



Department of Decision Sciences

Statistics Seminar

Bayesian cluster analysis: Point estimation and credible balls

Sara Wade

The University of Warwick

Thursday, 31st March 2016

12:30pm Room 3-E4-SR03 Via Röntgen 1 Milano

Abstract

Clustering is widely studied in statistics and machine learning, with applications in a variety of fields. As opposed to classical algorithms which return a single clustering solution, Bayesian nonparametric models provide a posterior over the entire space of partitions, allowing one to assess statistical properties, such as uncertainty on the number of clusters. However, an important problem is how to summarize the posterior; the huge dimension of partition space and difficulties in visualizing it add to this problem. In a Bayesian analysis, the posterior of a real-valued parameter of interest is often summarized by reporting a point estimate such as the posterior mean along with 95% credible intervals to characterize uncertainty. In this paper, we extend these ideas to develop appropriate point estimates and credible sets to summarize the posterior of clustering structure based on decision and information theoretic techniques.