

PCA 联合子空间理论的规范化与扩展

徐 斌 马 尽 文

(北京大学数学科学学院信息科学系和数学及其应用教育部重点实验室, 北京 100871)

摘 要: 对于高维数据的分类, 主成分分析(PCA)联合子空间可为每类数据建立更为细致的概率模型, 从而可有效地提高贝叶斯分类的准确性。本文首先对 PCA 联合子空间理论进行了规范化, 提出了两个基本假设, 并从理论上证明了残差子空间参数“代表特征根”的启发式取值正是其极大似然估计。本文进一步对样本残差的概率模型进行了扩展, 提出了扩展型逐类联合子空间算法。最后, 本文通过在真实数据上实验结果证明了扩展型逐类联合子空间算法的优越性。

关键词: 主成分分析(PCA); 贝叶斯分类; 联合子空间

中图分类号: TP181 **文献标识码:** A **文章编号:** 1003-0530(2013)12-1638-06

Theoretical Normalization and Generalization of PCA Joint Subspace Model

XU Bin MA Jin-wen

(Department of Information Science, School of Mathematical Sciences and LMAM,
Peking University, Beijing, 100871)

Abstract: For the classification of high-dimensional data, PCA joint subspace model can accurately describe the probability distribution of the sample data of each class and thus improve the classification accuracy of the corresponding Bayesian classifier. In this paper, we firstly make certain theoretical normalization of the PCA joint subspace. Particularly, its two basic assumptions are proposed. Moreover, it is proved that the used heuristic setting of the parameter referred to as “representative eigenvalue” for the residual subspace is just its maximum likelihood estimate. We further generalize the expression of the probability distribution of the residual subspace and establish the generalized class-wise joint subspace algorithm for Bayesian classification. Finally, the experimental results on several real-world datasets demonstrate the superiority of the generalized class-wise joint subspace algorithm.

Key words: principle component analysis (PCA); Bayesian classification; joint subspace

1 引言

对于高维数据的分类, 我们通常先采用主成分分析(PCA)方法^[1]将数据降到低维空间, 然后对每类数据通过一定的学习算法建立概率模型, 再获得数据属于各类的条件概率密度, 最后根据贝叶斯决策进行分类。实际上, PCA 是将特征(即数据)空间分解为互相正交的主成分子空间和残差子空间。我们一般仅使用样本在主成分子空间上的投影数

据, 而将样本在残差子空间的投影数据看作随机噪声并直接舍弃掉。然而, 大量事实证明样本在残差子空间的投影数据也会包含着一定的类别信息, 甚至可能出现一些极端的情况: 不同类的数据的区别主要表现在残差, 而非主成分子上。

基于上述不足, Moghaddam 和 Pentland 于 1997 年提出一种联合子空间模型^[2]。它实际上联合了数据在主成分子空间上的概率密度和在残差子空间上的概率密度, 将其相乘作为样本的近似概率密

度。该模型使用高斯混合模型^[3]来逼近模式类在主成分空间上的投影数据的概率分布,并用球形高斯密度来逼近模式类在残差子空间上的投影数据分布。因此,该模型有效地利用了数据残差的类别信息来提高分类的精度。Bernard 于 1998 年进一步提出了主成分高斯混合模型(M-PCG)^[4]。该模型直接使用高斯混合模型来逼近模式类数据的概率分布,且对每个高斯分量进行子空间分解,并以广义 Hebb 算法(GHA)等来学习参数,最后将主成分与残差子空间联合得到类的概率密度函数。随后,Liu 等^[5]在此基础上简化了 M-PCG 的参数学习方式,并提出一种共享子空间混合模型,即对于一个类的各分量寻找最能反映公共特性的子空间,并且可使得子空间参数与高斯混合模型参数同时在 EM 算法中迭代更新^[6]。这种联合子空间的方法也可用于防止 EM 算法在迭代过程中协方差矩阵出现奇异现象,并保持高斯分量的主分量不变,而基于贝叶斯正则化的 EM 算法^[7]则需要对协方差矩阵的所有特征根都增加一个正则化因子。

本文在上述研究的基础上,尝试将联合子空间理论进行规范化并加以扩展。首先,我们提出了联合子空间方法成立所暗示的两个假设,并进行了分析。然后,我们在两个基本假设下求出了残差子空间参数“代表特征根”的极大似然估计,从理论上证明了实际中所采用的启发式公式的合理性。对于样本残差,本文尝试了比球形高斯更细致的概率模型,提出了扩展型逐类联合子空间算法。最后,我们通过在真实数据上的实验结果证明了扩展型逐类联合子空间算法的优越性。

2 PCA 与联合子空间理论

2.1 PCA 的计算步骤

PCA 是一种非监督的特征提取方法。它的作用是使数据能由少量的重要特征来表示,尽量保持其有效信息,并消除掉噪声的干扰。其计算步骤如下:

给定一组 d 维样本 $\mathbf{X} = \{x_i\}_{i=1}^N$, 其均值为 $\mu = \frac{1}{N} \sum_{i=1}^N x_i$, 且其协方差矩阵为 $\Sigma = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T$ 。令 Σ 的特征值按降序排列,即为: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$, 所对应的特征向量构成一个 $d \times d$ 矩阵: $\Phi = [q_1, q_2, \dots, q_d]$ 。

按照特征值的累计贡献的实际需求 α 来选取

变换后的维数 $m: \sum_{j=1}^m \lambda_j / \sum_{j=1}^d \lambda_j \geq \alpha$, 例如取 $\alpha = 90\%$ 。

对所有样本中心化,即令 $\hat{x} = x - \mu$, PCA 将样本所在的特征空间分解成为两个正交的子空间:

主成分空间 F : 由 q_1, \dots, q_m 生成;

残差子空间 \bar{F} : 由 q_{m+1}, \dots, q_d 生成。

将 \hat{x} 在 F 上投影得 $y = [q_1, \dots, q_m]^T \hat{x}$ 。

这样便将样本数据 x 通过 PCA 转化为低维数据 y , 我们则可以通过 y 来分析 x 的特性, 因此 PCA 简化了数据的表示和结构。

2.2 高斯混合模型

当随机变量 Z 的概率分布是未知的时候, 我们总可以采用一个高斯混合模型(GMM)来逼近它, 即认为 Z 的概率密度具有如下形式:

$$P(z | \Theta) = \sum_{j=1}^k \alpha_j q(z | \mu_j, \Sigma_j) \quad (1)$$

其中 $q(z | \mu_j, \Sigma_j)$ 是第 j 个高斯(或正态)分量(即概率密度函数), μ_j, Σ_j 分别为该分量的均值和协方差矩阵, 具有下述表达式:

$$q(z | \mu_j, \Sigma_j) = \frac{1}{(2\pi)^{d/2} |\Sigma_j|^{1/2}} e^{-(1/2)(z - \mu_j)^T \Sigma_j^{-1} (z - \mu_j)} \quad (2)$$

另外, k 是分量个数, α_j 是第 j 个分量的混合比例系数, 满足 $\sum_{j=1}^k \alpha_j = 1$ 。

实际上, 高斯混合模型是一种很强的概率模型, 能够对一般数据的概率分布进行很好的逼近, 因此被广泛地应用到数据分析和信息处理中。当分量个数 k 已知时, 我们可采用 EM 算法^[8]来估计参数 Θ 。实际上, 这是一种基于最大似然估计的迭代算法。但当 k 是未知的时候, 我们则需要对 k 做出选择, 这就是所谓的模型选择问题。针对模型选择, Ma 等^{[9]-[12]}已经建立了一类很有效的自动模型选择算法。它们不仅能够根据样本数据合理地选择出高斯分量个数, 也同时得到了其他参数的有效估计。

2.3 贝叶斯决策

在实际分类问题中, 我们可对样本 X 经过 PCA 所得到的 Y 的每类数据建立一个 GMM, 即得到类的条件概率密度 $P(y | \omega_j)$ ($j = 1, 2, \dots, M$, M 为类别数, ω_j 代表第 j 类)。设第 ω_j 类的先验概率是 $P(\omega_j)$, 则根据贝叶斯公式:

$$P(\omega_i | y) = \frac{P(y | \omega_i)P(\omega_i)}{\sum_{j=1}^k P(y | \omega_j)P(\omega_j)} \quad (3)$$

可得到 y 属于第 ω_j 类的后验概率。贝叶斯决策的规则为:如果 $P(\omega_j | y) = \max_i P(\omega_i | y)$, 则 $y \in \omega_j$ 。

对于均衡分类情况,即各类的先验概率 $P(\omega_j)$ 都相等,这时贝叶斯决策的规则就转化为:如果 $P(y | \omega_j) = \max_i P(y | \omega_i)$, 则 $y \in \omega_j$ 。

当我们将 y 完成了贝叶斯分类后,则可将所对应的 x 归入相应的类中,完成输入数据的分类。

2.4 联合子空间理论

上述基于 PCA 的贝叶斯分类算法仅仅使用了样本在主成分空间的投影,而联合子空间的思想则是综合地考虑样本在主成分和残差子空间上的投影的概率分布,更好地逼近原数据的概率分布,即令

$$P(x | \Theta) \approx P_F(x | \Theta) \cdot P_{\bar{F}}(x | \Theta) \quad (4)$$

首先,我们以高斯分布为例来说明这种联合表达式的合理性。若 X 服从如下高斯分布:

$$P(x | \theta) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)} \quad (5)$$

对 X 进行 PCA, 设 Σ 的特征值按降序排列为: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$, 对应的标准正交的特征向量构成矩阵: $\Phi = [q_1, q_2, \dots, q_d]$, 由 $\lambda_1, \lambda_2, \dots, \lambda_d$ 构成的对角矩阵为 Λ , 则有:

$$\Sigma = \Phi \Lambda \Phi^T \quad (6)$$

令 $y = \Phi^T(x - \mu)$, 选取 m 维主成分, 则 x 在主成分空间 F 上的投影为 (y_1, \dots, y_m) , 在残差子空间 \bar{F} 的投影为 (y_{m+1}, \dots, y_d) , 定义残差平方和为:

$$\varepsilon^2(x) = \sum_{i=m+1}^d y_i^2 = \|\hat{x}\|^2 - \sum_{i=1}^m y_i^2 \quad (7)$$

令 $d(x)$ 为 x 的马氏距离, 即

$d(x) = (x - \mu)^T \Sigma^{-1} (x - \mu)$, 并利用 $\Sigma = \Phi \Lambda \Phi^T$, $y = \Phi^T(x - \mu)$, $\Phi^T = \Phi^{-1}$, 可得:

$$\begin{aligned} d(x) &= y^T \Lambda y = \sum_{i=1}^d \frac{y_i^2}{\lambda_i} \\ &= \sum_{i=1}^m \frac{y_i^2}{\lambda_i} + \sum_{i=m+1}^d \frac{y_i^2}{\lambda_i} \end{aligned} \quad (8)$$

统一用 ρ 代替较小的 $(d-m)$ 个特征根以逼近 $d(x)$, 则有:

$$\hat{d}(x) = \sum_{i=1}^m \frac{y_i^2}{\lambda_i} + \frac{1}{\rho} \sum_{i=m+1}^d y_i^2 = \sum_{i=1}^m \frac{y_i^2}{\lambda_i} + \frac{\varepsilon^2(x)}{\rho} \quad (9)$$

该 ρ 被称为残差子空间的代表特征根, Moghaddam 和 Pentland 使用了 ρ 的如下启发公式^[2]:

$$\rho^* = \frac{1}{d - m} \sum_{i=m+1}^d \lambda_i \quad (10)$$

现将 $\hat{d}(x)$ 代入 $P(x | \theta)$, 并用 ρ 代替最后 $(d-m)$ 个比较小的特征根, 我们则得到近似概率密度 $\hat{P}(x | \theta)$:

$$\begin{aligned} \hat{P}(x | \theta) &= \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\sum_{i=1}^m \frac{y_i^2}{\lambda_i} + \frac{\varepsilon^2(x)}{\rho})} \\ &\approx \frac{1}{(2\pi)^{m/2} \prod_{i=1}^m \lambda_i^{1/2}} e^{-\frac{1}{2}(\sum_{i=1}^m \frac{y_i^2}{\lambda_i})} \cdot \frac{1}{(2\pi\rho)^{(d-m)/2}} e^{-\frac{\varepsilon^2(x)}{2\rho}} \\ &= P_F(x | \theta) P_{\bar{F}}(x | \theta) \end{aligned} \quad (11)$$

从(11)式可以看出, 残差的分布是被一个球形的高斯函数 $P_{\bar{F}}(x | \theta)$ 所近似, 其作用可能很小, 但却起到了一定的“调制”作用, 反映了该类数据的一些补充信息。

对于服从一般概率分布的样本 X , 我们总可以用高斯混合模型的密度函数来逼近 $P_F(x | \Theta)$, 同时依然可用球形高斯函数或分布来近似 $P_{\bar{F}}(x | \Theta)$, 将二者相乘得到 X 的近似概率密度函数:

$$\hat{P}(x | \Theta) = \left(\sum_{j=1}^m \alpha_j q(x | \mu_j, \Sigma_j) \right) \frac{1}{(2\pi\rho)^{(d-m)/2}} \frac{\varepsilon^2(x)}{2\rho} \quad (12)$$

不同于传统的 PCA 贝叶斯分类方法直接对全体样本进行 PCA 降维, 基于联合子空间的 PCA 贝叶斯分类算法则需要计算出每类的近似概率密度。因此, 我们则需要逐类进行 PCA 和子空间分解, 当然 GMM 的参数学习和贝叶斯决策规则是不变的。

3 联合子空间理论的规范化与扩展

3.1 两个假设

从联合子空间的引入可以看出, (12)式的成立是需要一定的假设条件:

1) 假设随机向量 X (样本) 在主成分空间 F 上的投影 X_F 和在残差子空间 \bar{F} 上的投影 $X_{\bar{F}}$ 是独立的。由于 X_F 和 $X_{\bar{F}}$ 都是随机向量, 若令 $\hat{P}(x | \Theta) = P_F(x | \Theta) P_{\bar{F}}(x | \Theta)$, 当且仅当 X_F 和 $X_{\bar{F}}$ 是独立的, 这时才可用 $\hat{P}(x | \Theta)$ 来逼近 $P(x | \Theta)$ 。

从子空间的分解过程可知, F 和 \bar{F} 是正交的, 因此对于 X_F 的任意分量 x_{Fi} 和 $X_{\bar{F}}$ 的任意分量 x_{Fj} , 必然

有 $\text{cov}(x_{F_i}, x_{F_j}) = 0$, 即它们是统计不相关的。这是 X_F 和 $X_{\bar{F}}$ 独立的必要条件, 但这显然还不够充分。

2) 假设 X_F 服从球形高斯分布, 即各维方差相等的多元高斯分布。首先, 若 X_F 服从高斯分布, 则有

$$P_{\bar{F}}(x | \Theta) = \frac{1}{(2\pi)^{(d-m)/2} \prod_{i=m+1}^d \lambda_i^{1/2}} e^{-\frac{1}{2}(\sum_{i=m+1}^d \frac{y_i^2}{\lambda_i})} \quad (13)$$

从子空间的分解过程可知, X_F 以 0 为期望, 且方差都比较小, 故假设 $X_{\bar{F}}$ 服从单模态、对称的高斯分布可以得到对 $P_{\bar{F}}(x | \Theta)$ 的有效近似。

在上述基础上可做进一步简化。为了避免一一计算 $y_i (i=m+1, \dots, d)$, 假设 $\lambda_{m+1} = \dots = \lambda_d = \rho$ 后, 则有:

$$P_{\bar{F}}(x | \Theta) = \frac{1}{(2\pi\rho)^{(d-m)/2}} e^{-\frac{\varepsilon^2(x)}{2\rho}} \quad (14)$$

其中 $\varepsilon^2(x) = \|\hat{x}\|^2 - \sum_{i=1}^m y_i^2$ 。很显然, 这样使得贝叶斯分类的计算量大大减少了。

若上述两个假设成立, 并用高斯混合模型的密度函数来近似 $P_{\bar{F}}(x | \Theta)$, 就合理地得到(12)式, 完成了数据概率分布的 PCA 分解和联合。

3.2 代表特征根的极大似然估计

在(10)式中, 采用残差子空间各维特征根的均值作为代表特征根是一种启发式的想法。实际上, 我们下面可以证明这一公式是合理的, 因为在上述两个假设下, 它正是代表特征根的极大似然估计。

对于样本点 $x_i (i=1, 2, \dots, N)$, 其在残差子空间 \bar{F} 上的概率密度为(14)式, 则由这组样本点构成的对数似然函数为:

$$l(\rho, \mathbf{X}) = -\frac{\sum_{i=1}^N \varepsilon^2(x_i)}{2\rho} - \frac{(d-m)N}{2} \log \rho \quad (15)$$

其中省略了常数项。令

$$\frac{\partial l}{\partial \rho} = \frac{\sum_{i=1}^N \varepsilon^2(x_i)}{2\rho^2} - \frac{(d-m)N}{2\rho} = 0 \quad (16)$$

由此给出 ρ 的极大似然估计为

$$\hat{\rho} = \frac{\sum_{i=1}^N \varepsilon^2(x_i)}{(d-m)N} \quad (17)$$

将(7)式代入得

$$\hat{\rho} = \frac{\sum_{i=1}^N \sum_{j=m+1}^d y_{ij}^2}{(d-m)N} \quad (18)$$

y_{ij} 表示第 i 个样本 x_i 在第 j 个特征向量方向上的投影。又由于

$$\begin{aligned} \lambda_j &= \text{eig}_j(\text{cov}(X)) = \text{eig}_j\left(\frac{1}{N} \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T\right) \\ &= \text{eig}_j(\text{cov}(Y)) = \frac{1}{N} \sum_{i=1}^N y_{ij}^2 \end{aligned} \quad (19)$$

其中用到了 $Y = \Phi^T(X - \mu)$, $\Phi\Phi^T = I$ 。

综合(18)、(19)式, 我们则得到:

$$\hat{\rho} = \frac{\sum_{j=m+1}^d \lambda_j}{d-m} \quad (20)$$

因此, 我们便从理论上证明了(10)式中的 ρ^* 正是 ρ 的最大似然估计 $\hat{\rho}$, 即代表特征根的启发公式为其最大似然估计。

3.3 残差的扩展概率模型

对于残差子空间上的投影数据的概率分布, 球形高斯分布是最简单的假设, 具有各向同性、并沿径向指数衰减的特性。为了使其适应更复杂的情况, 我们可以尝试可塑性更强、结构稍复杂一些的概率模型。实际上, 我们可先进行下列概率分解:

$$P_{\bar{F}}(x) = P_{\bar{F}}(x, \varepsilon^2(x)) = P_{\bar{F}}(\varepsilon^2(x)) P_{\bar{F}}(x | \varepsilon^2(x)) \quad (21)$$

$$\text{当 } P_{\bar{F}}(x | \Theta) = \frac{1}{(2\pi\rho)^{(d-m)/2}} e^{-\frac{\varepsilon^2(x)}{2\rho}}, X_F = (y_{m+1}, \dots,$$

$y_d), y_i \sim N(0, \rho)$, 而 $\varepsilon^2(x) = \sum_{i=m+1}^d y_i^2$, 则自然有 $\varepsilon^2(x)$

$\sim \text{Gamma}(\frac{d-m}{2}, 2\rho)$, 即

$$P_{\bar{F}}(\varepsilon^2(x)) = \frac{\varepsilon^2(x)^{\frac{(d-m)-1}{2}} e^{-\frac{\varepsilon^2(x)}{2\rho}}}{\Gamma(\frac{d-m}{2})(2\rho)^{(d-m)/2}} \quad (22)$$

因此得到:

$$P_{\bar{F}}(x | \varepsilon^2(x)) = \frac{P_{\bar{F}}(x)}{P_{\bar{F}}(\varepsilon^2(x))} = \frac{2[\varepsilon^2(x)]^{1/2}}{S([\varepsilon^2(x)]^{1/2} | d-m)} \quad (23)$$

其中 $S(r | d-m)$ 是半径为 r 、 $(d-m)$ 维的球的表面积。

我们尝试保留 X_F 各向同性, 放宽沿径向指数衰减的假设, 采用一般的伽马分布来逼近 $P_{\bar{F}}(\varepsilon^2(x))$ 。

若保持(23)式不变, 采用 $\text{Gamma}(\alpha, \beta)$ 来逼近

$\varepsilon^2(x)$ 的概率分布,并对参数 α 和 β 进行下述矩估计:

$$\alpha = \frac{E(\varepsilon^2(x))^2}{\text{Var}(\varepsilon^2(x))} \quad (24)$$

$$\beta = \frac{E(\varepsilon^2(x))}{\text{Var}(\varepsilon^2(x))} \quad (25)$$

且得到 $P_F(\varepsilon^2(x))$ 的表达式:

$$P_F(\varepsilon^2(x)) = \frac{\varepsilon^2(x) e^{-\frac{\varepsilon^2(x)}{\beta}}}{\Gamma(\alpha) \beta^\alpha} \quad (26)$$

我们将(26)、(23)代入(21),则得:

$$P_F(x) = \frac{\Gamma\left(\frac{d-m}{2}\right) [\varepsilon^2(x)]^{(\alpha-\frac{d-m}{2})}}{\Gamma(\alpha) \pi^{(d-m)/2} \beta^\alpha} e^{-\frac{\varepsilon^2(x)}{\rho}} \quad (27)$$

当 $\alpha = \frac{d-m}{2}, \beta = 2\rho$ 时,(27)与(14)相同,这说明

后者是前者的扩展。为了便于描述,我们称使用(27)来代替(14)的相应算法为扩展型逐类联合子空间算法。由于(27)中的 α, β 的估计来自实际样本,更能真实地反映残差的概率分布,从而可提高对样本实际概率分布逼近的准确性。

4 实验结果和比较

为了评价PCA联合子空间算法在真实数据上的效果,我们特意在9个UCI数据集^[13]进行了实验。这些数据的维数从4到240不等。一些数据(optdigits、letter、satimage、pendigits)已经提供了标准的训练和测试集。而对于其它数据,我们使用全部数据参加训练和测试。数据的详细描述见表1,其中d表示数据的维数,class表示类数,train和test分别是训练和测试样本的个数。

为了叙述简便,我们将所使用的算法简称为:

- M0: PCA贝叶斯分类算法
- M1: 逐类联合子空间算法
- M2: 扩展型逐类联合子空间算法

这三种算法都有两个主要参数:特征根累计贡献率和GMM的分量个数,分别用 α 和 m 表示。对每个数据,利用EM算法进行高斯混合模型的参数估计,将每个算法运行50次,记录正确率的均值和标准差。根据这些数据的特性,这里没有采用复杂的模型选择算法,而是在保证较好分类结果的前提下,尽量选择较小的 α 和 m 以降低计算量并提高泛化能力。参数选取和实验结果已经在表2中列出,并且将联合子空间算法比PCA贝叶斯分类算法正

确率低的情况用黑色底纹标记。

表1 UCI数据集的特性描述

Tab. 1 The characteristic description of UCI datasets for the experiments

Dataset	d	class	train	test
Iris	4	3	150	150
Wine	13	3	178	178
optdigits	64	10	3823	1797
segment	19	10	2310	2310
mfeat-kar	64	10	2000	2000
mfeat-pix	240	10	2000	2000
letter	16	26	16000	4000
satimage	36	7	4435	2000
pendigits	16	10	7494	3498

表2 三种算法在UCI数据集上的分类正确率

Tab. 2 The classification accuracy rates of the three algorithms on the UCI datasets

Dataset	α	m	M0	M1	M2
iris	0.95	1	97.33% ± 0.00%	98% ± 0.00%	98% ± 0.00%
wine	0.60	1	97.75% ± 0.00%	99.44% ± 0.00%	98.88% ± 0.00%
optdigits	0.60	5	93.47% ± 0.36%	96.08% ± 0.06%	94.53% ± 0.12%
segment	0.80	5	93.85% ± 0.92%	87.34% ± 0.60%	87.85% ± 0.34%
mfeat-kar	0.50	2	97.16% ± 0.13%	98.18% ± 0.04%	98.23% ± 0.05%
mfeat-pix	0.50	5	96.53% ± 0.23%	98.56% ± 0.04%	98.65% ± 0.05%
letter	0.95	8	95.51% ± 0.28%	94.39% ± 0.19%	94.68% ± 0.21%
satimage	0.80	8	82.37% ± 0.17%	84.84% ± 0.16%	83.543% ± 0.29%
pendigits	0.80	5	93.88% ± 0.25%	94.15% ± 0.20%	95.10% ± 0.20%

表 2 给出了三种算法在 9 个 UCI 数据集上的分类正确率和标准差。从表 2 中可以看出:

- 1) 联合子空间算法的正确率几乎都高于 PCA 贝叶斯分类算法,并且更为稳定。但在 letter 数据集上,联合子空间算法稍弱于 PCA 贝叶斯分类算法,且在 segment 数据集上,联合子空间方法较明显地弱于 PCA 贝叶斯分类算法。这可能与这些数据的特殊结构有关。
- 2) 对这 9 个数据,扩展型逐类联合子空间算法有 5 个分类效果好于原算法,1 个不变,3 个变差,而分类结果的稳定性基本相同。故总体上,扩展型算法的分类效果要优于原算法。这说明对残差子空间使用可塑性更大的概率模型对于分类效果的提高确实具有一定的促进作用。

5 结论

本文首先对联合子空间理论进行规范化,给出了其成立的两个基本假设,并在极大似然估计的框架下证明了“代表特征根”的启发表达式的合理性。本文进一步将联合子空间模型进行了扩展,提出了扩展型逐类联合子空间算法,即对残差子空间使用比球形高斯分布更具可塑性的概率模型。基于 9 组真实数据的实验结果表明,PCA 联合子空间算法能够取得了比传统 PCA 贝叶斯分类算法更好的效果,并且本文所提出的扩展型逐类联合子空间算法也在一定程度上优于原算法。

参考文献

- [1] I Jolliffe. Principal Component Analysis[M]. New York: Springer-Verlag, 2002.
- [2] B Moghaddam, A Pentland. Probabilistic visual learning for object representation[J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 1997, 19(7): 696-710.
- [3] G J McLachlan, D Peel. Finite Mixture Models[M]. New York: John Wiley & Sons, 2000.
- [4] L Bernard. Mixtures of principal components Gaussians for density estimation in high dimension data spaces[J]. Lecture Notes in Computer Science, 1998, 1451: 952-959.
- [5] X H Liu, C L Liu, X W Hou. A pooled subspace mixture density model for pattern classification in high-dimensional spaces[C] // Proc. IJCNN 2008, Hong Kong: IEEE Press, 2008. 2467-2472.

- [6] X H Liu, C L Liu. Discriminative training of subspace Gaussian mixture model for pattern classification[J]. Lecture Notes in Computer Science, 2010, 6215: 213-221.
- [7] D Ormoneit, V Tresp. Improved Gaussian mixture density estimates using Bayesian penalty terms and network averaging[C] // Advances in Neural Information Processing Systems 8, Cambridge, MA: MIT Press, 1996. 542-548.
- [8] R A Render, H F Walker. Mixture densities, maximum likelihood and the EM algorithm[J]. SLAM Review, 1984, 26(2): 195-239.
- [9] J Ma, T Wang, L Xu. A gradient BYY harmony learning rule on Gaussian mixture with automated model selection[J]. Neurocomputing, 2004, 56: 481-487.
- [10] J Ma, L Wang. BYY harmony learning on finite mixture: adaptive gradient implementation and a floating RPCL mechanism[J]. Neural Processing Letters, 2006, 24(1): 19-40.
- [11] J Ma, J Liu. The BYY annealing learning algorithm for Gaussian mixture with automated model selection[J]. Pattern Recognition, 2007, 40: 2029-2037.
- [12] J Ma, X He. A fast fixed-point BYY harmony learning algorithm on Gaussian mixture with automated model selection[J]. Pattern Recognition Letters, 2008, 29(6): 701-711.
- [13] C Blake, E Keogh, C J Merz, UCI Repository of Machine Learning Databases [EB/OL]. <http://www.ics.uci.edu/~mllearn/MLRepository.html>, Department of Information and Computer Science, University of California, Irvine, 1998.

作者简介



徐 斌 男,1985 年生,湖北人。2012 年毕业于北京大学数学科学学院信息科学系,获得理学硕士学位。目前担任解放军 61915 部队助理工程师,主要研究方向为模式识别和信号处理。
E-mail: xu_bin_1985@126.com



马 尽 文 男,1962 年生,陕西人。1992 年毕业于南开大学数学系,获理学博士学位。现为北京大学数学科学学院信息科学系主任、教授、博士生导师,中国电子学会信号处理分会委员,中国工业与应用数学学会理事,主要从事智能信息处理、神经计算、模式识别、生物信息学等方面的研究。

E-mail: jwma@math.pku.edu.cn