

Niladri S. Chatterji

CONTACT INFORMATION	Department of Computer Science Stanford University	https://niladri-chatterji.github.io/ niladri@cs.stanford.edu
RESEARCH INTERESTS	<p>I am interested in theoretically understanding why current deep learning systems perform well, with an eye on improving them further.</p> <p>Generalization and Optimization of Overparameterized Models. Many neural network models generalize well despite perfectly fitting noisy training data. I am interested in understanding what leads to successful optimization and generalization.</p> <p>Robust Deep Learning. Many deep learning models are sensitive shifts in the training distribution. I develop techniques to build robust models.</p> <p>Sample and Model Size Tradeoffs. Large-scale models are trained on internet-scale data and need to tradeoff between the model size and the number of samples under a compute budget. My focus is to develop a theory to characterize these tradeoffs and to help make training more efficient.</p>	
CURRENT POSITION	SAIL Postdoctoral Fellow, Stanford University Advisors: Tatsunori Hashimoto & Percy Liang	(June 2021 – Present)
EDUCATION	University of California Berkeley Ph.D. in Physics Advisor: Peter Bartlett Thesis: Why do gradient methods work in optimization and sampling? Indian Institute of Technology Bombay Dual Degree (B.Tech. & M.Tech.) in Engineering Physics (Major) and Electrical Engg. (Minor).	(September 2015 – May 2021) (August 2010 – May 2015)
AWARDS	Stanford SAIL Postdoctoral Fellowship, 2020 Simons Berkeley Research Fellowship (declined), 2020 Institute Silver Medal, IIT Bombay, 2015 Institute Academic Prize, IIT Bombay, 2012–2014	
WORK EXPERIENCE	Research Intern, Google Brain Advisors: Behnam Neyshabur & Hanie Sedghi Research Intern, National Research Institute Japan Advisor: Hisao Nakamura	(Summer 2019) (Summer 2013)
JOURNAL PAPERS	<p>The interplay between implicit bias and benign overfitting in two-layer linear networks. Niladri Chatterji, Philip Long, Peter Bartlett. Journal of Machine Learning Research (JMLR), 2022.</p> <p>Foolish crowds support benign overfitting. Niladri Chatterji, Philip Long. Journal of Machine Learning Research (JMLR), 2022.</p>	

Oracle lower bounds for sampling algorithms.
Niladri Chatterji, Peter Bartlett, Philip Long.
Bernoulli, 2022.

Is there an analog of Nesterov acceleration for gradient-based MCMC?
Yi-An Ma, Niladri Chatterji, Xiang Cheng, Nicolas Flammarion, Peter Bartlett, Michael Jordan.
Bernoulli, 2021.

When does gradient descent with logistic loss find interpolating two-layer networks?
Niladri Chatterji, Philip Long, Peter Bartlett.
Journal of Machine Learning Research (JMLR), 2021.

Finite-sample analysis of interpolating linear classifiers in the overparameterized regime.
Niladri Chatterji, Philip Long.
Journal of Machine Learning Research (JMLR), 2021.

Enhancement of spin-transfer torque switching via resonant tunneling
Niladri Chatterji, Ashwin Tulapurkar, Bhaskaran Muralidharan.
Applied Physics Letters, 2014.

CONFERENCE PAPERS

Benign overfitting without linearity: Neural network classifiers trained by gradient descent for noisy linear data.
Spencer Frei, Niladri Chatterji, Peter Bartlett.
Conference on Learning Theory (COLT), 2022.

Is importance weighting incompatible with interpolating classifiers?
Ke Alexander Wang*, Niladri Chatterji*, Saminul Haque, Tatsunori Hashimoto.
International Conference on Learning Representations (ICLR), 2022.

On the theory of reinforcement learning with once-per-episode feedback.
Niladri Chatterji*, Aldo Pacchiano*, Peter Bartlett, Michael Jordan.
Advances in Neural Information Processing Systems (NeurIPS), 2021.

When does gradient descent with logistic loss interpolate using deep networks with smoothed ReLU activations?
Niladri Chatterji, Philip Long, Peter Bartlett.
Conference on Learning Theory (COLT), 2021.

OSOM: A simultaneously optimal algorithm for multi-armed and linear contextual bandits.
Niladri Chatterji, Vidya Muthukumar, Peter Bartlett.
International Conference on Artificial Intelligence and Statistics (AISTATS), 2020.

Langevin Monte Carlo without smoothness.
Niladri Chatterji*, Jelena Diakonikolas*, Michael Jordan, Peter Bartlett.
International Conference on Artificial Intelligence and Statistics (AISTATS), 2020.

The intriguing role of module criticality in the generalization of deep networks.
Niladri Chatterji, Behnam Neyshabur, Hanie Sedghi.
International Conference on Learning Representations (ICLR), 2020. **(Spotlight Talk)**

Online learning with kernel losses.
Niladri Chatterji*, Aldo Pacchiano*, Peter Bartlett.
International Conference on Machine Learning (ICML), 2019. **(Long Talk)**

Underdamped Langevin MCMC: A non-asymptotic analysis.
 Xiang Cheng*, Niladri Chatterji*, Peter Bartlett, Michael Jordan.
 Conference on Learning Theory (COLT), 2018.

On the theory of variance reduction for stochastic gradient Monte Carlo.
 Niladri Chatterji, Nicolas Flammarion, Yi-An Ma, Peter Bartlett, Michael Jordan.
 International Conference on Machine Learning (ICML), 2018.

Alternating minimization for dictionary learning: Local convergence guarantees.
 Niladri Chatterji, Peter Bartlett.
 Advances in Neural Information Processing Systems (NeurIPS), 2017.

PREPRINTS

Understanding scaling laws: Sample and model size tradeoffs in two-layer neural networks.
 Niladri Chatterji, Percy Liang, Tatsunori Hashimoto.
 In preparation, 2022.

Deep linear networks can benignly overfit when shallow ones do.
 Niladri Chatterji, Philip Long.
 arXiv preprint, 2022. Under review at JMLR.

Undersampling is a minimax optimal robustness intervention in nonparametric classification.
 Niladri Chatterji*, Saminul Haque*, Tatsunori Hashimoto.
 arXiv preprint, 2022. Under review.

Random feature amplification: Feature learning and generalization in neural networks.
 Spencer Frei, Niladri Chatterji, Peter Bartlett.
 arXiv preprint, 2022. Under review at JMLR.

On the opportunities and risks of foundation models.
 Rishi Bommasani and Niladri Chatterji et al.
 arXiv preprint, 2022. Under review at JMLR.

Sharp convergence rates for Langevin dynamics in the nonconvex setting.
 Xiang Cheng, Niladri Chatterji, Yasin Abbasi Yadkori, Peter Bartlett, Michael Jordan.
 arXiv preprint, 2018.

THESIS

Why do gradient methods work in optimization and sampling?
 Niladri Chatterji.
 Ph.D. Thesis, University of California Berkeley, 2021.

TEACHING EXPERIENCE

Linear Models (Statistics, UC Berkeley)	Spring 2020
<i>Graduate Student Instructor</i>	
Electricity and Magnetism (Physics, UC Berkeley)	Fall 2016, Spring 2017
<i>Graduate Student Instructor</i>	
Undergraduate Physics (Physics, UC Berkeley)	Summer 2016, Spring 2016, Fall 2015
<i>Graduate Student Instructor</i>	
Basic Electronics Laboratory (Physics, IIT Bombay)	Fall 2014
<i>Teaching Assistant</i>	
Electricity and Magnetism (Physics, IIT Bombay)	Spring 2014
<i>Teaching Assistant</i>	
Quantum Mechanics (Physics, IIT Bombay)	Fall 2013
<i>Teaching Assistant</i>	
Introduction to Numerical Analysis (Mathematics, IIT Bombay)	Summer 2012

Teaching Assistant

SERVICE

Journal Review

Journal of Machine Learning Research (JMLR), Bernoulli, Annals of Applied Probability, Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques, SIAM Journal on the Mathematics of Data Science.

Conference Review

NeurIPS (2018–), ICML (2020–), COLT (2018–), AISTATS (2019, 2020), ALT (2018, 2021), ICLR (2021), L4DC (2020, 2021).

October 19, 2022