Introduction

- Clustering separates samples into groups.
- One possible way is to analyze the similarity matrix.
- The matrix gives pairwise similarity between samples.
- A well-known model is spectral clustering.
- Such method could be applied to data of any form.
- There is limited work on generative models for the similarity matrix.
- We propose Block Mixture Model (BMM) for clustering.
- A probabilistic generative model.
- It finds the block-diagonal structure.
- Samples in the same cluster are similar.
- Samples from different clusters are different.

Method

- A similarity matrix $W$ is provided.
- $W_{ij}$ represents the similarity between $x_i$ and $x_j$.
- We generate the K-element cluster indicators:
  \[ z_n \sim \text{Categorical}(\pi) \]
  \[ \pi \sim \text{Dirichlet}(\alpha) \]
- Each matrix element follows a mixture of beta distributions.
  \[ p(W_{ij}|\theta_k, \beta_k, \lambda_k) = \text{lognormal}(a_k + \beta_k | \mu_k, \sigma_k^2) \]
- We assign priors for the beta distributions.

Experiments

- We generate a similarity matrix with structured noise.
- There might be more than one valid clustering solutions.
- Both $W^{(1)}$ and $W^{(2)}$ contain block-diagonal structures.

Conclusions

- We propose Block Mixture Model (BMM).
- Probabilistic generative model.
- It finds the block-diagonal structure.
- Each matrix element follows a mixture of beta distributions.
- We derive variational inference for BMM.
- We conduct experiments on synthetic and real data.
- BMM is more robust to structured noise.
- BMM outperforms spectral clustering in some real datasets.

Acknowledgements

This work was partially supported by NIH/NHLBI grants R01HL089856 & R01HL089857 and by NSF IIS-1546428.