Departamento Académico de Estadística
Río Hondo # 1, Col. Progreso Tizapán,
Ciudad de México, México C.P. 01080

Seminario Aleatorio

Sesión 332

Automated Scalable Bayesian Inference via Data Summarization

Tamara Broderich,

Statistics and Data Science Center at MIT, USA

Resumen

The use of Bayesian methods in large-scale data settings is attractive because of the rich hierarchical relationships, uncertainty quantification, and prior specification these methods provide. 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 in 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.

Viernes 28 de septiembre de 2018, 13:00 hrs.
Aula B-2, Plantel Río Hondo

El Seminario Aleatorio está destinado tanto a profesores como a estudiantes, por lo que el Departamento de Estadística agradece a los profesores que colaboren invitando a sus alumnos a estas sesiones.

En la red: http://estadistica.itam.mx/es/seminario-aleatorio-de-estad%C3%ADstica