Private On-device Learning via Bayesian Coresets

Dionysis Manousakas, Cecilia Mascolo, Rik Sarkar and Trevor Campbell

We lay theoretical foundations for big data summarization mechanisms that allow third-parties to efficiently store and run inference on sensitive data in memory constrained environments and at the same time ensure that the privacy of each individual contributing their information is not compromised. Combining the frameworks of Bayesian coresets and differential privacy, we propose an automated summarization methodology that enables scalable Bayesian data analysis applications with coherent posterior uncertainty quantification. We validate on synthetic and real datasets, demonstrating high-quality posterior approximations and significant reduction of inference cost.