

MELD

This repository contains a python package that implements *MELD*, a fast moment estimation method for generalized Dirichlet latent variable model. For the details of the method and the model, please see <http://arxiv.org/abs/1603.05324>.

Introduction

MELD stands for *Moment Estimation* for generalized *Latent Dirichlet* variable models.

The model

The generalized latent Dirichlet variable model in MELD assumes the p dimensional observation \mathbf{y}_i is a mixture of k latent components, with mixture weight denoted as a Dirichlet distributed latent variable \mathbf{x}_i . This model is also known as a *mixed membership model*. In contrast to previous mixed membership models, the new model allows each coordinate of \mathbf{y}_i to take different variable types, including categorical, continuous and integer-valued variables.

Parameter estimation

The parameter estimation method developed in MELD uses a moment method known as generalized method of moments (GMM). This method is in contrast to previous parameter estimation approaches such as MCMC or EM algorithms that require instantiations of latent variables. The new GMM approach does not require instantiations of latent variables. By encoding each coordinate of \mathbf{y}_i by a dummy variable, our new approach calculates the *cross* moment matrices or tensors among variables, with latent variables effectively marginalized out. Parameter estimation is conducted using a fast coordinate descent algorithm.

About the folders

- **data** contains the simulated categorical data with $n = \{50, 100, 200, 500, 1000\}$ observations. The number of categorical variables is 20. Each categorical variable could take 4 different categories.
- **code** contains the implementation of MELD and a python script showing how to use MELD.