

doi:10.3969/j.issn.1673-9833.2019.01.012

# 机器学习中的 PCA 降维方法研究及其应用

孙平安<sup>1, 2</sup>, 王备战<sup>3</sup>

(1. 武夷学院 实验室管理中心, 福建 武夷山 354300;  
2. 武夷学院 认知计算与智能信息处理福建省高校重点实验室, 福建 武夷山 354300;  
3. 厦门大学 软件学院, 福建 厦门 361005)

**摘要:** 机器学习中涉及大量以图像处理为代表的高维数据, PCA 作为有效的数据降维方法常被用于数据预处理阶段。讨论了 PCA 的 K-L 数据转换原理、具体降维处理过程、高维样本协方差矩阵的求解技巧、维数选择方法, 并在 ORL 人脸图样库上给出了基于 PCA 的人脸识别准确度分析。

**关键词:** PCA; K-L 变换; 线性降维; 人脸识别; 机器学习

**中图分类号:** TP391      **文献标志码:** A      **文章编号:** 1673-9833(2019)01-0073-06

**引文格式:** 孙平安, 王备战. 机器学习中的 PCA 降维方法研究及其应用 [J]. 湖南工业大学学报, 2019, 33(1): 73-78.

## A Research on PCA Dimension Reduction with Its Application in Machine Learning

SUN Ping'an<sup>1, 2</sup>, WANG Beizhan<sup>3</sup>

(1. Laboratory Management Center, Wuyi University, Wuyishan Fujian 354300, China;  
2. The Key Laboratory of Cognitive Computing and Intelligent Information Processing of Fujian Education Institutions, Wuyi University, Wuyishan Fujian 354300, China; 3. School of Software, Xiamen University, Xiamen Fujian 361005, China)

**Abstract:** Machine learning involves a large number of high-dimensional data represented by image processing. PCA, as an effective data dimension reduction method, is often applied for data preprocessing. A tentative inquiry has been made into the principle of K-L data conversion, the specific dimension reduction processing, the co-variance matrix of the high dimensional sample and the method of dimension selection, followed by an accuracy analysis of face recognition based on PCA from ORL face pattern database.

**Keywords:** PCA; K-L transform; linear dimension reduction; face recognition; machine learning

## 0 引言

随着大数据以及云计算技术的日渐成熟落地, 人类对图像数据的采集和存储能力得以大幅提升。

在图像的分析处理领域, 这些原始图像数据通常被看作是高维向量集合, 一张分辨率为 64 像素 × 64 像素的图像将其每个像素逐行排列之后可以表示为

收稿日期: 2018-03-27

基金项目: 福建省教育厅科技基金资助项目(JAT160524), 武夷学院转型发展教育教学改革与研究基金资助项目(xj2014018)

作者简介: 孙平安(1977-), 男, 陕西渭南人, 武夷学院教师, 硕士, 主要研究方向为数据库及数据挖掘, 机器学习, 图像处理, E-mail: wyuspa@wuyiu.edu.cn

通信作者: 王备战(1965-), 男, 陕西咸阳人, 厦门大学教授, 博士, 主要从事数据仓库与数据挖掘, 软件体系结构, 数字媒体技术等方面的研究, E-mail: wangbz@xmu.edu.cn

一个4 096维的向量；在机器学习的自然语言处理领域，一篇文章可以被看作一个大的向量；光谱分析处理领域，超光谱成像记录了全部波段的光谱信息，仅仅一个光谱分辨率为128个波段，包含有 $256 \times 256$ 个传感探测单元的超光谱成像，其向量化表示之后的超光谱数据高达 $256 \times 256 \times 128$ 维<sup>[1]</sup>。在数据维数和规模巨大时，传统数据挖掘技术所需的计算资源呈指数级增加，研究高效简易的数据表示方法，降低数据冗余度，建立高效率低成本的数据处理、存储和通信新技术是高维数据分析的一大难题<sup>[2]</sup>。

早在1957年，美国数学家R. Bellman在其介绍动态规划时就已经提出高维数据分析中面临的首要难题是“维度灾难 (curse of dimensionality)”<sup>[3]</sup>问题。事实上，在高维情形下出现的数据样本稀疏、距离计算困难等问题是所有机器学习方法共同面临的严重障碍。对高维数据的处理相当困难，因此缓解维度灾难的一个重要途径是降维，即通过某种数学变换将原始高维空间转变为一个低维“子空间”，并尽量保持高维空间所包含的有用信息。

## 1 数据降维方法分类

从不同的角度将数据降维方法分为以下几类。

1) 线性/非线性。降维后数据在低维空间依然能够保持数据之间的线性关系的称为线性降维，如线性判别分析 (linear discriminant analysis, LDA) 也称为Fisher's Linear Discriminant<sup>[4]</sup>、主成分分析 (primary component analysis, PCA)、局部保留投影 (locality preserving projection, LPP)<sup>[5-6]</sup>等；非线性降维是一类基于核 (Kernel) 或基于流形 (Manifold) 的降维方法，如核主成分分析 (kernel primary component analysis, KPCA)、等距映射 (Isometric Mapping, ISOMAP)<sup>[7]</sup>、局部线性嵌入 (local linear embedding, LLE) 等。

2) 监督/非监督。两者的主要区别在于数据样本是否存在类别标签信息。非监督降维方法的目标是在降维时使信息的损失最小，如PCA、LPP、ISOMAP、LLE、拉普拉斯特征映射 (Laplacian Eigenmaps) 等；监督式降维方法的目标是最大化类别间的辨别信息，如LDA。

3) 全局/局部。局部方法仅考虑样本集合的局部信息，即数据点与临近点之间的关系。局部方法以LLE为代表，还包括Laplacian Eigenmaps、LPP、局部切空间对齐 (local Tangent space alignment, LTSA)<sup>[6]</sup>。全局方法不仅考虑样本几何的局部信息还考虑样本集合的全局信息，即样本点与非临近点之

间的关系。全局算法有PCA、LDA、ISOMAP等。

## 2 主成分分析法

主成分分析思想最初由K. Pearson于1901年提出<sup>[9]</sup>。1933年，H. Hotelling进一步完善PCA的数学基础<sup>[10]</sup>，实质上其理论基础为Karhunen-Loeve变换 (简称K-L变换)。K-L变换以最小均方误差为衡量准则进行数据压缩，是最小均方误差意义下的最优正交变换。

### 2.1 K-L变换的基本原理

在数据处理方面，变换是寻求描述复杂事物的一种更简洁、高效的度量工具，它将复杂事物化成简单事物的组合或对其进行分解，进而分析其组成成分。K-L变换实质上是一种正交变换 (K-L变换有连续变换和离散两种，在此仅讨论离散变换)，即将向量 $\mathbf{x}$ 通过坐标系统变换将其描述为在另一种基向量坐标系下更有利于解决问题的表示。这组新的基向量 $\{\mathbf{u}_i\}$ ， $i=1, 2, \dots, \infty$ 是正交的，其中每个坐标基向量用 $\mathbf{u}_i$ 表示， $i=1, 2, \dots, \infty$ ， $\mathbf{x}$ 可看做被分解的信号，显然被分解后的信号可以看做是各成分之和 (共有无限个成分)，即向量 $\mathbf{x}$ 可表示为

$$\mathbf{x} = \sum_{i=1}^{\infty} c_i \mathbf{u}_i. \quad (1)$$

式中 $c_i$ 为坐标基向量 $\mathbf{u}_i$ 下的成分<sup>[11-12]</sup>。

对向量进行正交变换，可采用的正交坐标系有多种，选用何种坐标系的关键在于要有利于解决问题。式(1)中的无限维在实际中无法实现，为此将式(1)中的无限多维基向量坐标系改为如下有限维：

$$\hat{\mathbf{x}} = \sum_{i=1}^d c_i \mathbf{u}_i, \quad (2)$$

式(2)中 $\hat{\mathbf{x}}$ 为 $\mathbf{x}$ 的近似，则有

$$\mathbf{x} - \hat{\mathbf{x}} = \sum_{j=d+1}^{\infty} c_j \mathbf{u}_j. \quad (3)$$

为了使 $\hat{\mathbf{x}}$ 能够最好地近似表达 $\mathbf{x}$ ，即 $\hat{\mathbf{x}}$ 与 $\mathbf{x}$ 之间的均方误差最小，误差表示为

$$\varepsilon = E \left[ \left( \mathbf{x} - \hat{\mathbf{x}} \right)^T \left( \mathbf{x} - \hat{\mathbf{x}} \right) \right]. \quad (4)$$

找到最小的 $\varepsilon$ 就找到了 $\mathbf{x}$ 的最优近似 $\hat{\mathbf{x}}$ 。这一问题实质是一个条件极值问题，即求最佳的正交变换基 $\mathbf{u}_i$ ， $i=1, 2, \dots, \infty$ ，通常使用拉格朗日乘子法，将条件约束值转换为一个无条件极值问题。

由于 $\mathbf{u}_i$ ， $i=1, 2, \dots, \infty$ 为正交基，故

$$\mathbf{u}_i^T \mathbf{u}_j = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases}$$

将式(3)代入式(4)得:

$$\varepsilon = E \left[ \sum_{i=d+1}^{\infty} c_i u_i \sum_{j=d+1}^{\infty} c_j u_j \right] = E \left[ \sum_{k=d+1}^{\infty} c_k^2 \right]. \quad (5)$$

对于向量  $\mathbf{x}$  在某个坐标系下相应成分  $c_k$  的值, 通过  $\mathbf{x}$  与每一个基向量  $\mathbf{u}_k$  的内积来计算, 由于正交坐标系中各基向量是相互正交的, 故这个内积值实际上就是  $c_k$ , 即  $c_k = \mathbf{u}_k^T \mathbf{x} = \mathbf{x}^T \mathbf{u}_k$ , 代入式(5)得:

$$\varepsilon = E \left( \sum_{k=d+1}^{\infty} \mathbf{u}_k^T \mathbf{x} \mathbf{x}^T \mathbf{u}_k \right) = \sum_{k=d+1}^{\infty} \mathbf{u}_k^T E(\mathbf{x} \mathbf{x}^T) \mathbf{u}_k,$$

令  $\Sigma = E(\mathbf{x} \mathbf{x}^T)$ , 则有,

$$\varepsilon = \sum_{k=d+1}^{\infty} \mathbf{u}_k^T \Sigma \mathbf{u}_k. \quad (6)$$

求  $\varepsilon$  最小就变成一个确保正交变换下的  $\varepsilon$  最小化问题, 构造如下拉格朗日乘子式,

$$g(\mathbf{u}) = \sum_{k=d+1}^{\infty} \mathbf{u}_k^T \Sigma \mathbf{u}_k - \sum_{k=d+1}^{\infty} \lambda_k [\mathbf{u}_k^T \mathbf{u}_k - 1], \quad (7)$$

将式(7)中各向量对  $\mathbf{u}_k$  求导, 并令其为零, 可得:

$$(\Sigma - \lambda_k I) \mathbf{u}_k = 0, k=d+1, \dots, \infty. \quad (8)$$

由线性代数相关知识,  $\mathbf{u}_k$  为矩阵  $\Sigma$  的特征值  $\lambda_k$  所对应的特征向量, 满足  $\Sigma \mathbf{u}_k = \lambda_k \mathbf{u}_k$ , 故有

$$\varepsilon = \sum_{k=d+1}^{\infty} \mathbf{u}_k^T \Sigma \mathbf{u}_k = \sum_{k=d+1}^{\infty} \lambda_k \mathbf{u}_k^T \mathbf{u}_k = \sum_{k=d+1}^{\infty} \lambda_k. \quad (9)$$

也就是, 只要  $\mathbf{x}$  的  $k$  维后的特征值之和尽可能小, 就可使用  $\hat{\mathbf{x}}$  代替  $\mathbf{x}$ 。为此, 把  $\Sigma$  的各特征值按降序排序, 即  $\lambda_1 > \lambda_2 > \dots > \lambda_d > \dots$ , 取其最大的前  $d$  项特征值所对应的特征向量组成坐标系  $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d\}$ , 可使误差  $\varepsilon$  为最小。对向量  $\mathbf{x}$  的 K-L 降维与重构的表达如下:

$$\mathbf{z} = \mathbf{U}^T \mathbf{x}, \quad (10)$$

$$\mathbf{x} = \mathbf{U} \mathbf{z}. \quad (11)$$

## 2.2 PCA 降维过程

K-L 变换为 PCA 提供了理论基础依据。PCA 是一种基于线性变换进行向量降维的方法, 其核心思想是通过坐标旋转 (即寻找新的正交基), 将数据投影到使数据方差最大化的若干个坐标轴上, 得到数据在新空间的表示以消除原数据空间的多重共线性, 从而达到数据降维的目的。

通常, 欲获得低维子空间, 最简单的是对原始高维空间进行线性变换, 给定  $d$  维空间中的样本集  $\mathbf{X} = \{x_1, x_2, \dots, x_m\} \in \mathbf{R}^{d \times m}$ , 变换之后得到  $n$  维空间  $\mathbf{Z}$  中的样本, 其中  $n \ll d$ 。

$$\mathbf{Z} = \mathbf{U}^T \mathbf{X}. \quad (12)$$

式中:  $\mathbf{U} \in \mathbf{R}^{d \times m}$  为变换矩阵 (即找到的新正交基),  $\mathbf{U}$  是由  $\mathbf{X}$  的协方差的特征值最大的前  $n$  项所对应的

特征向量构成的正交矩阵;  $\mathbf{Z} \in \mathbf{R}^{d \times m}$  为样本在新空间中的投影, 是高维  $\mathbf{X}$  降维后的低维近似<sup>[12-15]</sup>。

PCA 降维算法伪代码描述如下。

输入:

样本集  $\mathbf{X} = \{x_1, x_2, \dots, x_m\}$ ,  $\mathbf{x}_i \in \mathbf{R}^d$ ; 低维空间维数  $n$ 。

处理过程如下。

1) 样本特征去均值化 (即数据的中心化)。对样本的每个特征, 使用当前特征的值减去样本集中该特征的均值  $\mu_j$ , 对第  $i$  个样本的第  $j$  个特征,  $\mathbf{x}_i^{(j)} = \mathbf{x}_i^{(j)} - \mu_j$ ,

其中  $\mu_j = \frac{1}{m} \sum_{k=1}^m \mathbf{x}_k^{(j)}$ , 全部特征均值向量为

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_d), \boldsymbol{\mu} \in \mathbf{R}^d;$$

2) 计算样本的协方差矩阵  $\Sigma = \mathbf{X} \mathbf{X}^T$ ;

3) 对协方差矩阵  $\Sigma = \mathbf{X} \mathbf{X}^T$  做特征值分解 (实践中常使用对  $\mathbf{X}$  的 SVD 分解来替代), 取最大的前  $l$  个特征值所对应的特征向量  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_l$  组成转换矩阵  $\mathbf{U}$ ;

4) 生成降维后的样本集  $\mathbf{Z} = \{z_1, z_2, \dots, z_m\}$ ,

$z_i = \mathbf{U}^T \mathbf{x}_i, i=1, 2, \dots, m$  原始样本集的近似  $\hat{\mathbf{D}} = \{\mathbf{U} z_i + \boldsymbol{\mu}\}, i=1, 2, \dots, m$ ;

输出:  $\mathbf{Z} \in \mathbf{R}^{l \times m}$ ;  $\hat{\mathbf{D}} \in \mathbf{R}^{d \times m}$ 。

PCA 仅需要保留特征向量组矩阵  $\mathbf{U}$  与样本的均值向量  $\boldsymbol{\mu}$ , 即可通过简单的向量减法和矩阵-向量的乘法将新样本投影至低维空间中。显然, 低维空间与原始的高维空间必然存在不同, 但实际中舍弃往往是必要的: 一方面, 舍弃部分信息后消除了多重共线性, 使样本的采集密度增大、数据更易使用, 降低很多算法的计算开销, 这也正是降维的重要动机; 另一方面, 当数据受到噪声影响时, 最小的特征值所对应的特征向量往往与噪声有关, 将它们舍弃能在一定的程度上起到去噪的效果。

## 2.3 PCA 中降维维度的选择

降维后的低维空间的维数通常由用户事先指定, 或通过在实验中取不同的值来确定。下面从另一个角度来思考维度选择。

确定主成分。由式(9)可得:

$$\varepsilon = \sum_{k=d+1}^{\infty} \mathbf{u}_k^T \Sigma \mathbf{u}_k = \sum_{k=d+1}^{\infty} \lambda_k \mathbf{u}_k^T \mathbf{u}_k = \sum_{k=d+1}^{\infty} \lambda_k.$$

对于协方差矩阵  $\Sigma = E(\mathbf{x} \mathbf{x}^T)$ , 存在正交变换矩阵  $\mathbf{U}$  将  $\Sigma$  化为对角阵 (也称为合同对角化), 即

$$\mathbf{U}^T \Sigma \mathbf{U} = \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix},$$

其中,  $\lambda_1 > \lambda_2 > \dots > \lambda_p$ , 矩阵  $\mathbf{U}$  的第  $i$  列对应于  $\mathbf{u}_i$ , 相应的  $\varphi_i$  为降维后的第  $i$  主成分<sup>[16]</sup>。

主成分的总方差:

由于  $\text{tr}(A) = \text{tr}(U^T \Sigma U) = \text{tr}(\Sigma U U^T) = \text{tr}(\Sigma)$ , 所

以,  $\sum_{i=1}^p \lambda_i = \sum_{i=1}^p \delta_{i,i}$  或  $\sum_{i=1}^p \text{var}(\varphi_i) = \sum_{i=1}^p \text{var}(\mathbf{x}_i)$ 。

其中,  $\delta_{i,i}$  是协方差矩阵  $\Sigma$  的第  $i$  对角元素。由此可以看出, 主成分分析把向量  $\mathbf{x}$  的  $p$  个维分量的总方差分解成了  $p$  个不相关维分量  $\varphi_1, \varphi_2, \dots, \varphi_p$  方差之和。使用协方差矩阵计算出来的特征值为其对应特征向量上样本集的方差, 方差越大, 说明数据集在该特征向量方向上越分散, 变化越大, 该方向就可以用来作为数据集的主方向。主成分  $\varphi_i$  对总方差的贡献比例

$\lambda_i / \sum_{j=1}^p \lambda_j$  称为主成分  $\varphi_i$  的贡献率, 各主成分贡献率的大小表明它解释  $\mathbf{x}$  的各维分量能力的强弱。PCA 的目的就是为了降低特征的个数, 通常选用具有较大方差的主成分而忽略一些有较小方差的主成分, 这样在降维的同时又不会给总方差带来大的影响。具有较大方差的前  $m$  个主成分在表达向量  $\mathbf{x}$  中的贡献率之

和  $\frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i}$  称为  $\varphi_1, \varphi_2, \dots, \varphi_m$  的累积贡献率, 它表明  $\varphi_1, \varphi_2, \dots, \varphi_m$  解释  $\mathbf{x}$  的能力。通常取使得累积贡献率达某个较高阈值的较小  $m$ , 此时即可用  $\varphi_1, \varphi_2, \dots, \varphi_m$  替代  $\mathbf{x}$  达到降维目的, 且信息损失不多。

### 3 实验过程与分析

#### 3.1 实验原理

一幅  $l \times n$  的图像可以看作一个  $l \times n$  的矩阵, 也可以将其拉长为  $l \times n$  维向量的数据样本。设训练样本集有  $m$  张人脸图像, 即样本集为  $X\{\mathbf{x}_i \in \mathbf{R}^{l \times n}, i=1, 2, \dots, m\}$ 。由于维数太高不利于分析处理, 需要进行特征降维, 使用较少的特征来表示所有样本。所有样本总协方差  $\Sigma = E(\mathbf{X}\mathbf{X}^T)$ , 其中  $\mathbf{X}$  是由所有样本去均值化后样本构成的  $(l \times n) \times m$  的矩阵。总协方差矩阵  $\Sigma$  为一个  $(l \times n) \times (l \times n)$  矩阵, 使用 PCA 降维需求解  $\Sigma$  的正交特征向量组, 因为维数很高计算量巨大, 直接计算比较困难。考察另一个矩阵  $T = \mathbf{X}^T \mathbf{X}$ ,  $T \in \mathbf{R}^{m \times m}$ 。通常  $m \ll (l \times n)$ 。矩阵  $T$  的特征方程为

$$T \mathbf{v}_i = \mathbf{X}^T \mathbf{X} \mathbf{v}_i = \lambda_i \mathbf{v}_i \quad (13)$$

式 (13) 两边同时乘以  $\mathbf{X}$  得,

$$\mathbf{X} \mathbf{X}^T \mathbf{X} \mathbf{v}_i = \lambda_i \mathbf{X} \mathbf{v}_i \quad (14)$$

设  $\mathbf{u}_i = \mathbf{X} \mathbf{v}_i$ , 则式 (14) 变为  $\Sigma \mathbf{u}_i = \lambda_i \mathbf{u}_i$ , 也就是  $\Sigma$  的特征方程。因此  $m \times m$  维矩阵  $\mathbf{X}^T \mathbf{X}$  和  $(l \times n) \times (l \times n)$  矩阵  $\mathbf{X} \mathbf{X}^T$  有相同的特征值, 特征向量具如下关系:

$$\mathbf{u}_i = \mathbf{X} \mathbf{v}_i \quad (15)$$

由于  $\Sigma$  的秩不会大于  $m$ , 因此  $\Sigma$  中最多有  $m$  个非零特征值。通过求解维数较低的矩阵  $\mathbf{X}^T \mathbf{X}$  的特征值与特征向量组实现对样本的 PCA 降维。

#### 3.2 实验预处理

ORL 人脸库包含 40 个人, 每个人包含 10 张人脸样本, 样本大小为  $112 \times 92$ , 展开成 10304 的向量, 组成  $10304 \times 400$  的矩阵。训练过程中, 得到了上述的正交变换矩阵  $\mathbf{U}$ , 用其对新的样本进行变换降维, 把 10304 维的人脸图降到 100 维以下, 以便进一步实施识别。训练样本从每人 10 张脸部图像中随机取 9 张, 其中某人的 9 张训练样本如图 1 所示; 对图像中各特征值进行中心化预处理, 目的是增加基向量的正交性, 结果如图 2 所示; 经过 PCA 降维到 100 维后获得的特征图如图 3 所示; 40 个人的均值脸如图 4 所示, 用于数据中心化预处理过程。

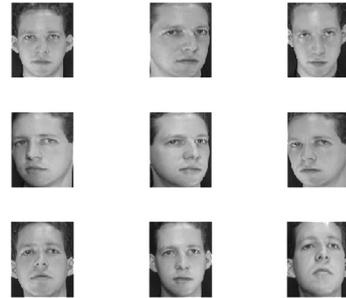


图 1 测试样本

Fig. 1 Training set



图 2 中心化处理后的训练样本

Fig. 2 Centered training set

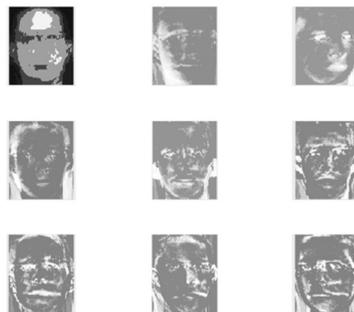


图 3 降维后的特征脸图

Fig. 3 Eigen face features after dimension reduction



图 4 均值脸

Fig. 4 Mean face features

### 3.3 实验过程

实验采用先从 40 个人中动态选人, 再从每个人的 10 张脸部图中随机选图的方式进行。最多使用每人 10 张图中的 9 张作为训练样本, 剩余的图像作为测试样本。实验过程中, 训练样本数分别从  $40 \times 1$  至  $40 \times 9$  (即训练样本从每人 1 张人脸样本逐渐增加到 9 张), 降维后维数从 1 至 100, 测试每种情况下的识别准确率, 其测试结果对比图如图 5 所示。

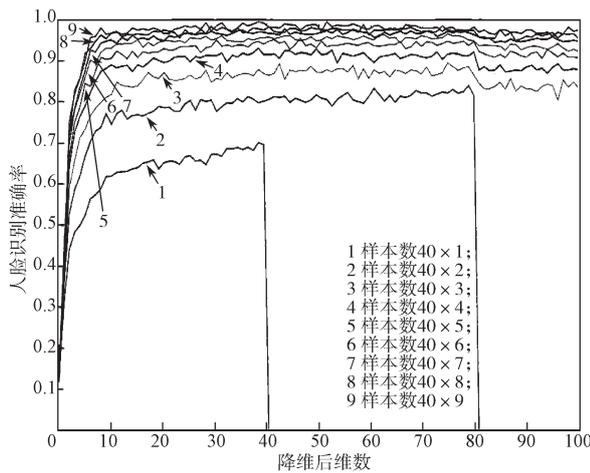


图 5 测试结果对比图

Fig. 5 Comparison of testing results

图 5 中, 当训练样本为 1 和 2 时, 表中数据出现断崖式归零, 是因为当从 40 个人, 每人 10 张脸样本图像中取 1 张、2 张人脸样本作为训练样本时, 特征值分解后的特征矩阵分别为  $40 \times 40$  维和  $80 \times 80$  维, 当降维维数分别超过 40 和 80 时, 实际上已经不是降维而是在升维, 计算将会出错。因此在实验程序中进行了判断处理, 如果出现类似情况直接置准确率为 0。从实验对比图中也可以看出, 当降维后维数超过 80 时识别准确率反而有所下降。

### 3.4 实验结果分析

取每种情况下运行 10 次的识别准确率均值作为其最终结果。本次实验结果为一个  $100 \times 9$  的表格, 因篇幅原因仅展示部分实验结果, 如表 1 所示。

表 1 部分测试结果

Table 1 Selected testing results %

降维后维数	训练样本数 / (人数·张数·人 <sup>-1</sup> )				
	40 × 1	40 × 3	40 × 5	40 × 7	40 × 9
30	67.2	84.8	93.5	97.2	97.2
31	67.3	85.7	95.1	96.7	98.5
32	66.8	85.5	93.6	96.5	99.3
33	68.7	87.2	93.9	96.2	97.7
34	68.6	85.8	93.5	96.7	97.2
35	68.9	86.5	93.9	97.3	98.0
36	68.4	86.9	93.9	97.4	99.2
37	69.0	86.9	93.4	96.9	98.5
38	68.2	86.4	94.4	96.7	99.0
39	70.1	87.9	94.7	97.1	99.5
40	69.6	86.9	94.1	97.0	99.2
41	0.0	86.5	94.8	96.7	98.2
42	0.0	87.0	95.2	96.6	98.2
43	0.0	89.3	94.3	96.3	98.0
44	0.0	85.5	94.7	97.8	99.0
45	0.0	87.8	92.8	96.5	98.8

通过实验可知, 训练样本为  $40 \times 3$  时, 即随机取每人 3 张图像作为训练样本, 共 120 张脸图, 当降维后维度为 30 时基于 PCA 降维的人脸识别就已经能达到 84% 以上的准确率; 如果将 10 张人脸样本中的 9 张作为训练样本, 则 1 张作为测试样本可达到 98% 以上的准确率。

使用数据降维工具箱 Drtoolbox<sup>[17]</sup>, 将 900 个 3 维样本降为 1 维, 8 种降维算法时间对比结果见表 2。

表 2 几种降维算法性能比较

Table 2 Comparison of different dimension reduction algorithm performance

算法	线性 / 非线性	全局 / 局部	耗时 / s
PCA	线性	全局	0.032 1
LDA	线性	全局	0.002 9
LPP	线性	全局	0.099 6
ISOMAP	非线性	全局	1.707 3
LLE	非线性	局部	0.230 8
Laplacian	非线性	局部	0.139 3
LTSA	非线性	局部	0.319 9
MVU	非线性	全局	1.206 5

表 2 的比较结果显示, PCA 在高维数据降维方面相当有效, 仅比常用于分类的有监督的线性判别算法 LDA 慢而优于其他算法。

## 4 结论

降维技术作为高维数据预处理步骤, 在将数据应用到其他算法之前对数据进行清洗, 使数据变得更加易用, 并且往往需去除数据中的噪声, 使其他机器学习的任务更加有效。PCA 能够识别数据中的主要特

征,它巧妙地利用协方差矩阵来计算样本集在不同方向上的分散程度,将方差最大的方向作为样本集的主方向。其主要处理过程是:首先利用样本矩阵计算得到协方差矩阵,然后计算协方差矩阵的特征值和特征向量,保留最大的前 $l$ 个特征值所对应的特征向量,组成正交基 $U$ ,最后将原始样本数据转换(投影)到新的基得到降维后的样本近似表示。

PCA降维虽然效率较高,但是也有其不足之处,例如,只适合数值型数据、可能会损失有用信息、容易受到噪声影响、未考虑样本固有的结构信息等。随着模式识别领域研究的深入,人们发现直接对图像的数值矩阵进行降维处理会比将其拉伸为向量再进行处理操作有更好的性能,在PCA的基础上发明了2DPCA、2DLDA、(2D)<sup>2</sup>PCA、Robust PCA等更有优势的算法。

#### 参考文献:

- [1] 何进荣. 几何观点下的线性降维算法[D]. 武汉: 武汉大学, 2014.  
HE Jinrong. Geometric Perspective on Linear Dimensionality Reduction Algorithms[D]. Wuhan: Wuhan University, 2014.
- [2] 李国杰. 大数据研究的科学价值[J]. 中国计算机学会通讯, 2012, 8(9): 8-15.  
LI Guojie. The Scientific Value of Big Data's Research[J]. Communications of The CCF, 2012, 8(9): 8-15.
- [3] BELLMAN R. Adaptive Control Process: A Guide Tour[M]. New Jersey: Princeton University Press, 1961: 203-218.
- [4] FISHER R A. The Use of Multiple Measurements in Taxonomic Problems[J]. Annals of Human Genetics, 1936, 7(2): 179-188.
- [5] HE X F, NIYOGI P. Locality Preserving Projections [C]// Advances in Neural Information Processing Systems 16 (NIPS 2003). Vancouver: NIPS, 2003: 153-160.
- [6] BELKIN M, NIYOGI P. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation[J]. Neural Computation, 2003, 15(6): 1373-1396.
- [7] TENENBAUM J B, SILVA V D, LANGFORD J C. A Global Geometric Framework for Nonlinear Dimensionality Reduction[J]. Science, 2000, 290: 2319-2323.
- [8] ROWEIS S T, SAUL L K. Nonlinear Dimensionality Reduction by Locally Linear Embedding[J]. Science, 2000, 290: 2323-2326.
- [9] PEARSON K. On Lines and Planes of Closest Fit to Systems of Point in Space[J]. Philosophical Magazine, 1901, 2(6): 559-572.
- [10] HOTELLING H. Analysis of a Complex of Statistical Variables into Principal Components[J]. Journal of Educational Psychology, 1933, 24: 417-441.
- [11] 王成. 基于谱正则化的线性降维方法研究[D]. 南京: 南京理工大学, 2012.  
WANG Cheng. The Research of Linear Dimensionality Reduction Base on Eigenspectrum Regularization[D]. NanJing: NanJing University of Science & Technology, 2012.
- [12] 张学工. 模式识别[M]. 3版. 北京: 清华大学出版社, 2010: 162-172.  
ZHANG Xuegong. Pattern Recognition[M]. 3rd ed. Beijing: Tsinghua University Press, 2010: 162-172.
- [13] 杉山将. 图解机器学习[M]. 许永伟, 译. 北京: 人民邮电出版社, 2015: 143-148.  
SUGIYAMA Masashi. Machine Learning Through Graphic[M]. XU Yongwei, Translated. Beijing: Posts & Telecom Press: 2015: 143-148.
- [14] HARRINGTON P. 机器学习实战[M]. 李锐, 译. 北京: 人民邮电出版社, 2013: 242-251.  
HARRINGTON P. Machine Learning in Action[M]. LI Rui, Translated. Beijing: Posts & Telecom Press: 2013: 242-251.
- [15] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 225-241.  
ZHOU Zhihua. Machine Learning[M]. Beijing: Tsinghua University Press, 2016: 225-241.
- [16] 薛毅, 陈立萍. 统计建模与R软件[M]. 北京: 清华大学出版社, 2006: 499-510.  
XUE Yi, CHEN Liping. Statistical Modeling and R Software[M]. Beijing: Tsinghua University Press, 2006: 499-510.
- [17] MAATEN L V D. Matlab Toolbox for Dimensionality Reduction[EB/OL]. [2016-05-04]. <http://lvdmaaten.github.io/drttoolbox/>.

(责任编辑: 申剑)