

Hoang Viet Chu

1050 North Mills Avenue, Claremont, CA | chu70643@gmail.com | 909-407-6637 | github.com/HoangChu-Claremont

EDUCATION

Pitzer College

Claremont, CA

Bachelor's Degree, Joint Computer Science and Mathematics (Honors)

Expected May 2024

- **GPA:** 3.7 / 4.0 (Major GPA: 3.9 -- Math and CS classes cross-registered at **Harvey Mudd College**)
- **Classes:** Machine Learning, Mathematics of Topic Modeling, Neural Network, Computer Systems, Linear Models.
- **Teaching Assistant -- 1 year:** Data Structures and Algorithms (in C++), Intermediate Probability, Abstract Algebra.

AWARDS

- **Math:** 1st Prize: Vietnam Mathematical Olympiad, Silver Medal: Asian Pacific Mathematical Olympiad, 3x AIME
- **Coding** (no high school experience): [1st Prize: Citadel Datathon](#), [Round 3: Google Code Jam](#), [Winner: SIG Challenge](#)

EXPERIENCE

Periwinkle Trading

August 2023 - Present

Quantitative Researcher - Contract (mentored by Scott Smallwood, CEO and former Partner at PDT)

Python

- Developed an end-to-end Generalized Regression model with Elastic Net regularization, achieving a $R^2 \sim 0.55$ and satisfying all Gauss-Markov assumptions with Weighted Correlation Matrix and Variance-Stabilizing Transformation.
- **Reduced 30% runtime** of the model pipeline by **avoiding matrix inversion**: removed correlated features with PCA, optimized Beta(s) with QR Factorization by Givens Rotations, and estimated Covariance Matrix with Factor Models.
- Consulted that, given a **10M+ tabular data with 1% outliers**, prioritized XGBoost over Recurrent Neural Network.
- Proactively trained a XGBoost prototype with a self-tuned Huber loss and got **0.6% relative returns** after 3 weeks.

University of Southern California

June 2023 - August 2023

Undergraduate Research Intern

Python, MATLAB

Topic: "Last-mile Delivery Optimization with Recurrent Neural Network"

- Designed a comprehensive feature selection that **reduced from 20 to 5** features needed while keeping $R^2 \sim 0.652$.
- Proposed and implemented a greedy Path Generation Algorithm in Python that **improved disparity score 10-20%**.
- **Improved** the team's RNN model accuracy **by 20%** in predicting a driver path by proposing a pairwise instead of a single pointer network, and optimized it to **real-time prediction**, allowing re-suggesting routes when drivers deviate.

Meta

May 2022 - August 2022

Engineering Intern (received return offer)

Python, Hack

- Proposed adaptation of two Microsoft Research papers on Vision Transformer (ViT) to Meta's ResNet50 model and customized a Window Selection algorithm addressing the ViT's scaling issue with **100,000,000+** Marketplace images.
- Independently built a prototype of the ViT model resulting in a **2% accuracy improvement** on 2,000,000 images.
- Initiated to implement and train Contrastive Captioners model, helping the team test new text generative algorithms.

Independent Research

January 2022 - May 2022

Topic: "Explore Constraints on Unitary Recurrent Neural Networks" ([paper](#))

Python

- Proved that the norm of Unitary Matrices are bounded, thus they can prevent RNN vanishing and exploding gradients.
- Exploited the orthogonality of Hermitian matrices to create a **new update rule** that **ensures the next weight matrix is always a unitary matrix** (complex spectral theorem) and **is computed efficiently** by eigen-decomposition.
- Implemented **from scratch** a Unitary Recurrent Neural Network with the above new update rule, adapting Arjovsky's paper, to demonstrate that the new model doesn't explode or vanish and can be used for LLM and Topic Modeling.

Cohost.ai ([cohost.ai](#))

June 2021 - August 2021

Engineering Intern

Python, C++

- Ideated a full-stack design and built from scratch an NLU-based conversational chatbot that serves 2000+ daily users.
- Collaborated with full-time engineers to build an IPC message queue to avoid message loss if chatbot crashes.
- Designed Tree Serialization algorithms reducing company web PageSpeed **from 30 secs to a consistent 0.8 second**.

PROJECTS

Citadel Datathon - LendingClub Risk Assessment (26,000,000 rows)

Python

- Ideated and built a customized Mahalanobis distance metric for K-Means algorithm, allowing for the final discovery.