Computing Non-negative Rank

Abstract

1. Introduction

2. Experiments

2.1. Synthetic experiments

In this section our goal is to show that the proposed algorithm discovers the correct non-negative rank (up to approximation) when the underlying dominant assumptions hold on the dataset. We present results under two noise models, Gaussian and Multinomial noise models. The dataset is assumed to be generated under dominant assumption.

Dominant data: Each column of C is generated from a symmetric Dirichlet distribution with hyper-parameter $\frac{1}{2k}$. Columns of B are also generated from Dirichlet by randomly selecting c features and putting weight $\propto \eta_0$ on these features. Concretely, let c be the number of dominant features for each basis vector and η_0 be the sum of the weights of the dominant features in each basis vector. Assume $\eta = (\frac{\eta_0}{1-\eta_0}) \cdot (\frac{d-c}{c})$. For the jth basis vector, we set $\widetilde{\alpha}_j = \mathbf{1}_{d\times 1}$ and multiply the elements of $\widetilde{\alpha}_j$ indexed from $(c(j-1)+1)$ to cj by η. jth column of B, is then generated from a Dirichlet with hyper-parameter $\widetilde{\alpha}_j$. We chose $c = 3, \eta_0 = 0.1$.

Gaussian Noise: Noise matrix N is generated with each entry from $\mathcal{N}(0, \sigma_i)$ where $\sigma_i = \beta \times B_{il}$. B_{il} is the maximum element in ith row of B. This is based on the subset noise. We chose $\beta \in [0, 0.5]$ as the noise level.

Multinomial Noise: The matrices B and C are normalized such that each of their column sums to one. $N_{.j}$, jth column of noise matrix N, is generated by picking m samples (a sample is a 1-of-d coding) from $[1 \cdots d]$ with prob(i) $= (BC)_{ij}$ and taking an average of the samples to find $\tilde{N}_{.j}$. Then $N_{.j}$ is set as $\tilde{N}_{.j} - (BC)_{.j}$. Lower m implies high noise and vice versa, $m \in \{10, 60, 100\}$.

Parameters: $edgeThreshold = 3 \times \frac{n}{s}$. $\gamma = 0.5$. For higher noise level, the number of groups s varies as per the table. In the experiments we consider $d = 1000, n = 2000, k = 20$. Table 1 presents the quality of non-negative rank discovered under Gaussian noise assumption. The rank presented is the median over 20 different datasets generated from the same underlying parameter.

Table 1: Non-negative rank discovered by proposed algorithm under different levels of Gaussian noise. (Is it really non-negative rank, how can we be sure?)

β **(Noise level)**	s **(Number of groups)**	**Discovered rank**
0	200	20
0.1	150	20
0.2	125	19
0.3	95	21
0.4	93	22
0.5	120	21

Table 2: Identifying non-negative rank under Multinomial noise. Smaller m implies higher noise.

m **(sample size)**	s **(Number of groups)**	**Discovered rank**
1000	140	21
500	340	19
200	1000	33

Table 2 presents the non-negative ranks discovered by the algorithm under multinomial assumption.

2.2. Experiments on real data

References

Tsoumakas, Grigorios and Katakis, Ioannis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining*, 3(3), 2006.