VEDANT SACHDEVA

(732) · 997 · 7488 ◇ sachdved@gmail.com ◇ github.com/sachdved

SUMMARY

Research scientist bridging statistical physics, information theory, and large-scale machine learning. Experienced in representation learning, generative modeling, and diagnostics for high-dimensional data.

EDUCATION

University of Chicago   *August 2017 - June 2022*
Thesis Title: Predictive Strategies in Time-Varying Environments
Adivsors: Arvind Murugan, Stephanie Palmer
Ph.D in Biophysical Sciences
William Rainey Harper Dissertation Fellow
Radix Trading Fellow

Rutgers University, New Brunswick   *September 2013 - May 2017*
Thesis Title: Linking Restricted Boltzmann Machines to Generalized Hopfield Models for Pattern Recognition
Advisor: Anirvan Sengupta (anirvans@physics.rutgers.edu)
B.S. in Physics with Highest Honors, Mathematics - Summa Cum Laude
Presidential Scholar

PROFESSIONAL EXPERIENCES

Senior Data Scientist   Jan '25 - Present
The Trade Desk

· Conceived and developed representation learning frameworks for high-cardinality categorical data using scalable hashing and embedding-based encoding architectures.

· Designed and trained client and audience embedding models that improved downstream forecasting accuracy by 10% and revealed transferable structure for KPI prediction.

· Originated a proposal to redefine forecasting targets, influencing ongoing work by the forecasting team.

· Built end-to-end model and data diagnostics to detect anomalies, drift, and instability in large-scale streaming data, establishing standardized debugging workflows across teams.

· Leveraged multi-GPU distributed systems for high-performance model training, parallel evaluation, and large-scale representation analysis.

Senior Data Scientist   Apr 24' - Jan '25
Valo Health

· Deployed deep learning models, such as transformers and variational autoencoders on patient electronic health record data, typically in a time series structure, to cluster and classify patients using PyTorch, scikit-learn, NumPy, and SciPy.

· Developed tools for datasets that are hundreds of gigabytes large to clean, filter, and select data for training using PySpark, PyArrow, and Polars.

· Evaluated models via classical ML methods, such as Random Forests and XGBoost.

Research Scientist   July '22 - Jan '24
Evozyne

- Trained and analyzed variational autoencoders to understand how latent spaces are organized for protein families using minimal data models to elucidate how VAEs learn, working primarily in Tensorflow and PyTorch
- Demonstrated the relationship between PCA and VAEs, and proposed a method to understand the relationship the covariance structure learned by such models that can be generalized to any dataset.
- Leveraged understanding from the above to propose and rigorously test a curriculum learning-based machine learning method
- Developed neural process regression as a potential generative model for new proteins
- Developing discrete diffusion as a potential generative model for new proteins
- Developed several analysis pipelines used by other members of the team to perform QC on data streams

Doctoral Researcher   Feb '18-June '22
University of Chicago - PhD Program in Biophysical Sciences

- Utilized information theoretic measures to make predictions on time-series data
- Used probabilistic graph models and Markov Chain Monte Carlo methods to model antibody binding affinities
- Analyzed high throughput sequencing data to identify mutations and explore selection functions for an experimental yeast system
- performed evolution experiments in novel experimental yeast system, OrthoRep
- Published 3 first author papers, with 1 more in preparation
- Co-referred a paper for Physical Review X
- Classroom experience with GPU optimization, and CUDA C++

Visitor   June '17-August '17
Center for Computational Biology - Flatiron Institute

- Used statistical physics tools to explore phase transitions in machine learning algorithms
- Participated in journal clubs connecting machine learning, neuroscience, and physics

PUBLICATIONS

K.B. Husain*, **V. Sachdeva***, R. Ravasio, M. Peruzzo, W. Liu, B.H. Good, A. Murugan. Direct and Indirect Selection a Proofreading Polymerase. Published to Biorxiv in October 2024.

M. Falk, J. Wu, A. Matthews, **V. Sachdeva**, N. Pashine, M. Gardel, S. Nagel, A. Murugan. Learning to learn: Non-equilibrium design protocols for adaptable materials. Published in Proceedings of the National Academy of Sciences in June 2023.

V. Ngampruetikorn*, **V. Sachdeva***, J. Torrence, J. Humplik, D. Schwab, S.E. Palmer. Inferring Couplings in Networks Across Order-Disorder Phase Transitions. Published in Physical Review Research in June 2022.

V. Sachdeva, A. Walczak, T. Mora, S.E. Palmer, Optimal Prediction with Resource Constraints Using the Information Bottleneck. Published in PLoS Computational Biology in March 2021.

V. Sachdeva*, K. Husain* J. Sheng, S. Wang, A. Murugan. Tuning Environmental Timescales to Evolve and Maintain Generalists. Published in Proceedings of the National Academy of Sciences in May 2020.

V. Sachdeva, T. Mora, A. Walczak, S.E. Palmer. Kalman Filtering with Efficient Sensory Encoding Schemes. In Preparation.

SKILLS

Coding: Python, MatLab, some proficiency with C++, R
Mathematics: Information Theory, Computational Modeling, Statistics, Statistical Physics, and Regression Methods
Experiment: Flow cytometry, yeast culturing, cell transformations, and high throughput sequencing
Additional language proficiencies: Hindi and French

PRESENTATIONS AND TALKS

Finding the Optimal Observer Using the Information Bottleneck Framework. Talk. CNRS-UChicago Workshop. April 2021.

Predicting the future from the past in visual object motion: optimal representations of mixed stochastic/deterministic trajectories. Talk. American Physical Society, March Meeting. March 2020.

Predicting the future from the past in visual object motion: Optimal representations of mixed stochastic/deterministic trajectories. Poster. Society for Neuroscience. October 2019.

Tuning Evolution Towards Generalists via Resonant Environmental Cycling. Talk. American Physical Society, March Meeting. March 2019.

Pattern Recognition in Restricted Boltzmann Machines. Poster. Quantitative Biology, June 2017.