Literature Survey
Non-negative Matrix Factorization

Joel A. Tropp

`<jtropp@ticam.utexas.edu>`

Texas Institute for Computational and Applied Mathematics

The University of Texas at Austin

Austin, TX 78712

Multi-DSP

EE 381K

5 March 2003

INTRODUCTION

- Suppose we have a set of N non-negative, d-dimensional data vectors. Form the $d \times N$ matrix $\boldsymbol{X}$ whose columns are the data points.
- We would like to find a set of $K \ll N$ representatives. We hope that linear combinations of representatives will approximate the data well.
- Problem can be expressed as factoring $\boldsymbol{X} \approx \boldsymbol{V}\boldsymbol{H}$ with $\boldsymbol{V}^{d \times K}$ and $\boldsymbol{H}^{K \times N}$.
- Applications: Dimensionality reduction, feature extraction, matrix approximation, compression.

L'ANCIEN RÉGIME

Principal Component Analysis (PCA)

- The goal of PCA is to find the best K-dimensional projection of the data in the least-squares sense.
- How? First, form the (scaled) covariance matrix

$$\boldsymbol{S} = \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{\mu})(\boldsymbol{x}_n - \boldsymbol{\mu})^t.$$

  Then select the top K eigenvectors of $\boldsymbol{S}$ as the basis for the projection subspace.
- Pros: Straightforward to implement; provably optimal in terms of least-squares.
- Cons: The eigenvectors lack intuitive meanings; optimality is achieved through cancelation effects.

Source: R. O. Duda, P. E. Hart and D. G. Stork. *Pattern Classification*. 2nd ed. New York: John Wiley and Sons, 2001.

An Alternative

Non-negative Matrix Factorization (NNMF)

- To avoid cancelation effects, constrain the factors $\boldsymbol{V}$ and $\boldsymbol{H}$ to have non-negative entries.
- Lee and Seung have shown empirically NNMF yields a parts-based representation for images.
- For text, NNMF seems able to distill semantic groups.

Source: D. D. Lee and H. S. Seung. ”Learning the parts of objects by non-negative matrix factorization.” *Nature*. Vol. 401, pp. 788–791. 21 October 1999.

IMPLEMENTATIONS

Least Squares

- Measures quality of approximation with $\|\boldsymbol{X} - \boldsymbol{V}\boldsymbol{H}\|_F$.
- Update rules:

$$\boldsymbol{H}_{jk} \leftarrow \boldsymbol{H}_{jk} \frac{(\boldsymbol{V}^t \boldsymbol{X})_{jk}}{(\boldsymbol{V}^t \boldsymbol{V}\boldsymbol{H})_{jk}} \qquad \boldsymbol{V}_{ij} \leftarrow \boldsymbol{V}_{ij} \frac{(\boldsymbol{X}\boldsymbol{H}^t)_{ij}}{(\boldsymbol{V}\boldsymbol{H}\boldsymbol{H}^t)_{ij}}.$$

Kullback-Leibler Divergence

- Measures quality of approximation with

$$\sum_{i,k} \left(\boldsymbol{X}_{ik} \log \frac{\boldsymbol{X}_{ik}}{(\boldsymbol{V}\boldsymbol{H})_{ik}} - \boldsymbol{X}_{ik} + (\boldsymbol{V}\boldsymbol{H})_{ij} \right).$$

- Update rules:

$$\boldsymbol{H}_{jk} \leftarrow \boldsymbol{H}_{jk} \frac{\sum_i \boldsymbol{V}_{ij} \boldsymbol{X}_{ik} / (\boldsymbol{V}\boldsymbol{H})_{ik}}{\sum_i \boldsymbol{X}_{ij}}$$

$$\boldsymbol{V}_{ij} \leftarrow \boldsymbol{V}_{ij} \frac{\sum_k \boldsymbol{H}_{jk} \boldsymbol{X}_{ik} / (\boldsymbol{V}\boldsymbol{H})_{ik}}{\sum_i \boldsymbol{X}_{ij}}.$$

Source: D. D. Lee and H. S. Seung. ”Algorithms for non-negative matrix factorization.” *NIPS*. pp. 556–562. 2000.