Ronak Mehta

ronakrm.github.io
ronakrm@gmail.com
203-969-5613

Education

Computer Sciences, PhD *2016 - 2022*
University of Wisconsin-Madison
Machine Learning and Computer Vision Research
Thesis: Identifying Feature, Parameter, and Sample Subsets in Machine Learning and Image Analysis
Minor in Statistics

Computer Sciences, MS *2014 - 2016*
University of Wisconsin-Madison
Selected Coursework: Statistical Machine Learning, Computational Statistics, Nonconvex Optimization

Computer Engineering, B.S.E. *2010 - 2014*
University of Michigan-Ann Arbor

Experience

Machine Learning and Theory Scholars Program *Berkeley, CA*
Research Scholar *Summer 2024*

- Working on theoretical and practical solutions for identifying and accounting for worst-case model behaviors.
- Applying classical optimization schemes such as Lipsschitz optimization and mirror descent to find jailbreaks and identify regions in the inut sequence space that may exhibit outlier behaviors.
- Exploring heuristic estimators to identify an abstraction that enables understanding existing estimators such as Gaussian processes, as well others that may better explain how neural-network models aggregate information.

Orca DB, Inc. *Boston, MA*
Member of Technical Staff *September 2023 - Present*

- Founding scientist and engineer building out core ML business solutions and models enabling direct control and interpretability via memory inspection and editing.
- Working on memory augmentation for machine learning models ranging from large language models to simpler classifiers and regression models for non-generative use cases.

Redwood Research *Berkeley, CA*
REMIX Research Resident *January 2023*

- Participated in research program on mechanistic interpretability for large language models.
- Worked on grounding topical mechanistic interpretability methods in theoretical foundations from mainstream machine learning research, connecting ideas in interpretability hypothesis testing to classical probabilistic measures of conditional independence.

Computer Sciences Department, UW-Madison *Madison, WI*
Graduate Research Assistant *2015-2022*

- Collaborated on machine learning and computer vision research projects, with applications in modeling preclinical development of Alzheimer's disease with the Wisconsin Alzheimer's Disease Research Center.
- Focused on Selection Problems in Machine Learning: Which features, samples, or models are minimally sufficient or important based on a specified measure of interest (accuracy, fairness, model size, etc.)
- Publications in a number of top machine learning and computer vision conferences and journals.

Skills

Model Experience: Off-the-shelf LLMs, RNNs (GRUs, LSTMs, Transformers), CNNs (U-Nets, Flow-based methods), Bayesian Methods, Neural Architecture Search, Mixed Effects Regression, Kernel SVMs
Programming Languages: Python, R, C++, MATLAB, Julia, HTML/JavaScript
Scientific Tools: Scikit-Learn, Tensorflow, PyTorch, Lme4, GGPlot, Pandas/NumPy/SciPy