Research Note 2

Jiaxin Hu

Email: jhu267@wisc.edu

Date: 2019.2.21

1 Supervised Clustering with Normal Setting

Settings and Assumptions Consider a tensor $\chi \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ with continuous variables and a covariate matrix of the first mode $X \in \mathbb{R}^{d_1 \times p}$. And the noise in this setting would be i.i.d and follow a normal distribution with equal variance σ^2

Models and Estimation

- **Only consider the supervise part in the loss function**

  Under this idea, we still have three membership matrix A, B, C on each mode of the tensor, however, we want to let A close to X.

  First, we construct the model:

$$\chi = \mathcal{U} \times_1 A^T \times_2 B^T \times C^T + \epsilon$$

  where $\mathcal{U} \in \mathbb{R}^{k_1 \times k_2 \times k_3}$ in which each element μ_{ijk} refers to the mean of each cluster and $A \in \mathbb{R}^{d_1 \times k_1}, B \in \mathbb{R}^{d_2 \times k_2}, C \in \mathbb{R}^{d_3 \times k_3}$ are the membership matrix on each mode.

  Then the optimization of loss function is:

$$\min_{\mathcal{U},A,B,C} \|\chi - \mathcal{U} \times_1 A^T \times_2 B^T \times C^T\| + \lambda \|A - X\|_{0,1,2}$$

  where $\|\cdot\|_{0,1,2}$ refers to different penalty regularization.

- **Consider the covarite matrix as a part of factorization**

  Under this idea, we may only do clustering on direction of mode B, C (the second and third mode). The supervised data on the first mode facilitate the clustering on the other mode.

  Construct the model:

$$\chi = \mathcal{U} \times_1 X^T \times_2 B^T \times C^T + \epsilon$$

  where $\mathcal{U} \in \mathbb{R}^{p \times k_2 \times k_3}$ in which each element μ_{pjk} refers to the mean of each cluster on mode B, C and how it effected by X. And $B \in \mathbb{R}^{d_2 \times k_2}, C \in \mathbb{R}^{d_3 \times k_3}$ are the membership matrix on each mode.

  Then the optimization of loss function is:

$$\min_{\mathcal{U},B,C} \|\chi - \mathcal{U} \times_1 X^T \times_2 B^T \times C^T\|$$

2 Supervised Clustering with Binary Setting

Settings and Assumptions Consider a tensor $\chi \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ with binary variables *i.e.* the elements in $\chi \in \{0, 1\}$ and a covariate matrix of the first mode $X \in \mathbb{R}^{d_1 \times p}$.

Model and Estimation

- **Only consider the supervise part in the loss function**

  The utilization idea of supervised data has introduced above. Now assume the variables in tensor follow Bernoulli distribution and construct a model:

$$
\begin{aligned}
&\chi_{ijk} \sim_{i.i.d} Ber(p_{ijk}),\ \mathbb{E}\chi_{ijk} = p_{ijk},\ P = \{p_{ijk}\} \in \mathbb{R}^{d_1 \times d_2 \times d_3} \\
&logit(P) = log\frac{p_{ijk}}{1 - p_{ijk}} = \mathcal{U} \times_1 A^T \times_2 B^T \times C^T
\end{aligned}
$$

  where $\mathcal{U} \in \mathbb{R}^{k_1 \times k_2 \times k_3}$ in which each element μ_{ijk} refers to the mean of each cluster and $A \in \mathbb{R}^{d_1 \times k_1}, B \in \mathbb{R}^{d_2 \times k_2}, C \in \mathbb{R}^{d_3 \times k_3}$ are the membership matrix on each mode. The optimization would use penalty likelihood function:

$$
\begin{aligned}
&\max_{\mathcal{U},A,B,C} \|\chi \circ \mathcal{U} \times_1 A^T \times_2 B^T \times C^T - log(1 + exp(\mathcal{U} \times_1 A^T \times_2 B^T \times C^T))\|_F \\
&\quad -\lambda \|A - X\|_{0,1,2}
\end{aligned}
$$

- **Consider the covarite matrix as a part of factorization**

  The assumption of tensor and the idea of using convariate matrix substitute the membership matrix are the same with previous part. Construct the model:

$$
\begin{aligned}
&\chi_{ijk} \sim_{i.i.d} Ber(p_{ijk}),\ \mathbb{E}\chi_{ijk} = p_{ijk},\ P = \{p_{ijk}\} \in \mathbb{R}^{d_1 \times d_2 \times d_3} \\
&logit(P) = log\frac{p_{ijk}}{1 - p_{ijk}} = \mathcal{U} \times_1 X^T \times_2 B^T \times C^T
\end{aligned}
$$

  where $\mathcal{U} \in \mathbb{R}^{p \times k_2 \times k_3}$ in which each element μ_{pjk} refers to the mean of each cluster on mode B, C and how it effected by X. And $B \in \mathbb{R}^{d_2 \times k_2}, C \in \mathbb{R}^{d_3 \times k_3}$ are the membership matrix on each mode.

  The optimization would use penalty likelihood function:

$$
\max_{\mathcal{U},B,C} \|\chi \circ \mathcal{U} \times_1 A^T \times_2 B^T \times C^T - log(1 + exp(\mathcal{U} \times_1 X^T \times_2 B^T \times C^T))\|_F
$$

t.b.c