Title: Simplicity Bias in Deep Learning

Speaker: Prateek Jain, adjunct faculty (CSE, IITK)

Date and Time: 18th February 2022 (Friday), 4:00 PM

Venue: Online

Abstract:

While deep neural networks have achieved large gains in performance on benchmark datasets, their performance often degrades drastically with changes in data distribution encountered during real-world deployment. In this work, through systematic experiments and theoretical analysis, we attempt to understand the key reasons behind such brittleness of neural networks in real-world settings.

More concretely, we demonstrate through empirical+theoretical studies that (i) neural network training exhibits "simplicity bias" (SB), where the models learn only the simplest discriminative features and (ii) SB is one of the key reasons behind non-robustness of neural networks. We will then briefly outline some of our (unsuccessful) attempts so far on fixing SB in neural networks illustrating why this is an exciting but challenging problem.

About the speaker:

Prateek Jain is a Senior Staff Research Scientist at Google AI as well as an adjunct faculty member with our department (CSE, IITK) as well as an alumnus of IIT Kanpur (BT CSE 2004). He was previously a Senior Principal Researcher at Microsoft Research India and obtained his PhD from UT Austin. Prateek works in the areas of large-scale and non-convex optimization, high-dimensional statistics, and ML for resource-constrained devices. His works have won multiple awards including best student paper awards at ICML 2007 and CVPR 2008 and the 2020 best paper award by the IEEE Signal Processing Society. Prateek has also received the Young Alumnus Award 2021 from IIT Kanpur and the ACM India Early Career Researcher Award 2021. He regularly serves on the senior program committees of top ML conferences, is an action editor for JMLR, and an associate editor for the SIMODS journal.