NAKUL VERMA

Columbia University ◇ Computer Science Department ◇ New York, USA

verma@cs.columbia.edu ◇ (212) 853-8471 ◇ https://www.cs.columbia.edu/~verma

EDUCATION

University of California, San Diego, La Jolla, California, USA

Doctor of Philosophy	Machine Learning	September 2012
Master of Science	Machine Learning	June 2008
Bachelor of Science	Computer Science	December 2004

EMPLOYMENT

Columbia University, New York City, New York USA　　　　Fall 2017-Present
Teaching Professor

· Teaching graduate level courses in Machine Learning, High-dimensional Statistics and CS theory.
· Heading research work in understanding and exploiting the intrinsic structure in data (eg. manifold or sparse structure) to design effective machine learning algorithms in missing- or weakly labelled data regimes.

Janelia Research Campus, Ashburn, Virginia USA　　　　2013-2017
Research Specialist

· Developing biologically plausible statistical models to understand organism behavior from observational data.
· Conduct machine learning research to understand the mathematical structure of behavior.

Amazon.com Inc., Seattle, Washington USA　　　　2012-2013
Research Scientist

· Developing and analyzing automated models for detecting fraudulent transactions in real time.
· Part of the fraud analytics team; responsible for all live European digital orders.

Yahoo! Labs, Bangalore, India　　　　Summer 2011
Visiting Researcher

· Developed learning models for improved classification and visualization for hierarchical multi-class data.

Qualcomm Incorporated, San Diego, California USA　　　　Summer 2008
Visiting Researcher

· Applying machine learning and data mining concepts to design provably correct algorithms for phone personalization in a memory constrained environment.
· Part of the Advanced Technologies Group in the Office of the Chief Scientist.

HONORS AND AWARDS

Research & professional service related
- NeurIPS one of the top scoring reviewers, 2019
- ICML reviewer award, 2015, 2020
- Best paper, Wireless Health 2012.

Teaching related
- Janelia Teaching Fellowship, 2015.

Programming related
- Second place in UCSD Programming Contest, 2005.
- Fourth place in Southern California Regional ACM Programming Contest, 2004.
- Third place in UCSD Programming Contest, 2004. (top undergraduate finisher)
- Fourth place in UCSD Datamining Contest, 2004.

University of California, San Diego

- Provost honors, 2001-2004.

GRANTS

Collaboratory Fellows Fund   2022
Co-Principal Investigator

· Developing a cross-disciplinary project-based course where teams of students with a diverse skillset work together on creating products or services to address important social problems.
· Received funding to create an interdisciplinary high-impact course.

Provost Teaching and Learning Grant   2020, 2021
Principal Investigator

· Investigating data-driven methodologies to improve teaching and learning in tech-heavy classes.
· Awarded funding to develop ML models to learn effective representations of unstructured student solution data and use it to do predictive analytics to improve student performance.

Alliance Joint Project Grant   2019-20
co-Principal Investigator

· Investigating data-driven methodologies to automatically discover physical laws in mechanics and astrophysics data.
· Awarded seed-funding to conduct high-impact research and foster transatlantic collaborations with researchers.

Data Science Institute Scholars Grant   2019, 2020
Principal Investigator

· Investigating non-traditional embedding techniques for modern large-scale data.
· Awarded funding twice to support students to work on research projects representation learning.

INVITED TALKS

Modern supervised learning techniques for neuroscience

- CAJAL Neuroscience Institute, Champalimaud Foundation (host Kristin Branson), 2023.

Hidden potential of non-Euclidean representations for machine learning

- Bloomberg LP, CTO Datascience Speaker Series (host Kai-Zhan Lee), 2019.

Distance preserving embeddings for Riemannian manifolds

- Carnegie Mellon University, Machine Learning Department (host Aarti Singh), 2015.
- IBM Research, Almaden (host Ken Clarkson), 2014.
- University of Washington, Math Department (host Marina Meila), 2013.
- Yahoo Labs, Bangalore (host Dhruv Mahajan), 2012.

Sample complexity of metric learning

- IBM Research, Almaden (host Ken Clarkson), 2015.
- Machine Learning seminar, Janelia Research, HHMI (host Kristin Branson), 2015.

A tutorial on metric learning with some recent advances

- Bay Area Machine Learning Group (host Tony Tran), 2014.

An introduction to statistical theory of learning

- Neurotheory seminar, Janelia Research, HHMI (host Shaul Druckmann), 2015.

PATENTS

D. Krishnaswamy, R. Daley, N. Verma. Delivery of targeted content related to a learned and predicted future behavior based on spatial, temporal, and user attributed and behavioral constraints. *US Patent Application # 20110282964*, May 2010.

D. Krishnaswamy, N. Verma, V. Bychkovsky. Method and system using keyword vectors and associated metrics for learning and prediction of user correlation of targeted content messages in a mobile environment. *US Patent Application # 20090125462*, May 2009.

D. Krishnaswamy, N. Verma, V. Bychkovsky. Method and system for keyword correlation in a mobile environment. *US Patent Application # 20090125517*, May 2009.

EDUCATION ACTIVITIES

❖ *Courses taught at Columbia*

Machine Learning (COMS 4771) 2017-2024

· Took ownership of the course. One of the most in-demand courses. Have streamlined the course and made multiple sections available every semester.

Unsupervised Learning (COMS 4774) 2018-2024

· Designed a new graduate-level special topics course. This course introduces modern techniques in unsupervised learning.

Data-Driven Design for Social Innovation (COMS 4995) 2024

· Designed a new interdisciplinary course that brings students from various departments to work together on creating products or services to address important real-world social problems.

Analysis of Algorithms (CSOR 4231) 2024

· Offered the introductory Algorithms design course.

Computational Linear Algebra (COMS 3251) 2021-23

· Revamped the classic Linear Algebra course to include practical applications and specific focus for Computer Science audience.

Automata and Complexity Theory (COMS 3261) 2017

· Offered the introductory Computer Science theory course.

❖ *Outreach Activities*

Professor Advisor to the IEEE Student Chapter Columbia 2023–present

· Serving as a professor advisor to help undergraduate and graduate students network with people in industry, academia, and among universities for opportunities and connections.

Guest speaker at Dalton High-School Robotics Club Spring 2022

· Invited speaker to generate interest among high-school students to study data analysis and machine learning and the importance of fundamental mathematics to such quantitative fields.

Guest speaker on Columbia Data Science Society Podcast Summer 2020

· Invited speaker to discuss the successes and challenges of machine learning in modern society, and how young professionals can enhance their datascience and analysis skills.

Served on Columbia ADI faculty panel Fall 18

· Invited for a panel discussion to introduce students to different career paths in academia, help build an understanding of research, and bridge the gap between faculty and students.

Hour of Code Volunteer for Grade-school Students 2017

· Taught grade-school students introductory programming skills once a month.

California Summer School for Math and Science (COSMOS) program Summer 06

· Taught high-school students develop creative multi-media applications, and introductory programming skills.

❖ *Courses taught at UC San Diego*

Teaching Assistant and Course Tutor   2003–2007

· Advanced Data Structures – WI04, SP05, SP06, FA06
· Compiler Construction – WI05, FA05, WI06, FA06, WI07
· Design and Analysis of Algorithms – SP05
· Discrete Mathematics – SP04
· Intro to Computer Science and OOP: Java – WI03, SP03, FA03, FA04, WI07

PROFESSIONAL ACTIVITIES

Reviewer and program committee member:

- Nature Methods
- Journal of Machine Learning Research (JMLR)
- Machine Learning Journal (MLJ)
- IEEE Transactions on Information Theory (T-IT)
- Discrete and Computational Geometry (DCG)
- Journal of the Royal Statistical Society (JRSS)
- Transactions on Machine Learning Research (TMLR)
- Conference on Learning Theory (COLT)
- Neural Information Processing Systems (NeurIPS)
- International Conference on Machine Learning (ICML)
- International Conference on Artificial Intelligence and Statistics (AISTATS)
- Association for the Advancement of Artificial Intelligence (AAAI)
- Neural Computation
- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
- IEEE Transactions on Signal Processing (TSP)
- IEEE Signal Processing Letters (SPL)
- International Journal of Computer Vision (IJCV)
- EURASIP Journal of Wireless Communications and Networking (JWCN)
- ACM SIGKDD Conference on Knowledge Data Discovery and Data Mining (KDD)
- IEEE International Conference on Data Mining (ICDM) High Dimensional Mining.
- Asian Conference on Machine Learning (ACML)
- International Conference on Distributed Computing and Networking (ICDCN)

Grant reviewer for:
- Columbia University's Research Initiatives in Science & Engg. (RISE) competition, 2022
- Columbia University's Data Science and Health Initiative (DASHI) program, 2022
- Israel Science Foundation, 2022

PUBLICATIONS

(1800+ citations; h-index: 19)

E. Ri, F. Lee, N. Verma. Contrastive Loss is All You Need to Recover Analogies as Parallel Lines. Association for Computational Linguistics (ACL) workshop on Representation Learning, 2023.

I. Drori, S. Zhang, Z. Chin, R. Shuttleworth, A. Lu, L. Chen, B. Birbo, M. He, P. Lantigua, S. Tran, G. Hunter, B. Feng, N. Cheng, R. Wang, Y. Hicke, S. Surbehera, A. Raghavan, A. Siemenn, N. Singh, J. Lynch, A. Shporer, N. Verma, T. Buonassisi, A. Solar-Lezama. A Dataset for Learning University STEM Courses at Scale and Generating Questions at a Human Level. *Educational Advances in Artificial Intelligence (EAAI)*, 2023.

X. Yu, N. Verma. Improving Model Training via Self-learned Label Representations. *Computing Research Repository (CoRR) abs/2209.04528*, 2022.

I. Drori, S. Zhang, R. Shuttleworth, L. Tang, A. Lu, E. Ke, K. Liu, L. Chen, S. Tran, N. Cheng, R. Wang, N. Singh, T. Patti, J. Lynch, A. Shporer, N. Verma, E. Wu, and G. Strang. A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level. *Proceedings of National Academy of Sciences (PNAS)*, 2022.

L. Flokas, W. Wu, J. Wang, N. Verma, E. Wu. How I stopped worrying about training data bugs and started complaining. *Proceedings of the Sixth Workshop on Data Management for End-To-End Machine Learning (DEEM)*, 2022.

L. Tang, E. Ke, N. Singh, B. Feng, D. Austin, N. Verma, I. Drori. Solving Probability and Statistics Problems by Program Synthesis at Human Level and Predicting Solvability. *International Conference on Artificial Intelligence in Education (AIED)*, 2022

L. Flokas, W. Wu, Y. Liu, J. Wang, N. Verma and E. Wu. Complaint-Driven Training Data Debugging at Interactive Speeds. *International Conference on Data Management (ACM SIGMOD)*, 2022.

F. Lee, C. Kedzie, N. Verma and K. McKeown. An analysis of document graph construction methods for AMR summarization. *Computing Research Repository (CoRR) abs/2111.13993*, 2021.

I. Drori and N. Verma. Solving Linear Algebra by program synthesis, *Computing Research Repository (CoRR) abs/2111.08171*, 2021.

H. Xing, A. Salleb-Aouissi, N. Verma. Automated Symbolic Law Discovery: A Computer Vision Approach. *Association for the Advancement of Artificial Intelligence (AAAI)*, 2021.

B. Cowgill, F. Dell'Acqua, A. Chaintreau, N. Verma and S. Deng. Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics. *ACM Conference on Economics and Computation (EC)*, 2020.

B. Sauerbrei, J. Guo, J. Cohen, M. Mischiati, W. Guo, M. Kabra, N. Verma, B. Mensh, K. Branson, A. Hantman. Cortical pattern generation during dexterous movement is input-driven. *Nature*, 2019.

A. Lamy, Z. Zhong, A. Menon, N. Verma. Noise-tolerant fair classification. *Neural Information Processing Systems (NeurIPS)*, 2019.

M. Aalto, N. Verma. Metric Learning on Non-Euclidean Representations. *Computing Research Repository (CoRR) abs/1902.01738*, 2019.

D. Im, Y. Jiang, N. Verma. Model-Agnostic Meta-Learning using Runge-Kutta Methods. *Computing Research Repository (CoRR) abs/1910.07368*, 2019.

Y. Jiang, N. Verma. Meta-Learning to Cluster. *Computing Research Repository (CoRR) abs/1910.14134*, 2019.

J. Im, N. Verma, K. Branson. Stochastic Neighbor Embeddings under f-divergences. *Computing Research Repository (CoRR) abs/1811.01247*, 2018.

S. Kpotufe, N. Verma. Time-accuracy tradeoffs in Kernel prediction: controlling prediction quality. *Journal of Machine Learning Research (JMLR)*, 2017.

N. Verma, K. Branson. Sample complexity of learning Mahalanobis distance metrics. *Neural Information Processing Systems (NIPS)*, 2015.

N. Verma. Distance preserving embeddings for general n-dimensional manifolds. *Journal of Machine Learning Research (JMLR)*, 2013.

B. Milosevic, J. Yang, N. Verma, S. Tilak, P. Zappi, E. Farella, L. Benini, T. Rosing. Efficient energy management and data recovery in sensor networks using latent variables based tensor factorization. *Sixteenth IEEE International Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, 2013.

K. Patrick, B. Griswold, G. Norman, F. Raab, J. Kerr, B. Demchak, I. Krueger, S. Dasgupta, N. Nikzad, C. Ziftci, N. Verma, P. Zappi, E. Bales, N. Quick. Mobile Technologies and the Exposome: Continuous Assessment of Environmental Exposures Critical to Health. *Lecture Notes on Wireless Healthcare Research*, pp. 25–32, 2013.

N. Verma. Learning from data with low intrinsic dimension. Ph.D. thesis, *Dept. of Computer Science and Engineering, UC San Diego*, 2012.

N. Verma. Distance preserving embeddings for low-dimensional manifolds. *Conference on Learning Theory (COLT)*, 2012.

N. Verma, D. Mahajan, S. Sellamanickam, V. Nair. Learning hierarchical similarity metrics. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

N. Nikzad, N. Verma, C. Ziftci, E. Bales, N. Quick, P. Zappi, K. Patrick, S. Dasgupta, I. Krueger, T. Rosing, W. Griswold. CitiSense: Improving geospatial environmental assessment of air quality using a wireless personal exposure monitoring system. *Wireless Health*, 2012. (**best paper**)

C. Ziftci, N. Nikzad, N. Verma, P. Zappi, E. Bales, I. Krueger, W. Griswold. Citisense: mobile air quality sensing for individuals and communities. *Conference on Systems, Programming, and Applications: Software for Humanity*, 2012.

N. Verma. A note on random projections for preserving paths on a manifold. *UC San Diego Tech. Report CS2011-0971*, 2011.

N. Verma, P. Zappi, T. Rosing. Latent variables based data estimation for sensing applications. *IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, 2011.

B. Babenko, N. Verma, P. Dollar, S. Belongie. Multiple instance learning with manifold bags. *Twenty-Eighth International Conference on Machine Learning (ICML)*, 2011.

N. Nikzad, C. Ziftci, P. Zappi, N. Quick, P. Aghera, N. Verma, B. Demchak, K. Patrick, H. Shacham, T. Rosing, I. Krueger, W. Griswold, S. Dasgupta. CitiSense - Adaptive Services for Community-Driven Behavioral and Environmental Monitoring to Induce Change. *UC San Diego Tech. Report CS2011-0961*, 2011.

N. Verma, S. Kpotufe, S. Dasgupta. Which spatial trees are adaptive to intrinsic dimension?. *Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI)*, 2009.

N. Verma. Mathematical advances in manifold learning. *UC San Diego Tech. Report*, 2008.

Y. Freund, S. Dasgupta, M. Kabra, N. Verma. Learning the structure of manifolds using random projections. *Twenty-First Conference on Neural Information Processing Systems (NIPS)*, 2007.

S. Dasgupta, D. Hsu, N. Verma. A concentration theorem for projections. *Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI)*, 2006.