyen**, Nhat Ho**, Alessandro Rinaldo**. On Least Square Estimation in Softmax Gating Mixture of Experts. *Proceedings of the International Conference on Machine Learning, 2024.*

14. **Huy Nguyen**, Pedram Akbarian, Nhat Ho. Is Temperature Sample Efficient for Softmax Gaussian Mixture of Experts? *Proceedings of the International Conference on Machine Learning, 2024.*

13. **Huy Nguyen**, Pedram Akbarian, TrungTin Nguyen, Nhat Ho. A General Theory for Softmax Gating Multinomial Logistic Mixture of Experts. *Proceedings of the International Conference on Machine Learning, 2024.*

12. **Huy Nguyen**, Pedram Akbarian, Fanqi Yan, Nhat Ho. Statistical Perspective of Top-K Sparse Softmax Gating Mixture of Experts. *In International Conference on Learning Representations, 2024.*

11. Fanqi Yan*, **Huy Nguyen***, Dung Le*, Pedram Akbarian, Nhat Ho. Understanding Expert Structures on Minimax Parameter Estimation in Contaminated Mixture of Experts . *In International Conference on Artificial Intelligence and Statistics, 2025.*

10. **Huy Nguyen***, TrungTin Nguyen*, Khai Nguyen, Nhat Ho. Towards Convergence Rates for Parameter Estimation in Gaussian-gated Mixture of Experts. *In International Conference on Artificial Intelligence and Statistics, 2024.*

9. **Huy Nguyen**, Khai Nguyen, Nhat Ho. On Parameter Estimation in Gaussian Deviated Mixture of Experts. *In International Conference on Artificial Intelligence and Statistics, 2024.*

8. **Huy Nguyen**, TrungTin Nguyen, Nhat Ho. Demystifying Softmax Gating Function in Gaussian Mixture of Experts. *Advances in Neural Information Processing Systems, 2023 (Spotlight, Top 3.6% out of 12343 submissions).*

7. Dat Do*, **Huy Nguyen***, Khai Nguyen, Nhat Ho. Minimax Optimal Rate for Parameter Estimation in Multivariate Deviated Models. *Advances in Neural Information Processing Systems, 2023.*

6. Dung Le*, **Huy Nguyen***, Khai Nguyen*, Trang Nguyen*, Nhat Ho. Fast Approximation of the Generalized Sliced-Wasserstein Distance . *IEEE International Conference on Acoustics, Speech and Signal Processing, 2024.*

5. Khai Nguyen, Tongzheng Ren, **Huy Nguyen**, Litu Rout, Tan Nguyen, Nhat Ho. Hierarchical Sliced Wasserstein Distance. *In International Conference on Learning Representations, 2023.*

4. **Huy Nguyen***, Khang Le*, Dung Le*, Dat Do, Tung Pham, Nhat Ho. Entropic Gromov-Wasserstein between Gaussian Distributions. *Proceedings of the International Conference on Machine Learning, 2022.*

3. **Huy Nguyen***, Khang Le*, Khai Nguyen, Tung Pham, Nhat Ho. On Multimarginal Partial Optimal Transport: Equivalent Forms and Computational Complexity. *In International Conference on Artificial Intelligence and Statistics, 2022.*

2. **Huy Nguyen***, Khang Le*, Quang Minh Nguyen, Tung Pham, Hung Bui, Nhat Ho. On Robust Optimal Transport: Computational Complexity and Barycenter Computation. *Advances in Neural Information Processing Systems, 2021.*

1. Thu Nguyen, Duy H. M. Nguyen, **Huy Nguyen**, Binh T. Nguyen, Bruce A. Wade. EPEM: Efficient Parameter Estimation for Multiple Class Monotone Missing Data. *Information Sciences Journal, Volume 567, page 1-22.*

Preprints

7. **Huy Nguyen***, Fanqi Yan*, Pedram Akbarian, Nhat Ho**, Alessandro Rinaldo**. Sigmoid Self-Attention is Better than Softmax Self-Attention: A Mixture-of-Experts Perspective. *Under review, arXiv:2502.00281.*
6. Tuan Truong*, Chau Nguyen*, **Huy Nguyen***, Minh Le, Trung Le, Nhat Ho. RepLoRA: Reparameterizing Low-rank Adaptation via the Perspective of Mixture of Experts. *Under review, arXiv:2502.03044.*
5. Minh Le, Anh Nguyen, **Huy Nguyen**, Chau Nguyen, Nhat Ho. Adaptive Prompt: Unlocking the Power of Visual Prompt Tuning. *Under review, arXiv:2501.18936.*
4. Nghiem Tuong Diep*, **Huy Nguyen***, Chau Nguyen, Minh Le, Duy Minh Ho Nguyen, Daniel Sonntag, Mathias Niepert, Nhat Ho. On Zero-Initialized Attention: Optimal Prompt and Gating Factor Estimation. *Under review, arXiv:2502.03029.*
3. **Huy Nguyen***, Xing Han*, Carl Harris, Suchi Saria**, Nhat Ho**. On Expert Estimation in Hierarchical Mixture of Experts: Beyond Softmax Gating Functions. *Under review, arXiv:2410.02935.*
2. Pedram Akbarian*, **Huy Nguyen***, Xing Han*, Nhat Ho. Quadratic Gating Functions in Mixture of Experts: A Statistical Insight . *Under review, arXiv:2410.11222.*
1. Quang Pham, Giang Do, **Huy Nguyen**, TrungTin Nguyen, Chenghao Liu, Mina Sartipi, Binh T. Nguyen, Savitha Ramasamy, Xiaoli Li, Steven Hoi, Nhat Ho. CompeteSMoE - Effective Training of Sparse Mixture of Experts via Competition . *Under review, arXiv:2402.02526.*

Teaching Experience

The University of Texas at Austin, Austin, TX, USA.
Teaching Assistant at the Department of Statistics and Data Sciences.

- SDS302F - Foundations of Data Analysis. Fall 2022
- SDS322E - Elements of Data Science. Spring 2023
- SDS320E - Elements of Statistics. Fall 2023

Professional Services

Program Committee/Reviewer at

- the Electronic Journal of Statistics (EJS).
- the Transactions on Machine Learning Research (TMLR)
- the International Conference on Machine Learning (ICML) 2022-2025.
- the Conference on Neural Information Processing Systems (NeurIPS) 2022-2024.
- the International Conference on Artificial Intelligence and Statistics (AISTATS) 2022-2025.
- the International Conference on Learning Representations (ICLR) 2024-2025.
- the Association for the Advancement of Artificial Intelligence (AAAI) 2025.

Honors and Awards

- Top Reviewer at NeurIPS 2024. 2024
- AISTATS 2024 Registration Grant. 2024
- ICLR 2024 Travel Award. 2024
- NeurIPS 2023 Scholar Award. 2023
- Doctoral Fellowship of the University of Texas at Austin. 2022

Technical Skills

- *System*: MacOS, Linux, Windows.
- *Programming Languages*: Python (Pytorch, Sci-kit Learn, Numpy, Matplotlib), R, MATLAB.
- *Softwares*: LaTeX, Microsoft Offices.

References

- **Nhat Ho.** Email: minhnhat@utexas.edu (Advisor).
- **Alessandro Rinaldo.** Email: alessandro.rinaldo@austin.utexas.edu (Advisor).