Automated Scalable Bayesian Inference via Data Summarization

Tamara Broderick, PhD

ITT Career Developments Assistant Professor
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Abstract: Bayesian methods are attractive for analyzing large-scale data due to in part to their coherent uncertainty quantification, ability to model complex phenomena, and ease of incorporating expert information. Many standard Bayesian inference algorithms are often computationally expensive, however, so their direct application to large datasets can be difficult or infeasible. Other standard algorithms sacrifice accuracy in the pursuit of scalability. We take a new approach. Namely, we leverage the insight that data often exhibit approximate redundancies to instead obtain a weighted subset of the data (called a "coreset") that is much smaller than the original dataset. We can then use this small coreset as input to existing Bayesian inference algorithms without modification. We provide theoretical guarantees on the size and approximation quality of the coreset. In particular, we show that our method provides geometric decay in posterior approximation error as a function of coreset size. We validate on both synthetic and real datasets, demonstrating that our method reduces posterior approximation error by orders of magnitude relative to uniform random subsampling.

Bio: Tamara Broderick is the ITT Career Development Assistant Professor in the Department of Electrical Engineering and Computer Science at MIT. She is a member of the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL), the MIT Statistics and Data Science Center, and the Institute for Data, Systems, and Society (IDSS). She completed her Ph.D. in Statistics at the University of California, Berkeley in 2014. Previously, she received an AB in Mathematics from Princeton University (2007), a Master of Advanced Study for completion of Part III of the Mathematical Tripos from the University of Cambridge (2008), an MPhil by research in Physics from the University of Cambridge (2009), and an MS in Computer Science from the University of California, Berkeley (2013). Her recent research has focused on developing and analyzing models for scalable Bayesian machine learning. She has been awarded an NSF CAREER Award (2018), a Sloan Research Fellowship (2018), an Army Research Office Young Investigator Program award (2017), Google Faculty Research Awards, the ISBA Lifetime Members Junior Researcher Award, the Savage Award (for an outstanding doctoral dissertation in Bayesian theory and methods), the Evelyn Fix Memorial Medal and Citation (for the Ph.D. student on the Berkeley campus showing the greatest promise in statistical research), the Berkeley Fellowship, an NSF Graduate Research Fellowship, a Marshall Scholarship, and the Phi Beta Kappa Prize (for the graduating Princeton senior with the highest academic average).

MIDAS gratefully acknowledges Wacker Chemie AG for supporting the MIDAS Seminar Series.  

For more information see: http://midas.umich.edu/events