

带结构的主成分分析

赵知临 胡琪滢 林坤昱

中山大学数据科学与计算机学院

中国科学院大学夏季学期课程011M7012H

2016年7月13日

目录

1 简介

- 问题介绍
- 数学分析

2 算法实现

- 奇异值分解 (SVD)
- 非负矩阵分解 (NMF)
- ADMM求解的非负矩阵分解 (NMF-ADMM)
- 提出的算法

3 实验部分

- 实验结果

4 总结

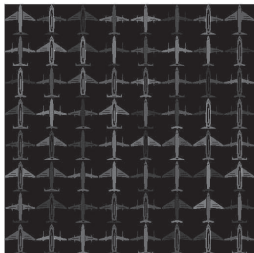
5 参考文献

问题介绍

带结构的主成分分析：主成分分析是找到给定数据的结构信息的主要手段之一。然而，对于某些特殊的输入数据传统的主成分分析，并不能真正得到问题的主成分。这时我们在建模时就需要考虑到这些特殊的结构。请设计合适的主成分分析模型，并提供相应算法，使得当输入为下面两幅图左边信息时，可以返回右边的主成分元素。

问题介绍

Dataset: 64*2-digit numbers*



True PCs



True PCs



数学分析

假设 $V \in \mathbb{R}^{m \times n}$ 是一个描述两张图的高维度矩阵， m 表示的是两张图中特征的数量， n 是样本数量，这里的特征值是每一个像素点的灰度值，区间为 $[0, 1]$ 。为了到寻找两张图的主成分分析，我们利用降维的算法找到 V 的低维度的基矩阵和系数矩阵。

设 k 为基矩阵中基的数量，在实验过程中 $k = 8$ ， $U \in \mathbb{R}^{m \times k}$ 表示的是从高维空间映射到低维空间的基矩阵， $V \in \mathbb{R}^{k \times n}$ 表示的就是每一个样本对应每一个基矩阵的系数，我们将利用四个算法来寻找到最合适的 U 和 M 的值。

奇异值分解 (SVD)

利用奇异值分解法可以用两个较低维度的矩阵 U 和矩阵 M 的乘积来拟合已有的高纬度矩阵 X 。可以用一下目标函数进行表示：

$$\min_{U, M} \sum_{i=1}^m \sum_{j=1}^n \frac{1}{2} l_{i,j} (V_{i,j} - U_{i,j} M_{i,j})^2 + \lambda_u \sum_{i=1}^m \|U_i\|_2^2 + \lambda_m \sum_{j=1}^n \|V_j\|_2^2 \quad (1)$$

其中 λ_u 、 λ_m 是回归系数， $\|\cdot\|_2^2$ 是二范数，用来做函数的回归项。 $l_{i,j}$ 是一个指示函数，如果 $V_{i,j}$ 的值不为0，那么 $l_{i,j}$ 的值为1，反之为0。

奇异值分解 (SVD)

利用梯度下降法不断更新 U 和 V 的值，以得到目标函数值最小时的 U 和 M 的最优解：

$$\begin{aligned}\nabla U_i &:= -\frac{\partial E}{\partial U_i} = \sum_{j=1}^m l_{i,j}((V_{i,j} - UM)M_j) - \lambda_u U_i, i = 1, \dots, n, \\ \nabla M_j &:= -\frac{\partial E}{\partial M_j} := \sum_{i=1}^n l_{i,j}((V_{i,j} - UM)U_i) - \lambda_m M_j, j = 1, \dots, m,\end{aligned}\quad (2)$$

$$U := U - \alpha \nabla U$$

$$M := M - \alpha \nabla M.$$

其中 α 是学习率，控制每一次更新的程度。

非负矩阵分解

SVD的最大缺点在于，它是对整体做了一个主成分分析，而没有考虑到数据是有内部结构的。因为原先的特征值是非负的，低维空间中的基矩阵和相应系数也应是非负的。

基于此，我们使用了非负矩阵分解来进行建模，其最大的特性就是在计算过程中可以保证基矩阵和相应系数值一直非负：

$$\begin{aligned} \min \quad & \|V - UM\|_2^2, \\ \text{s.t.} \quad & U \geq 0, V \geq 0. \end{aligned} \tag{3}$$

非负矩阵分解

不同于使用学习率控制的梯度下降法，NMF中使用的是如下的方式更新两个矩阵：

$$\begin{aligned}U_{i,\alpha} &:= U_{i,\alpha} \frac{(VM^T)_{i,\alpha}}{(UMM^T)_{i,\alpha}}, \\M_{\alpha,j} &:= U_{i,\alpha} \frac{(U^TV)_{\alpha,j}}{(U^TUM)_{\alpha,j}},\end{aligned}\tag{4}$$

这样的更新可以保证只要 U ， M 的初始化值为非负，那么更新过后它们的值可以一直非负。

ADMM求解的非负矩阵分解（NMF-ADMM）

除了一般的NMF，还可以使用交替方向乘子法（ADMM）来求非负的矩阵分解。ADMM 通过分解协调过程，将大的全局问题分解为多个较小、较容易求解的局部子问题，并通过协调子问题的解而得到大的全局问题的解。首先，使用划分将上述形式的目标函数转变为一下形式：

$$\begin{aligned} \min \quad & \frac{1}{2} \|X - V\|_F^2 + I_+(U) + I_+(M) \\ \text{s.t} \quad & X - UM = 0 \end{aligned} \tag{5}$$

其中 I_+ 是标示非负矩阵的指示方程。

ADMM求解的非负矩阵分解 (NMF-ADMM)

ADMM能分解原函数和扩增函数，以便于在更一般的假设条件下并行优化。其更新迭代过程如下：

$$\begin{aligned}(X^{k+1}, U^{k+1}) &:= \arg \min_{X, U \geq 0} (\|X - V\|_F^2 + (\rho/2)\|X - UM^k + U^k\|_F^2) \\ M^{k+1} &:= \arg \min_{M \leq 0} \|X^{k+1} - U^{k+1}M + W^k\|_F^2 \\ W^{k+1} &:= W^k + X^{k+1} - U^{k+1}M^{k+1}\end{aligned}\tag{6}$$

提出的算法

考虑到最终的成像可以看做是在分别在四个向量中选一个向量，然后做加权和来得到每一个块的图像，我们可以将基矩阵和系数矩阵进行划分。令 F_1 为前4个基矩阵， F_2 为后4个基矩阵， G_1 ， G_2 是相应的前4个和后四个系数矩阵。首先，可以写出如下目标函数：

$$\begin{aligned} \text{minimize} \quad & E = \|X - F_1 G_1^T - F_2 G_2^T\|^2 \\ \text{subject to} \quad & G_1 G_1^T = I, \quad G_2 G_2^T = I, \\ & F_1 \geq 0, \quad F_2 \geq 0, \quad G_1 \geq 0, \quad G_2 \geq 0 \end{aligned} \quad (7)$$

其次我们要介绍更新 F_1 和 F_2 的方法。通过目标方程对 F_1 的求导，可以得出：

$$\frac{\partial E}{\partial (F_1)_{ij}} = [(X - F_1 G_1^T - F_2 G_2^T) G_1]_{ij} \quad (8)$$

这样就可以来更新 F_1 :

$$(F_1)_{ij} := (F_1)_{ij} + \eta_{ij}[(X - F_2 G_2^T)G_1 - F_1 G_1^T G_1]_{ij} \quad (9)$$

我们用如下形式来求出学习率 η :

$$\eta_{ij} = \frac{(F_1)_{ij}}{F_1 G_1^T G_1} \quad (10)$$

从而对 F_1 的更新, 可以写作如下形式:

$$(F_1)_{ij} := (F_1)_{ij} \frac{(X - F_2 G_2^T)G_1}{F_1 G_1^T G_1} \quad (11)$$

相似的, 对于 F_2 我们也可以有相应的更新形式:

$$(F_2)_{ij} := (F_2)_{ij} \frac{(X - F_1 G_1^T)G_2}{F_2 G_2^T G_2} \quad (12)$$

接下来考虑 G_1 和 G_2 的更新。我们使用拉格朗日乘子可以讲一个带约束的优化问题变成一个无约束的优化问题，基于此，目标函数可以写作以下形式：

$$\begin{aligned} L &= \|X - F_1 G_1^T - F_2 G_2^T\|^2 + \text{Tr}[\lambda_1(G_1^T G_1 - I)] + \text{Tr}[\lambda_2(G_2^T G_2 - I)] \\ &= \text{Tr}[XX^T - 2X^T(F_1 G_1^T + F_2 G_2^T) + (F_1 G_1^T + F_2 G_2^T)^T(F_1 G_1^T + F_2 G_2^T)] \quad (13) \\ &\quad + \text{Tr}[\lambda_1(G_1^T G_1 - I)] + \text{Tr}[\lambda_2(G_2^T G_2 - I)] \end{aligned}$$

同样的，我们可以计算 目标函数 L 的导数：

$$\frac{\partial L}{\partial G_1} = -2X^T F_1 + 2G_2 F_1^T F_2 + 2G_2 F_2^T F_1 + 2G_1 F_1^T F_1 + 2\lambda_1 G_1 \quad (14)$$

根据KKT条件，我们有如下的方程：

$$(-2X^T F_1 + 2G_2 F_1^T F_2 + 2G_2 F_2^T F_1 + 2G_1 F_1^T F_1 + 2\lambda_1 G_1)_{ij} (G_1)_{ij} = 0 \quad (15)$$

$$(-2X^T F_1 + 2G_2 F_1^T F_2 + 2G_2 F_2^T F_1 + 2G_1 F_1^T F_1 + 2\lambda_1 G_1)_{ij} (G_1)_{ij}^2 = 0 \quad (16)$$

我们可以更新 G_1 ：

$$(G_1)_{ij} := (G_1)_{ij} \sqrt{\frac{(X^T F_1 - G_2 F_1^T F_2 - G_2 F_2^T F_1)_{ij}}{[G_1(F_1^T F_1 + \lambda_1)]_{ij}}} \quad (17)$$

更新 G_2 ：

$$(G_2)_{ij} := (G_2)_{ij} \sqrt{\frac{(X^T F_2 - G_1 F_2^T F_1 - G_1 F_1^T F_2)_{ij}}{[G_2(F_2^T F_2 + \lambda_2)]_{ij}}} \quad (18)$$

我们需要求解出 λ_1 and λ_2 。根据下面的公式以及约束 $G_1 G_1^T = I$ 可以求出 λ_1 ：

$$\lambda_1 = G_1^T (X F_1 - G_2 F_2^T F_1 - G_1 F_1^T F_1) - F_1^T F_1 \quad (19)$$

替换更新 G_1 的公式中的 λ_1 ，我们可以得到：

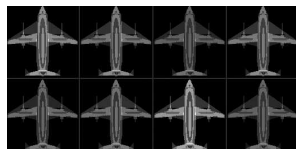
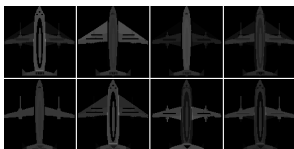
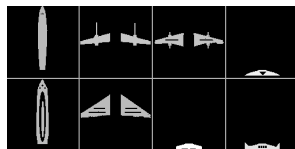
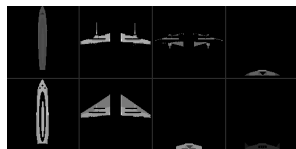
$$(G_1)_{ij} := (G_1)_{ij} \sqrt{\frac{(X^T F_1 - G_2 F_1^T F_2 - G_2 F_2^T F_1)_{ij}}{[G_1 G_1^T (X^T F_1 - G_2 F_1^T F_2 - G_2 F_2^T F_1)]_{ij}}} \quad (20)$$

同样的，我们也可以得到 G_2 的更新形式：

$$(G_2)_{ij} := (G_2)_{ij} \sqrt{\frac{(X^T F_2 - G_1 F_2^T F_1 - G_1 F_1^T F_2)_{ij}}{[G_2 G_2^T (X^T F_2 - G_1 F_2^T F_1 - G_1 F_1^T F_2)]_{ij}}} \quad (21)$$

实验结果







数据1的主成分分析结果，下图顺序为：左上：我们提出的算法，右上：结合ADMM的NMF，左下：NMF，右下SVD。



总结

针对这个带结构的主成分分析问题，我们小组尝试了上述四种方法，分别为NMF，SVD，NMF+ADMM，以及我们自己提出的一个新的算法，其中，前两种算法的效果比较一般，提取出来的主成分与标准答案相差较大，而后两种方法可以得到较好的实验结果，其中基于ADMM的NMF算法效果最好。

参考文献

-  Ma C C, Ma C C. A Guide to Singular Value Decomposition for Collaborative Filtering[J]
-  Ding C, Li T, Jordan M I. Nonnegative Matrix Factorization for Combinatorial Optimization: Spectral Clustering, Graph Matching, and Clique Finding[C], In: Proceedings of the 8th IEEE International Conference on Data Mining (ICDM, pp. 183-192, Pisa, (2008)
-  Lee D D. Algorithms for Non-negative Matrix Factorization[J], Advances in Neural Information Processing Systems, vol. 13(6), pp. 556–562, (2001)
-  Boyd S, Parikh N, Chu E, et al. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers[J], In: Foundations and Trends in Machine Learning, vol. 3(1), pp. 1–122, (2011)
-  Chen M, Lin Z, Ma Y, et al. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices[J], In: Eprint Arxiv, 9, (2009)
-  Ding C, Li T, Peng W, et al. Orthogonal nonnegative matrix t-factorizations for clustering[C], In: Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 126–135, Philadelphia (2006)