Nonnegative matrix factorization and its applications

(Due at mid-night Dec-13-2017)

Instruction: Every group (with up to 3 students) only need to submit ONE copy of files, which should include the source code file, output on the provided test instance, and one report file. In the report file, you can summarize what you observe or difficulties you have. **You will be evaluated based on correctness of your algorithms, accuracy and efficiency of algorithms, performance on applications, and also the report**.

1 Problem description

The nonnegative matrix factorization (NMF) [4, 3] aims to find two nonnegative matrices $W \in \mathbb{R}^{m\times r}$ and $H \in \mathbb{R}^{n\times r}$ such that their product $WH^\top$ approximates a given nonnegative matrix $X \in \mathbb{R}^{m\times n}$. Assuming Gaussian prior on the noise, we can use Euclidean norm and model NMF as

$$\underset{W,H}{\text{minimize}} \frac{1}{2}\|WH^\top - X\|_F^2 + \phi_1(W) + \phi_2(H), \text{ s.t. } W \in \mathbb{R}_+^{m\times r}, H \in \mathbb{R}_+^{n\times r}, \tag{1}$$

where ϕ_1 and ϕ_2 are regularization terms to encourage certain structures of the solution.

As $r \ll \min(m, n)$, NMF can be used for dimension reduction. Due to nonnegativity, it usually has better interpretability than the principal component analysis (PCA) [3]. It has found many applications including hyperspectral data analysis [5], face recognition [1], document clustering [2], and so on.

2 Algorithm design

In general, it is hard [6] to find a global optimal solution of (1). A heuristic method is to alternatingly update W and H by minimizing the objective with respect to one variable while the other one is fixed. Starting from an initial guess (W^0, H^0), we can iteratively update W and H for $k = 0, 1, 2, \ldots$ by

$$W^{k+1} = \operatorname*{arg\,min}_{W \in \mathbb{R}_+^{m \times r}} \frac{1}{2} \|W(H^k)^\top - X\|_F^2 + \phi_1(W) \tag{2}$$

$$H^{k+1} = \operatorname*{arg\,min}_{H \in \mathbb{R}_+^{n \times r}} \frac{1}{2} \|W^{k+1} H^\top - X\|_F^2 + \phi_2(H). \tag{3}$$

If ϕ_1 and ϕ_2 are convex functions, both subproblems in (2) and (3) are convex programs and can be solved to global optimality.

3 Requirements

To have a complete algorithm for (1), we need to solve (2) and (3) for every k. Your **first task** for this project is to make an algorithm (e.g., gradient-type or interior-point method) and implement it to solve the nonnegative least squares problem

$$\min_Z \frac{1}{2} \|AZ - B\|_F^2 + \alpha \|Z\|_1, \text{ s.t. } Z \in \mathbb{R}_+^{p \times n}, \tag{4}$$

where $A \in \mathbb{R}^{m \times p}$ and $B \in \mathbb{R}^{m \times n}$ are two given matrices, and $\|Z\|_1 \triangleq \sum_{i,j} |z_{ij}|$. The **second task** is to make an algorithm and implement it to solve the sum-to-one constrained least squares problem

$$\min_Z \frac{1}{2} \|AZ - B\|_F^2, \text{ s.t. } Z \in \mathbb{R}_+^{p \times n}, \sum_{i=1}^p z_{ij} = 1, \forall j. \tag{5}$$

Your **third task** is to make an algorithm to solve the sparse and sum-to-one constrained NMF:

$$\operatorname*{minimize}_{W,H} \frac{1}{2} \|WH^\top - X\|_F^2 + \alpha \|W\|_1, \text{ s.t. } W \in \mathbb{R}_+^{m \times r}, H \in \mathbb{R}_+^{n \times r}, \sum_{j=1}^r h_{ij} = 1, \forall i = 1, \ldots, n, \tag{6}$$

namely, in (1), we take

$$\phi_1(W) = \alpha \|W\|_1, \quad \phi_2(H) = \iota_{\mathcal{H}}(H)$$

where $\iota_{\mathcal{H}}$ is the indicator function on $\mathcal{H}$ and

$$\mathcal{H} = \left\{ H \in \mathbb{R}^{n \times r} : \sum_{j=1}^r h_{ij} = 1, \forall i = 1, \ldots, n \right\}.$$

The **fourth task** is to apply your algorithm for (6) to one or two applications such as unmixing hyperspectral data and face recognition (by dimension reduction and then classification).

References

[1] David Guillamet and Jordi Vitria. Classifying faces with nonnegative matrix factorization. In *Proc. 5th Catalan conference for artificial intelligence*, pages 24–31, 2002. 1

[2] Ehsan Hosseini-Asl and Jacek M Zurada. Nonnegative matrix factorization for document clustering: A survey. In *International Conference on Artificial Intelligence and Soft Computing*, pages 726–737. Springer, 2014. 1

[3] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999. 1

[4] Pentti Paatero and Unto Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994. 1

[5] V Paul Pauca, Jon Piper, and Robert J Plemmons. Nonnegative matrix factorization for spectral data analysis. *Linear algebra and its applications*, 416(1):29–47, 2006. 1

[6] Stephen A Vavasis. On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization*, 20(3):1364–1377, 2009. 1