Infinite Latent Feature Models and the Indian Buffet Process

Youyuan Zhang
Department of Statistical Science
Duke University
youyuan.zhang@duke.edu

Lin Xiao
Department of Statistical Science
Duke University
lin.xiao@duke.edu

Abstract

By unveiling the latent variables that can generate observed properties of objects is one of the most fundamental issues in unsupervised learning. However, one of the crucial problems of unsupervised learning algorithms is to detect the latent structure. In K-means problem for example, we need to determine the number of clusters. One classic way is by performing model selection, while the other way is to use a Bayesian nonparametric method. One important method of Bayesian nonparametric method is the Indian buffet process (IBP), which is a stochastic process that provides a probability distribution over equivalence classes of binary matrices of bounded rows and potentially infinite columns. In this report, we first implement the Indian buffet process by Gibbs sampling and Metropolis-Hasting algorithm in python. For improvement in efficiency, we perform matrix calculation optimization and utilize JIT, Cython and parallel programming decrease the computation duration. Finally, we use Code Test for checking the validity and effectiveness of our acceleration and optimization.