“Towards Efficient Learning under Label Noise: From Dawid-Skene to Deep Neural Networks”

Friday, June 30, 2023 • 2:00PM • CREOL 102/103 • Zoom

Abstract:
With the rise of the AI revolution, there is an exceptionally high demand for data labeling. However, most data labeling is done through crowdsourcing techniques, where non-expert annotators often provide labels. This results in considerably noisy labels, which severely degrades the performance of AI systems trained on such data. This talk introduces a range of performance-guaranteed approaches aimed at efficiently learning from noisy crowdsourced labels. The talk begins by revisiting the Dawid-Skene model---one of the most influential models in crowdsourcing---and demonstrates how to advance learning this model, addressing critical aspects such as model identifiability, scalability, sample efficiency, and provable algorithms. The talk also covers how to leverage end-to-end deep learning techniques to enhance the robustness of the learning systems in adverse crowdsourcing scenarios including those with dependent annotators, no expert annotators, or annotators having data-dependent confusions. In essence, the talk offers an interesting exploration into finding synergy between classical factorization tools and deep learning methods in this domain.

Bio:
Shahana Ibrahim is a Ph.D candidate at the School of Electrical Engineering and Computer Science at Oregon State University. She earned her Master's degree from Oregon State University in 2019 and her Bachelor's degree from the National Institute of Technology, Calicut, India, in 2012. Her research interests lie in developing methods and theories for robust machine learning systems, with a particular focus on weakly supervised learning problems, such as learning under label noise, observation errors, and missing data. Her work also spans the areas of tensor decomposition, nonnegative matrix factorization, and stochastic optimization algorithms, and has been published in top-tier venues, including NeurIPS, ICML, ICLR, and IEEE TSP. She has also held positions in the industry, including at NVIDIA and Texas Instruments.