Nonnegative Sparse PCA

Yunfei Wang

Department of Computer Science & Technology
Huazhong University of Science & Technology

November 22, 2012

Table of contents

PCA

The decomposition performed by PCA is a linear combination of the input coordinates where the coefficients of the the principal vectors form a low-dimensional subspace that corresponds to the direction of maximal variance in the data.

$$\begin{aligned} &\max_U \frac{1}{2}\|U^T X\|_F^2 = \max_U \frac{1}{2} Tr[U^T X X^T U] \\ &s.t.\ U^T U = I \end{aligned} \tag{1}$$

PCA is attractive for a number of reasons.

- Maximum variance property provides a way to compress the data with minimal information loss.
- Representation of the data in the projected space is uncorrelated.
- PCA decomposition can be achieved via an eigenvalue decomposition of the data covariance matrix or SVD.

Nonnegative (Semi-Disjoint) PCA

Adding a nonnegativity constraint to PCA:

$$\begin{aligned} &\max_U \frac{1}{2}\|U^T X\|_F^2 \\ &s.t.\ U^T U = I, U \geq 0 \end{aligned} \tag{2}$$

While disjoint principal component may be considered as a kind of sparseness, it is too restrictive for most problems.We wish to allow some overlap among the principal vectors.$\| I - U^T U \|_F^2$ is typically used as a measure for orthonormality and the relaxed version of Eq(2) becomes,

$$\begin{aligned} &\max_U \frac{1}{2}\|U^T X\|_F^2 - \frac{\alpha}{4}\|I - U^T U\|_F^2 \\ &s.t.\ U^T U = I, U \geq 0 \end{aligned} \tag{3}$$

Nonegative Sparse PCA(NSPCA)

While semi-disjoint principal components can be considered sparse when the number of coordinates is small, it may be too dense when the number of coordinates highly exceeds the number of principal vectors. Minimizing the number of non-zero elements directly:

$$\max_U \frac{1}{2}\|U^T X\|_F^2 - \frac{\alpha}{4}\|I - U^T U\|_F^2 - \beta\|U\|_{L_0} \qquad (4)$$
$$s.t.\ U \geq 0$$

The L_0 norm could be relaxed by replacing it with a L_1 term

$$\begin{aligned} &\max_U \frac{1}{2}\|U^T X\|_F^2 - \frac{\alpha}{4}\|I - U^T U\|_F^2 - \beta\|U\|_{L_1} \\ &= \max_U \frac{1}{2}Tr[U^T X X^T U] - \frac{\alpha}{4}Tr[U^T U U^T U - 2U^T U] \\ &\quad - \beta 1^T U 1 - \frac{\alpha}{4}k \end{aligned} \qquad (5)$$
$$s.t.\ U \geq 0$$

Algorithm of NSPCA

The objective function of u_{rs}(the r row and s column of U) is:

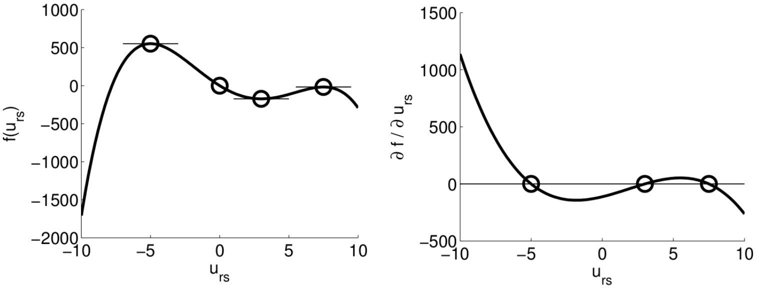
$$f(u_{rs}) = -\frac{\alpha}{4}u_{rs}^4 + \frac{c_2}{2}u_{rs}^2 + c_1 u_{rs} + const \tag{6}$$

where const stands for terms that do not depend on u_{rs}. Setting the derivative with respect to u_{rs} to zero we obtain a cubic equation:

$$\frac{\partial f}{\partial u_{rs}} = -\alpha u_{rs}^3 + c_2 u_{rs} + c_1 = 0 \tag{7}$$

Evaluating Eq.(6) for the nonnegative roots of Eq.(7) and zero, the nonnegative global maximum of $f(u_{rs})$ can be found.

How to find maximum?



(a)4th order polynomial (b) corresponding derivative

In order to find the global nonnegative maximum, the function has to be inspected at all nonnegative extrema (where the derivative is zero) and at $u_{rs} = 0$.

Modifications on LSDA

Generally,we project the data set $X = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n) \in R^{n \times m}$ to d directions and get a set of points $Y = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n)^T \in R^{n \times d}$.

Objective function of within-class graph becomes:

$$\min Tr[Y^T L_w Y] \tag{8}$$

Objective function of between-class graph becomes:

$$\max Tr[Y^T L_b Y] \tag{9}$$

Objective function of sparseness becomes:

$$\min Tr[CY] \tag{10}$$

where C is a $d \times n$ matrix with all ones.
Constraint C: $Y \geq 0$

Finally,we obtain our objective functions:

$$\begin{aligned}\min_{Y} F(Y) =& Tr[Y^T L_w Y] - \alpha Tr[Y^T L_b Y] + \beta Tr[CY] \\ & Tr[Y^T (L_w - \alpha L_b) Y] + \beta Tr[CY]\end{aligned} \tag{11}$$

$$s.t. Y \geq 0$$

$$F(Y_{rs}) = (L_w - \alpha L_b)_{rr} Y_{rs}^2 + \left[\sum_{i=1, i\neq r} (L_w - \alpha L_b)_{ri} Y_{is} + \beta \right] Y_{rs} \tag{12}$$

$$F'(Y_{rs}) = 0 \Longrightarrow Y_{rs} = max\{ \frac{\sum_{i=1,i\neq r} (L_w - \alpha L_b)_{ri} Y_{is} + \beta}{2(\alpha L_b - L_w)_{rr}}, 0\} \tag{13}$$