

一种新的局部判别投影方法

谢 钧¹⁾ 刘 剑²⁾

¹⁾(解放军理工大学指挥自动化学院 南京 210007)

²⁾(解放军理工大学通信工程学院 南京 210007)

摘 要 为解决多模数据的分类问题,局部化思想被引入到判别分析中,称为局部判别分析.该文以人工数据为例深入分析了近年来提出的较为成功的两种局部线性判别分析方法:LFDA(Local Fisher Discriminant Analysis)和MFA(Marginal Fisher Analysis)的不足.为克服这两种方法中没有充分考虑异类样本近邻关系的缺点,文中提出了一种新的局部判别投影方法.该方法采用与LFDA和MFA不同的局部化方法,其基本思想是寻找投影方向使同类近邻样本在投影后尽量紧凑,而异类近邻样本在投影后尽量分开.针对该思想,文中提出了两种优化目标(一种用样本间距离平方和来表示,另一种用样本类内与类间散度来表示)并做了分析和比较.实验结果表明,该文方法有效地克服了LFDA和MFA存在的固有问题,在人工数据集、UCI、USPS手写数字标准数据集和IDA标准数据集上均取得较好效果.

关键词 分类;降维;线性判别分析;局部保持;多模

中图法分类号 TP391 DOI号: 10.3724/SP.J.1016.2011.02243

A New Local Discriminant Projection Method

XIE Jun¹⁾ LIU Jian²⁾

¹⁾(Institute of Command Automation, PLA University of Science and Technology, Nanjing 210007)

²⁾(Institute of Communication Engineering, PLA University of Science and Technology, Nanjing 210007)

Abstract To solve the problem of multimodal data classification, the idea of localization is introduced into discriminant analysis, known as local discriminant analysis. In this paper, we first illustrated, by some synthetic data as examples, the drawbacks of LFDA and MFA, two recently proposed and successfully used local linear discriminant analysis methods. We then proposed a new local discriminant projection method to overcome the drawback of LFDA and MFA that neighbor relationships between samples of different classes are not fully taken into consideration. The underlying idea of the new method, different from LFDA and MFA, is that the desired projection should make neighbors of the same class close and neighbors of different classes apart. Based on this idea, we proposed two optimal functions, one is represented by the sum of squared distance between samples, the other is represented by within class scatter and between class scatter. Analyses and comparisons on the two optimal functions are also included in this paper. Experiment results show that the new method overcomes the shortcomings of LFDA and MFA, and achieves good performance on the synthetic data, USPS, UCI standard handwriting digital data sets and IDA standard data sets.

Keywords classification; dimensionality reduction; linear discriminant analysis; locality preserving; multimodal

1 引言

降维是通过某种变换把高维空间中的数据变换到低维空间,在降低数据维数的同时尽可能地保留原始数据的本质信息.降维技术是高维数据压缩、可视化和分类的重要预处理手段,在计算机视觉、机器学习和模式识别等领域应用广泛.

降维方法可分为线性降维(采用线性变换)和非线性降维(采用非线性变换).以 ISOMap^[1]、Laplacian 特征映射^[2]、LLE(Locally Linear Embedding)^[3] 等为代表的非线性降维方法虽然在处理特定的人工数据时表现出很好的性能,但面对实际数据时却并没有明显优势^[4].此外,一些基于流形学习的非线性降维方法通常不能得到一个明确的变换,因而无法处理新出现的样本,使得此类方法在分类中的应用受到限制.在分类中更常用的是各种线性降维方法,包括传统的 PCA(Principle Component Analysis)^[5]、LDA(Linear Discriminant Analysis)^[5] 及近年来提出的 LPP(Local Preserving Projections)^[6]、NPE(Neighborhood Preserving Embedding)^[7]、LEA(Locally Linear Embedding)^[8]、ISO 投影^[9]、SPP(Sparsity Preserving Projections)^[10] 等等.此类方法通过训练得到的投影矩阵可直接将未见样本投影到低维空间.

降维又可分为无监督降维和有监督降维,前者关注如何更好地表示高维数据,通常不借助样本的类别信息;而后者是为了有效地实现分类,将类别信息用于降维过程中.两种降维也分别称作表示性降维和判别性降维.在线性判别降维方法中 LDA 较为经典,它以最小化类内散度和最大化类间散度为优化目标. LDA 在模式识别领域得到广泛应用,但对于多模数据(同一类样本包含几个不同分布的簇)的分类却并不理想^[11].要解决多模数据的分类问题,需要在降维过程中借助局部信息,尽可能地保留数据的流形特性.为此,一些研究聚焦在如何将局部化思想引入到判别性降维方法中,例如近年提出的 LFDA(Local Fisher Discriminant Analysis)^[11] 和 MFA(Marginal Fisher Analysis)^[12].

本文深入分析了 LFDA 和 MFA 的不足,并在此基础上提出了一种新的局部线性判别降维方法,采用与 LFDA 和 MFA 不同的局部化方法,有效地克服了两种方法存在的固有问题,在人工数据、USPS、UCI 手写数字标准数据集和 IDA 标准数据

集上的实验中均取得较好效果.

2 局部化线性判别降维方法

由 n 个样本组成的样本集记作 $\mathbf{X}=[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]_{d \times n}$, \mathbf{x}_i 是第 i 个样本,类标号为 c_i ,维数为 d .线性降维是按照某种优化准则求得一个 $d \times m$ ($m < d$) 的变换矩阵 \mathbf{W} ,将样本集投影到低维空间得到 $\mathbf{Y}=\mathbf{W}^T \mathbf{X}=[\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]_{m \times n}$,其中 \mathbf{y}_i 是 \mathbf{x}_i 在低维空间的投影,维数为 m .

2.1 LDA 和 LPP

经典线性判别降维方法 LDA 的基本思想是最小化类内散度和最大化类间散度,并以散布矩阵行列式作为样本散度的度量.其优化目标为

$$\min_{\mathbf{W} \in \mathbb{R}^{d \times m}} \frac{|\mathbf{W}^T \mathbf{S}^{(w)} \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}^{(b)} \mathbf{W}|}$$

$$\mathbf{S}^{(w)} = \sum_{l=1}^c \sum_{i: c_i=l} (\mathbf{x}_i - \boldsymbol{\mu}_l)(\mathbf{x}_i - \boldsymbol{\mu}_l)^T \quad (1)$$

$$\mathbf{S}^{(b)} = \sum_{l=1}^c n_l (\boldsymbol{\mu}_l - \boldsymbol{\mu})(\boldsymbol{\mu}_l - \boldsymbol{\mu})^T$$

其中 c 是样本类别总数, $\boldsymbol{\mu}$ 是所有样本的均值, $\boldsymbol{\mu}_l$ 是第 l 类样本的均值, $\mathbf{S}^{(w)}$ 称为总类内样本散布矩阵, $\mathbf{S}^{(b)}$ 称为总类间散布矩阵.

为便于与 LFDA 和 MFA 比较,把 LDA 的 $\mathbf{S}^{(w)}$ 和 $\mathbf{S}^{(b)}$ 写成下面的等价形式^[11]:

$$\mathbf{S}^{(w)} = \frac{1}{2} \sum_{i,j=1}^n p_{ij} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T = \mathbf{X}(\mathbf{D} - \mathbf{P})\mathbf{X}^T,$$

$$\mathbf{S}^{(b)} = \frac{1}{2} \sum_{i,j=1}^n \tilde{p}_{ij} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T = \mathbf{X}(\tilde{\mathbf{D}} - \tilde{\mathbf{P}})\mathbf{X}^T,$$

$$p_{ij} = \begin{cases} 1/n_l, & c_i = c_j = l \\ 0, & c_i \neq c_j \end{cases},$$

$$\tilde{p}_{ij} = \begin{cases} 1/n - 1/n_l, & c_i = c_j = l \\ 1/n, & c_i \neq c_j \end{cases} \quad (2)$$

其中 $\mathbf{P}=[p_{ij}]_{n \times n}$ 和 $\tilde{\mathbf{P}}=[\tilde{p}_{ij}]_{n \times n}$ 是权值矩阵, \mathbf{D} 和 $\tilde{\mathbf{D}}$ 是 $n \times n$ 的对角阵,对角元素 $d_i = \sum_j p_{ij}$, $\tilde{d}_i = \sum_j \tilde{p}_{ij}$,以下均采用这种记号.从式(2)不难看出,权值 p_{ij} 和 \tilde{p}_{ij} 中体现了判别信息,即权值与 \mathbf{x}_i 和 \mathbf{x}_j 是否同类有关,然而 p_{ij} 和 \tilde{p}_{ij} 中没有体现局部信息,即权值与 \mathbf{x}_i 和 \mathbf{x}_j 是否相邻无关.

LPP^[6] 是一种旨在保留局部信息的表示性降维方法,在降维过程中能较好地保持流形特性,是 Laplacian 特征映射^[2] 的线性近似. LPP 的基本思想是希望在高维空间中相距较近的样本点在投影后的

低维空间中仍比较近. 其优化目标是

$$\begin{aligned} \min_{\mathbf{W} \in \mathbb{R}^{d \times m}} \sum_{i,j=1}^n p_{ij} \|\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j\|^2 \\ \text{s. t. } \mathbf{W}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{W} = \mathbf{I}_m \quad (3) \\ p_{ij} = \begin{cases} A_{ij}, & i \in N_k(j) \\ 0, & \text{其它} \end{cases} \end{aligned}$$

其中 $N_k(j)$ 表示样本 \mathbf{x}_j 的 k 近邻集合, A_{ij} 是样本 \mathbf{x}_i 和 \mathbf{x}_j 之间的一种相似性度量, 如果 \mathbf{x}_i 和 \mathbf{x}_j 差距大则 A_{ij} 小. 例如 A_{ij} 可定义为 $A_{ij} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma_i \sigma_j}\right)$, 以 σ_i, σ_j 为参数. 可见 LPP 中的权值 p_{ij} 体现了局部信息, 与 \mathbf{x}_i 和 \mathbf{x}_j 是否相邻有关.

2.2 LFDA 和 MFA

LFDA 和 MFA 有效结合了 LDA 和 LPP 的思想, 在 LDA 权值的定义中融入局部信息, 在数据表示和分类中均取得较好效果. 两者可与 LDA 纳入到同一个优化框架中

$$\min_{\mathbf{W} \in \mathbb{R}^{d \times m}} \frac{|\mathbf{W}^T \mathbf{X} (\mathbf{D} - \mathbf{P}) \mathbf{X}^T \mathbf{W}|}{|\mathbf{W}^T \mathbf{X} (\tilde{\mathbf{D}} - \tilde{\mathbf{P}}) \mathbf{X}^T \mathbf{W}|} \quad (4)$$

在 LFDA 中权值定义为

$$\begin{aligned} p_{ij} = \begin{cases} A_{ij}/n_l, & c_i = c_j = l \\ 0, & c_i \neq c_j \end{cases}, \\ \tilde{p}_{ij} = \begin{cases} A_{ij}(1/n - 1/n_l), & c_i = c_j = l \\ 1/n, & c_i \neq c_j \end{cases} \quad (5) \end{aligned}$$

其中 A_{ij} 表示 \mathbf{x}_i 和 \mathbf{x}_j 之间的一种相似性度量, 通常在两个样本不是近邻时取 0.

在 MFA 中权值定义为

$$\begin{aligned} p_{ij} = \begin{cases} 1, & \mathbf{x}_i \in N_{k_1}^+(j) \text{ 或 } \mathbf{x}_j \in N_{k_1}^+(i) \\ 0, & \text{其它} \end{cases}, \\ \tilde{p}_{ij} = \begin{cases} 1 & (i, j) \in P_{k_2}(c_i) \text{ 或 } (i, j) \in P_{k_2}(c_j) \\ 0, & \text{其它} \end{cases} \quad (6) \end{aligned}$$

其中 $N_{k_1}^+(j)$ 是样本 \mathbf{x}_j 的同类 k_1 近邻的集合, $P_{k_2}(c)$ 是集合 $\{(i, j) | i \in \pi_c, j \notin \pi_c\}$ 中前 k_2 个最近的样本对, π_c 表示类别号为 c 的所有样本的标号集合.

2.3 LFDA 和 MFA 的不足

从式(5)和(6)不难看出, LFDA 和 MFA 都在同类样本间的权值定义中引入了局部信息, 同类近邻样本之间权值非 0, 同类非近邻样本之间权值为 0. 两种方法的主要区别在于对异类样本权值的选择上.

LFDA 不考虑异类样本之间的近邻关系, 只要两个样本异类, 两者之间的权值 \tilde{p}_{ij} 就是 $1/n$, 这会带来两个缺点: 一是对野值敏感, 例如存在一个野值 \mathbf{x}_o , 则所有与野值异类的样本 \mathbf{x}_i , 差值项

$(\mathbf{x}_i - \mathbf{x}_o)(\mathbf{x}_i - \mathbf{x}_o)^T$ 都将出现在目标函数中, 这样优化得到的投影矩阵 \mathbf{W} 将受 \mathbf{x}_o 的影响过大. 第二, 不考虑异类样本之间的近邻关系, 会导致相距较远的异类样本在目标函数中占据较大比重, 以致在处理某些多模问题时出现错误. 例如后节图 1 所示的多模数据分类问题, “o”和“+”代表的两类数据分别有左和右两个模态, LFDA 选择的投影直线接近水平, 导致不同类的数据重叠在一起. 这是因为“o”类的左(右)模态和“+”类的右(左)模态数据是异类样本, 为了使这些异类样本在投影之后距离较远, 优化过程选择了投影到水平方向. 其根本原因是 LFDA 在定义异类样本权值时没有考虑近邻关系.

MFA 排列出每类样本到异类样本之间的前 k_2 个距离最近的样本对, 并将这些样本对之间的权值设为 1, 这相当于仅考虑了部分位于边界的样本与异类样本之间的关系. 这将带来两个问题: 一是参数 k_2 较难选取, 因为它依赖于样本个数和数据维数, 样本越多、维数越大, 位于边界的样本点就越多, 就要选择更大的 k_2 才能充分描述边界样本和异类样本之间的关系. 二是在多模问题中可能会出现一类中某一个模态的样本比本类其它模态的样本到异类样本之间的距离要远, 那么该模态的样本与异类样本之间的距离将被忽视, 导致出现错误的结果. 例如图 3 所示的“o”类样本有左上和右下两个模态, 因为右下模态与“+”类样本较近, 因而异类样本距离的前 k_2 个距离最近的样本对中就没有左上模态的样本, 也就是说“o”类样本左上模态与“+”类样本之间的距离完全不在优化目标考虑的范围内, 那么 MFA 在选择投影方向时除了同类近邻之间的距离以外, 只考虑到“o”类样本右下模态与“+”类样本之间的距离要远, 以至于选择了向水平方向投影, 导致“o”类样本左上模态与“+”类样本完全混叠.

3 一种新的局部判别投影方法

为解决多模数据的分类问题, LFDA 和 MFA 将局部化思想融入到 LDA 当中, 但它们对异类样本间权值的设定方法使它们不能正确处理某些多模数据的分类问题. 本文提出了一种新的线性局部判别投影方法 LDP(Local Discriminant Projection), 有效地解决了 LFDA 和 MFA 存在的问题.

LDP 同时考虑了同类样本和异类样本间的近邻关系, 其基本思想是希望同类近邻样本在投影后尽量紧凑, 而异类近邻样本在投影后尽量分开. 这样

的思想可以通过两种目标函数来刻画,第一种是用样本间距离平方和来刻画;第二种是用样本类内和类间加权散度来刻画. 本文分别给出了这两种目标函数的分析和实验结果的比较.

设样本集记作 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]_{d \times n}$, $d \times m$ 变换矩阵 \mathbf{W} , 变换后样本 $\mathbf{Y} = \mathbf{W}^T \mathbf{X}$ 记作 $\mathbf{Y} = [y_1, y_2, \dots, y_n]_{m \times n}$.

第 1 种目标函数是最小化同类近邻样本之间的距离,同时最大化异类近邻样本之间的距离. 同类近邻样本之间的距离平方和为

$$f_1(\mathbf{W}) = \sum_{i,j=1}^n p_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 = \sum_{i,j=1}^n p_{ij} \|\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j\|^2, \quad (7)$$

$$p_{ij} = \begin{cases} 1 & \mathbf{x}_j \in N_k^+(i) \\ 0, & \text{其它} \end{cases}$$

其中 $N_k^+(j)$ 表示样本 \mathbf{x}_j 的同类 k 近邻的集合. 对 $f_1(\mathbf{W})$ 变形得到

$$\begin{aligned} f_1(\mathbf{W}) &= \sum_{i,j=1}^n p_{ij} \|\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j\|^2 \\ &= \sum_{i,j=1}^n p_{ij} (\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j)^T (\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j) \\ &= \text{tr} \left[\sum_{i,j=1}^n p_{ij} (\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j) (\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j)^T \right] \\ &= \text{tr} \left[\mathbf{W}^T \sum_{i,j=1}^n p_{ij} (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{W} \right] \\ &= \text{tr} [\mathbf{W}^T \mathbf{X}(\mathbf{D} - \mathbf{P})\mathbf{X}^T \mathbf{W}] \quad (8) \end{aligned}$$

其中 $\mathbf{P} = [p_{ij}]_{n \times n}$ 是权值矩阵, tr 表示矩阵的迹, \mathbf{D} 与 \mathbf{P} 的关系同式(2). 类似地, 异类样本之间的距离平方和为

$$f_2(\mathbf{W}) = \sum_{i,j=1}^n \tilde{p}_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 = \sum_{i,j=1}^n \tilde{p}_{ij} \|\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j\|^2, \quad (9)$$

$$\tilde{p}_{ij} = \begin{cases} 1 & \mathbf{x}_j \in N_k^-(i) \\ 0, & \text{其它} \end{cases}$$

其中 $N_k^-(j)$ 表示样本 \mathbf{x}_j 的异类 k 近邻的集合. 化简可得

$$f_2(\mathbf{W}) = \text{tr} [\mathbf{W}^T \mathbf{X}(\tilde{\mathbf{D}} - \tilde{\mathbf{P}})\mathbf{X}^T \mathbf{W}] \quad (10)$$

按照最小化同类近邻样本间距离,最大化异类近邻样本之间距离的思路,我们可把目标函数定义为

$$\min_{\mathbf{W} \in \mathbb{R}^{d \times m}} \frac{f_1(\mathbf{W})}{f_2(\mathbf{W})} = \min_{\mathbf{W} \in \mathbb{R}^{d \times m}} \frac{\text{tr} [\mathbf{W}^T \mathbf{X}(\mathbf{D} - \mathbf{P})\mathbf{X}^T \mathbf{W}]}{\text{tr} [\mathbf{W}^T \mathbf{X}(\tilde{\mathbf{D}} - \tilde{\mathbf{P}})\mathbf{X}^T \mathbf{W}]}$$

$$p_{ij} = \begin{cases} 1 & \mathbf{x}_j \in N_k^+(i) \\ 0, & \text{其它} \end{cases} \quad (11)$$

$$\tilde{p}_{ij} = \begin{cases} 1 & \mathbf{x}_j \in N_k^-(i) \\ 0, & \text{其它} \end{cases}$$

其中 $\mathbf{P} = [p_{ij}]_{n \times n}$ 和 $\tilde{\mathbf{P}} = [\tilde{p}_{ij}]_{n \times n}$ 是权值矩阵, \mathbf{D} 与 \mathbf{P}

的关系同式(2). 这个优化问题可以用文献[13]提出的迭代算法求得唯一最优解.

第二种目标函数是最小化类内加权散度,最大化类间加权散度. 散度反映了样本散布的离散程度,散度的一个简单度量就是散布矩阵行列式. 类内和类间散布矩阵可写成式(2)中 $\mathbf{S}^{(w)}$ 和 $\mathbf{S}^{(b)}$ 的形式,修改其中的系数 p_{ij} 和 \tilde{p}_{ij} 得到的 $\mathbf{S}^{(w)}$ 和 $\mathbf{S}^{(b)}$ 称为加权散布矩阵. 本文采用式(11)中定义的 p_{ij} 和 \tilde{p}_{ij} 得到加权散布矩阵,并采用加权散布矩阵的行列式作为散度的度量,那么优化目标就可纳入到框架式(4)中,写成

$$\min_{\mathbf{W} \in \mathbb{R}^{d \times m}} \frac{|\mathbf{W}^T \mathbf{X}(\mathbf{D} - \mathbf{P})\mathbf{X}^T \mathbf{W}|}{|\mathbf{W}^T \mathbf{X}(\tilde{\mathbf{D}} - \tilde{\mathbf{P}})\mathbf{X}^T \mathbf{W}|}$$

$$p_{ij} = \begin{cases} 1, & \mathbf{x}_j \in N_k^+(i) \\ 0, & \text{其它} \end{cases} \quad (12)$$

$$\tilde{p}_{ij} = \begin{cases} 1, & \mathbf{x}_j \in N_k^-(i) \\ 0, & \text{其它} \end{cases}$$

其中 $N_k^+(j)$ 表示样本 \mathbf{x}_j 的同类 k 近邻集合, $N_k^-(j)$ 表示样本 \mathbf{x}_j 的异类 k 近邻集合, $\mathbf{P} = [p_{ij}]_{n \times n}$ 和 $\tilde{\mathbf{P}} = [\tilde{p}_{ij}]_{n \times n}$ 是权值矩阵, \mathbf{D} 与 \mathbf{P} 的关系同式(2).

比较目标函数(11)和(12)不难看出,前者是类内加权散布矩阵和类间加权散布矩阵的迹的比值,而后者是两个矩阵行列式的比值. 散布矩阵的各特征值正比于样本散布超椭球体的各轴长,考虑到矩阵的迹是各特征值之和,而矩阵的行列式是各特征值之积,因此散布矩阵的迹正比于超椭球体各轴长之和,而散布矩阵的行列式正比于各轴长之积(超椭球体的体积). 尽管这两个指标都一定程度上反映出样本散布的程度,但相对而言还是超椭球体的体积反映得更加准确,实验结果也证实了这一点,在多数情况下,目标函数(12)获得了比(11)更好的识别结果.

4 实验结果与讨论

为评估各种算法的性能,本文完成了 3 类实验. 第 1 类是为了说明各种算法在原理上的区别而构造的 3 种多模人工数据;第 2 类是标准手写数字库 USPS 和 UCI,我们把 10 个数字分成两类,每类都是多模数据;第 3 类是 IDA 标准数据库共 13 个数据集,其中有 3 个数据集的数据呈现多模分布. 3 类实验中我们分别比较了 LDA、LPP(drtoolbox^[4])、LFDA(文献[11]提供代码)、MFA 以及本文算法 LDP 的工作原理. LDP 中选择近邻个数 $k=8$, MFA 中同类近邻个数 $k_1=8$, 异类距离最近的样本对个

数取 $k_2=10$, LPP、LFDA 选用源码中的默认参数. 实验结果表明无论是在人工数据还是标准数据库, 本文算法都具有优势.

4.1 人工数据

每个人工数据实验随机产生 200 组数据(3 个实验中分别有一组数据显示在图 1、图 2 和图 3 中, 图例中的“LDP”是以式(12)为目标函数, 在各表中记为“LDP2”), 每组数据包含两类各 100 个样本. 每次实验取一组作为训练, 另一组作为测试, 共进行 100 次. 对每种方法(LDA, LPP, LFDA, MFA, 本文算法 LDP2)计算了 100 次实验识别率的均值, 如表 1 所示. 3 个人工数据实验中的数据均产生自二元正态分布, 实验 1 中二元数据的协方差阵为二阶单位阵, 两类共计 4 个模态的数据均值分别为 $(-7, 3)$ $(7, 3)$ $(-7, -3)$ $(7, -3)$. 实验 2 中, 数据的协方差阵是对角元素为 1 和 36 的对角阵, 中间一类数据的均值为 $(0, 1)$, 两边一类的两个模态均值分别为 $(-5, 0)$ 和 $(5, 0)$. 实验 3 中数据协方差同实验 2, 右下角一类数据的均值为 $(-3, -5)$, 另一类两个模态数据的均值为 $(-3, 3)$ 和 $(3, -5)$.

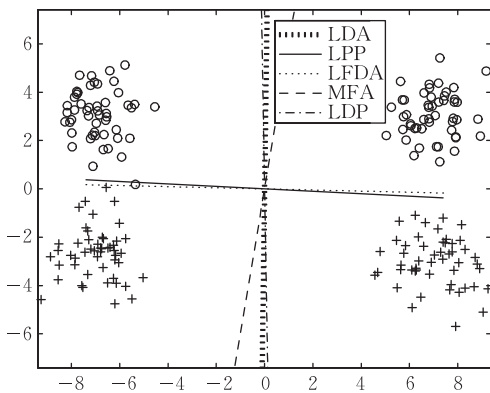


图 1 人工数据实验 1 的一组数据

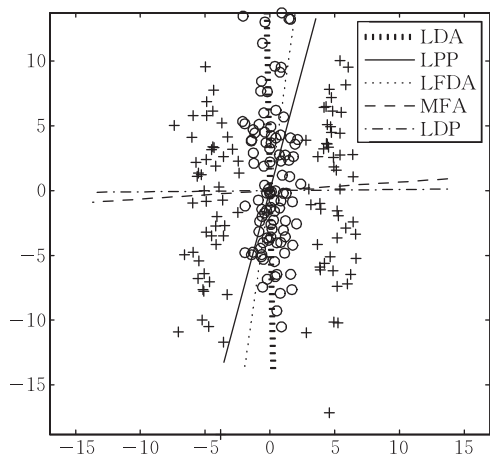


图 2 人工数据实验 2 的一组数据

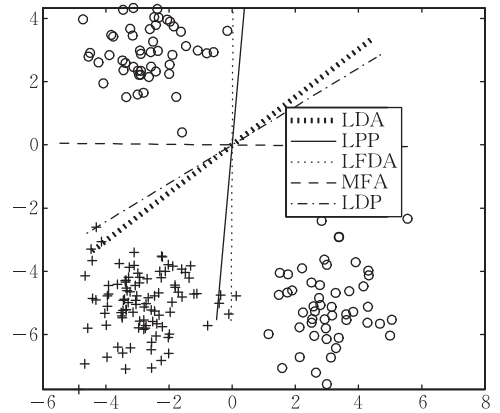


图 3 人工数据实验 3 的一组数据

表 1 各种算法在人工数据上的识别率

	识别率			平均值
	人工数据实验 1	人工数据实验 2	人工数据实验 3	
LDA	0.988	0.520	0.987	0.832
LPP	0.519	0.506	0.712	0.579
LFDA	0.512	0.514	0.686	0.571
MFA	0.953	0.972	0.780	0.902
LDP2	0.995	0.981	0.889	0.955

在给出定量的实验结果之前, 我们首先详细地分析 3 种人工数据实验数据的特点, 并讨论每种方法在处理此类数据时的行为以及各自成败的原因.

图 1 所示的多模数据分类问题中, “o” 和 “+” 代表的两类数据分别有左和右两个模态, 并且同类两个模态样本之间的距离大于不同类别相同模态样本之间的距离. 显然正确的投影方向是垂直方向, 实验结果显示 LDA、MFA 和 LDP 获得了正确结果, 而 LFDA 和 LPP 没有获得正确结果. LDA 的目标是最小化类内散度, 同时最大化类间散度, 显然向垂直方向投影, 把同类的两个模态样本投影到一起同时满足这两个需求. MFA 要使到处于边界的样本点与异类样本之间的距离较远, 本文算法 LDP 要使异类近邻样本之间的距离较远, 因此在这两种方法的目标函数中都会出现部分 “o” 类的左(右)模态和 “+” 类的左(右)模态样本之间的距离. 这些项的存在使优化过程选择了向垂直方向投影(如果向水平方向投影会使这些异类样本之间的距离变小). “o” 类的左(右)模态和 “+” 类的右(左)模态数据是异类样本, LFDA 没有考虑异类样本之间的近邻关系, 因此为了使所有这些异类样本在投影之后的距离平方和较远而选择了向水平方向投影. LPP 的优化目标是在投影后样本范数的加权平方和为常数的条件下, 希望相距较近的样本点在投影后仍比较近, 因此投影到数据方差较大的方向(本例中是水平方向)能

获得更优的目标函数。

图 2 所示的多模数据分类问题可简记做“夹心问题”，“o”类数据被夹在“+”类数据之间，而且每个模态分布都呈现为狭长的流型。显然正确的投影方向是水平方向，实验结果显示 MFA 和 LDP 获得了正确结果，而 LDA、LFDA 和 LPP 没有获得正确结果。同实验 1 类似，在 MFA 和 LDP 的目标函数中分别出现了“o”类数据和“+”类数据左和右两个模态之间样本的距离，这些项的存在导致了优化过程选择向水平方向投影，以防止这些异类样本之间距离变近。本例中两类样本各自的均值位置以及所有样本的均值位置十分接近，由于“o”类样本稍微偏上，因此 LDA 的投影方向为垂直方向。LPP 属于无监督投影，对它而言样本 3 个模态的数据就是一个狭长分布的数据团，而向数据方差较大的方向投影可以获得更大的样本范数的加权平方和，因此 LPP 会选择向垂直方向投影。如果数据不是这样狭长并且靠得很近，LFDA 和 LPP 是可以获得正确结果的（如文献[11]中 Figure1(c)所示），而对于狭长的数据 LFDA 之所以出错还是因为这种狭长数据中相距较远的异类样本的距离和在目标函数中占据了过大的比重，是 LFDA 在异类中不考虑近邻关系的结果。

图 3 所示的多模分类问题中，“+”类数据和“o”类数据左上和右下两个模态成钝角之势。为了使两类数据在投影后分开，正确的投影方向应该是近似 45° 直线或 135° 直线。只有 LDA 和本文算法 LDP 获得了正确结果。LDP 能获得正确结果仍然是得益于目标函数中分别出现了“+”类数据样本和“o”类数据左上和右下两个模态之间样本的距离，这些项导致优化过程选择了正确的投影方向。LDA 倾向于将同类不同模态的样本投影到一起以获得较小的类内散度，因而选择接近 45° 的直线。LPP 选择向垂直方向投影是因为垂直方向数据的方差比其它方向更大，而且“+”类数据样本和“o”类右下数据之间存在近邻关系，要求投影后保持近邻关系。MFA 出现错误结果是因为异类样本距离的前 k_2 个距离最近的样本对中并没有“o”类数据左上模态的样本，也就是说“o”类样本左上模态与“+”类样本之间的距离完全不在优化目标考虑的范围之内，以至于选择了向水平方向投影，以保证同类近邻样本之间距离近以及“+”类样本与“o”类样本右下模态之间的距离较远。LFDA 选择了向垂直方向投影是为了保持“+”类样本与“o”类样本左上模态之间这些异类样本距离较远。

表 1 是各种方法在 3 个人工数据实验（每个实验进行了 100 组）中的平均识别率（投影后用最近邻分类器进行分类），从以上人工数据的分析以及表 1 不难看出，本文算法 LDP 有效地克服了 LFDA 和 MFA 在处理一些特殊多模问题时暴露出的缺点，实验结果具有明显优势。

4.2 标准手写数字库

4.2.1 标准库简介

UCI Multiple Features Data Set 手写数字库是 UCI 机器学习数据集^① (University of California, Irvine Machine Learning Repository) 中的一个标准数据库，包含了‘0’~‘9’ 10 类数字的每类 200 个共计 2000 个样本。每个样本提取出包括付立叶系数、轮廓相关函数、KL 系数、像素值均值、泽尔尼克 (Zernike) 矩、形态特征 6 类共计 649 个特征。

美国国家邮政局 USPS 手写数字库^② 包括‘0’~‘9’ 10 类数字的共计 7291 个训练样本和 2007 个测试样本，每个样本用 256 个灰度值来表示。

4.2.2 在标准库上的实验

在 UCI 和 USPS 手写数字库上我们对 LDP1 (目标函数(11))、LDP2 (目标函数(12))、LDA、LFDA 和 MFA 进行了比较。从标准库中每个数字取 100 个样本作为训练样本，取 100 个样本作为测试样本。为公平比较，每个样本先用 PCA 降到相同维数（对于 UCI，每个样本先用 PCA 降维到 150 维；对于 USPS，每个样本先用 PCA 降维到 100 维），再用各算法降到 10 维，最后用最近邻分类器进行分类。实验结果见表 2 第 2 列和第 3 列，本文算法 LDP2 均获得次优解，并与最优结果十分接近。

表 2 各种算法标准手写数字库上的识别率

	识别率								
	UCI	USPS	UCI1	UCI2	UCI3	USPS1	USPS2	USPS3	
LDP1	0.967	0.861	0.971	0.987	0.965	0.901	0.898	0.915	
LDP2	0.983	0.893	0.990	0.998	0.986	0.943	0.950	0.939	
LDA	0.985	0.876	0.956	0.950	0.945	0.802	0.897	0.817	
LFDA	0.982	0.895	0.983	0.995	0.978	0.877	0.906	0.900	
MFA	0.983	0.870	0.896	0.990	0.861	0.834	0.851	0.824	

为比较在多模数据上的性能，实验中用多类手写数字构造两类多模数据。实验 UCI1 和 USPS1 中的两类是数字 [0 1 2 3 4] 和 [5 6 7 8 9]（实验结果见表 2 第 4 列和第 7 列），实验 UCI2 和 USPS2 中的两类是数字 [1 3 5 7 9] 和 [0 2 4 6 8]（实验结果见

① <http://archive.ics.uci.edu/ml>

② <http://www-i6.informatik.rwth-aachen.de/~keyser/usps.html>

表 2 第 5 列和第 8 列), 实验 UCI3 和 USPS3 中的两类是[0 1 2 6 9]和[3 4 5 7 8](实验结果见表 2 第 6 列和第 9 列). 实验结果显示本文算法 LDP2 在多模数据上有较明显的性能优势. 同时实验结果还表明, 目标函数(12)要明显优于目标函数(11), 这也验证了我们此前对于两种目标函数的分析.

4.3 IDA 标准数据库

4.3.1 标准库简介

IDA 标准库^①最早在文献[14]中使用, 后来成为模式识别中广泛使用的标准数据. 它包含 13 个标准数据集以及每个数据集训练样本和测试样本的多个标准划分, 每个数据集都是两类数据, 其中表 3 中用“*”标记的 3 个数据集中每类数据呈现多模分布. 表 3 给出了各数据集的基本信息.

表 3 IDA 标准数据库的基本信息

数据集	样本维数	训练样本个数	测试样本个数	标准划分个数
banana*	2	400	4900	100
breast-cancer	9	200	77	100
diabetis	8	468	300	100
flare-solar	9	666	400	100
german	20	700	300	100
heart	13	170	100	100
image	18	1300	1010	20
ringnorm	20	400	7000	100
splice	60	1000	2175	20
thyroid*	5	140	75	100
titanic	3	150	2051	100
twonorm	20	400	7000	100
waveform*	21	400	4600	100

4.3.2 在 IDA 数据库上的实验

为公平比较, 本文实验中将 13 个标准集上的数据都降至 2 维, 然后采用最近邻分类器进行分类. 在表 4 中最后一列给出了不降维直接用最近邻分类器

表 4 各种算法在 IDA 标准库上的识别率

	识别率					
	LDA	LPP	LFDA	MFA	LDP2	不降维
banana*	0.8635	0.8637	0.8625	0.8636	0.8637	0.8636
breast-cancer	0.6608	0.6464	0.6534	0.6339	0.6500	0.6730
diabetis	0.6890	0.6392	0.6630	0.6172	0.6784	0.6988
flare-solar	0.6074	0.6026	0.6113	0.5935	0.6095	0.6078
german	0.6880	0.6144	0.6792	0.6066	0.6594	0.7054
heart	0.7784	0.7641	0.7800	0.6316	0.7717	0.7684
image	0.7809	0.8275	0.8342	0.7611	0.8442	0.9662
ringnorm	0.7268	0.6493	0.7319	0.6590	0.7285	0.6497
splice	0.7956	0.6220	0.8351	0.6203	0.8370	0.7115
thyroid*	0.8560	0.9451	0.9336	0.9385	0.9517	0.9564
titanic	0.6692	0.6687	0.6687	0.6647	0.6700	0.6689
twonorm	0.9642	0.9628	0.9657	0.8709	0.9653	0.9332
waveform*	0.8144	0.8828	0.8831	0.6627	0.8798	0.8417
平均	0.7611	0.7453	0.7771	0.7018	0.7776	0.7727

的分类结果, 该结果没有参与挑选最佳识别率. 结果表明本文算法 LDP 在 13 类数据中有 5 类达到最优, 在 3 类多模数据中有 2 类达到最优, 此外还在 13 类的平均识别率上获得了最优.

5 总 结

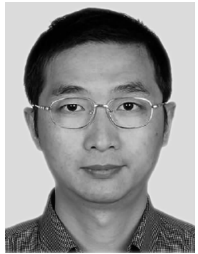
为解决多模数据的分类问题, LFDA 和 MFA 都将局部化思想融入到 LDA 当中. LFDA 在构造权值矩阵时完全不考虑异类样本间的局部性, 导致远距离的异类样本严重影响最终的投影方向. 由于多模数据通常有不止一条边界, 而 MFA 仅考虑最近异类样本间的关系, 相当于只考虑了部分边界. 本文深入分析了两种算法在工作原理上的弱点和问题, 提出了一种新的线性局部判别投影方法 LDP. 该方法在设计异类样本间权值时也考虑到近邻的因素, 其基本思想是希望同类近邻样本在投影后尽量紧凑, 而异类近邻样本在投影后尽量分开. 该算法有效解决了 LFDA 和 MFA 存在的问题, 在人工数据集和标准数据集上都取得较好效果.

参 考 文 献

- [1] Tenenbaum J B. Mapping a manifold of perceptual observations//Proceedings of the 10th Advances in Neural Information Processing Systems. Colorado, USA, 1998: 682-688
- [2] Belkin M, Niyogi P. Laplacian eigenmap and spectral techniques for embedding and clustering//Proceedings of the 14th Advances in Neural Information Processing Systems. British Columbia, Canada, 2001: 585-591
- [3] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. Science, 2000, 290(5500): 2323-2328
- [4] Maaten L J, Postma E O, Herik H J. Dimensionality reduction: A comparative review. Netherlands: Tilburg University, TiCC-TR: 2009-005, 2009
- [5] Theodoridis S, Koutroumbas K. Pattern Recognition. 4th Edition. USA: Elsevier Science, 2009
- [6] He X, Niyogi P. Locality preserving projections//Proceedings of the 16th Advances in Neural Information Processing Systems, Vancouver, Canada, 2003: 153-160
- [7] He X, Cai D, Yan S et al. Neighborhood preserving embedding//Proceedings of the 10th IEEE International Conference on Computer Vision. Beijing, China, 2005: 1208-1213

① <http://www.fml.tuebingen.mpg.de/Members/raetsch/benchmark>

- [8] Fu Y, Huang T S. Locally linear embedded eigenspace analysis. UIUC, USA: IFP-TR UIUC2005, 2005
- [9] Cai D, He X, Han J. Isometric projection//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver, Canada, 2007; 528-533
- [10] Qiao L, Chen S, Tan X. Sparsity preserving projections with applications to face recognition. Pattern Recognition, 2010, 43(1): 331-341
- [11] Sugiyama M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. Journal of Machine Learning Research, 2007, 8: 1027-1061
- [12] Yan S, Xu D, Zhang B et al. Graph embedding and extensions: A general framework for dimensionality reduction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1): 40-51
- [13] Wang H, Yan S, Xu D et al. Trace ratio vs. ratio trace for dimensionality reduction//Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. Minnesota, USA, 2007; 17-22
- [14] Ratsch G, Onoda T, Muller K R. Soft margins for adaboost. Machine Learning, 2001, 42(3): 287-320



XIE Jun, born in 1973, Ph. D., associate professor. His research interests include intelligent information processing and network management.

LIU Jian, born in 1976, lecturer. Her research interests include intelligent information processing and pattern recognition.

Background

Dimensionality reduction is the transformation of data from a high-dimensional space into a low-dimensional space, preserving most of “intrinsic information” contained in the data. It is an important preprocessing step in high-dimensional data compression, visualization and classification. Dimensionality reduction can be classified into unsupervised dimensionality reduction and supervised dimensionality reduction, the latter is also called discriminant analysis. The most successfully used linear discriminant analysis method is LDA (Linear Discriminant Analysis). However, LDA does not work well in the classification of multimodal data, which is the focus of our work.

To solve the problem of multimodal data classification, the idea of localization is introduced into discriminant analysis, known as local discriminant analysis. LFDA and MFA are two recently proposed and widely used local linear discriminant analysis methods. However, neighbor relationships between samples of different classes are not fully taken into consideration in the two methods, which results in errors in the classification of some types of multimodal data. To overcome the shortcomings of LFDA and MFA, we proposed a new local discriminant projection method. The un-

derlying idea of the new method is that the desired projection should make neighbors of the same class close and neighbors of different classes apart. Experiment results show that the new method overcomes the drawbacks of LFDA and MFA, and achieves good performance on the synthetic data, USPS, UCI standard handwriting digital data sets and IDA standard data sets.

The work is supported by National Natural Science Foundation of China “Research on Semantic Image Segmentation with Tree Structured Model” (No. 61101202). Semantic image segmentation is one of the most important and challenging problems in image understanding and computer vision, it has been receiving extensive attention. CRF based framework, in which the unary potentials are usually defined directly by the probability outputs of some classifiers, is the most popular approach to semantic image segmentation. The final accuracy of segmentation is mainly determined by accuracy of the unary classifiers. As we know, effective feature extraction is crucial to reliable classification, so we pay special attention to various feature extraction methods, including dimension reduction techniques, which is the topic this paper addresses.