Peiru (Jenny) Xu (She/Her/Hers)

xpr2019@berkeley.edu | 510-561-5734 | https://xupeiru2000.github.io | http://linkedin.com/in/jenny-peiru-xu

EDUCATION

University of California, Berkeley   Berkeley, CA
- Bachelor of Arts, Applied Mathematics & Philosophy (GPA 3.902)   Aug. 2019 – May. 2023

PUBLICATION

- Y. Zhou., **P. Xu.**, G. Hooker., "A Generic Approach for Statistical Stability in Model Distillation". *submitted, arXiv:2211.12631* [arXiv][pdf]
- **P. Xu.**, J. Hong., H. Wang., M. Wang., "Neural Directed Evolution for Protein Sequence Optimization". *in preparation*

RESEARCH EXPERIENCE

Student Researcher, Department of Statistics, UC Berkeley   Advisor: Prof. Giles Hooker
A Generic Approach for Statistical Stability in Model Distillation   May. 2022 – Current
- Explore the statistical stability and interpretability of learned explanations for black-box models via model distillation, which is the process of first train the black box teacher model and generate pseudo-dataset with corresponding teacher predictions, and finally train the student model with easy interpretability.
- Develop a generic approach for stable model distillation based on central limit theorem, construct a multiple hypothesis testing framework to select a large enough pseudo sample size to ensure that the best distilled student model would be selected under different pseudo samples.
- Investigate application of the proposed distillation algorithm on decision tree, falling rule list and symbolic regression, illustrate the testing procedure through a theoretical analysis with Markov process, and conduct simulation experiments on real-world datasets to demonstrate effectiveness of the proposed algorithm.

Machine Learning Researcher Intern, ABio-X Holdings   Advisor: Prof. Mengdi Wang, Prof. Huazheng Wang
Neural Directed Evolution for Model-guided Protein Sequence Optimization   July. 2022 – Current
- Propose a Monte Carlo Tree Search-based Directed Evolution framework for sequence optimization, which utilizes upper-confidence bound (UCB) algorithm that effectively search for closely related protein mutants with high fitness values and relatively low mutation counts. Formalize a bandit model for Directed Evolution as an initial population evolves via uniform mutation and directed recombination of parent sequences, and evaluates the fitness value of the next children generation.
- Simulate the ground-truth protein fitness landscape with a pre-trained black-box oracle model in replacement of real-world wet-lab experiments and measurements, use TAPE embedding inputs to simulate and train the black-box oracle model.
- Evaluate the proposed MCTS-based DE on benchmark protein sequence datasets including GB1, AAV and WW domain, achieve substantial improvement of performance over baseline algorithms (AdaLead, DyNA PPO, DbAS, CbAS, PEX).

Student Researcher, US Air Force Research Laboratory   Advisor: Dr. Alvaro Velasquez
Automaton Distillation for Non-Markovian Knowledge Transfer in Deep RL   Aug. 2021 – Dec. 2021
- Proposed automaton distillation as a form of symbolic transfer learning to mitigate issues of non-Markovian decision process with traditional transfer learning methods. Introduced a new distillation approach which first trains a teacher agent using standard Deep Q-Learning, then distills teacher Q-value estimates into the learning of the student in the target environment.
- Leveraged the automaton representation of an objective in the proposed distillation approach to convey information from the teacher to the student by distilling teacher Q-value to the objective automaton, used a modified DQN loss function that incorporates the automaton Q-values for training the student.
- Evaluated the proposed automaton distillation algorithm on various grid-world environments with smaller teacher environment and larger target environment, achieved better performance and faster convergence compared to existing baseline methods. [slides]

RESEARCH INTERESTS AND SKILLS

- Research interests include **interpretable Machine Learning**, **statistical inference** with ML models, **multi-armed bandits** in Reinforcement Learning, distillation for Transfer Learning, and ML in general.
- **Machine Learning** skills: Multiple courses and research projects on deep learning and reinforcement learning
- **Programming** skills: Python(3 years), R(4 years), MATLAB(3 years), SQL(2 years)
- **Data Analysis** skills: sufficient knowledge and experience in EDA and data visualization

INTERNSHIP

Essence Securities Research Center, Shanghai, China   Manager: Zhicheng Ye, Dr. Sijiang Lu
Quantitative Research Intern   July. 2021 – Aug. 2021
- Built the quantitative model for multi-factor investment strategy of convertible bonds, developed a customized back-test framework using python that used historical data of convertible bonds in China trading market to assess multiple investment strategies
- Conducted simulation experiments on portfolio adjustment and back-tested existing multi-factor investment strategies of convertible bonds, with different entry and exit points and position size
- Conducted research of single factor and multi-factor models on convertible bonds, refined existing trading algorithms to achieve better performance in back-testing

Deloitte, Shanghai, China   Manager: Mingchao Xin
Data Analysis Intern   June. 2021 – July. 2021
- Conducted financial analysis and fraud detection for clients using data analysis and machine learning with higher efficiency and accuracy compared to traditional methods
- Implemented SQL queries on Microsoft SQL Server Database to perform data analysis and generate insights for clients on their finance conditions
- Engaged in client presentation with Tableau to visualize analysis results

RELEVANT COURSEWORK

- Multivariable Calculus(A), Linear Algebra(A+), Real Analysis, Complex Analysis(A), Abstract Algebra(A), Numerical Analysis(A)
- Probability Theory(A), Statistical Theory(A), Stochastic Process, Data Science Principles(A)
- Machine Learning(A), Deep Learning, Reinforcement Learning

LEADERSHIP AND AWARD

- Member of MUSA (Mathematics Undergraduate Student Association) and Women in Math
- INTEL ISEF 2019: Joined team China in INTEL ISEF 2019 and awarded with finalist
- DEAN'S HONORS LIST - College of Letter & Science 2019|2020|2021|2022
- HONORS TO DATE (Distinction in General Scholarship) 2019|2020|2021|2022