

基于最大散度差鉴别准则的 自适应分类算法¹⁾

宋枫溪¹ 张大鹏² 杨静宇³ 高秀梅⁴

¹(哈尔滨工业大学深圳研究生院 深圳 518055)

²(香港理工大学)

³(南京理工大学 南京 210094)

⁴(淮阴师范学院 淮阴 223001)

(E-mail: fx_song@sina.com)

摘 要 首先证明了, 当类内散布矩阵非奇异时, 特定参数值 c_0 下最大散度差的最优鉴别方向等同于 Fisher 最优鉴别方向; 其次, 给出了最大散度差分类算法的识别率随参数 C 变化的曲线. 该曲线通常为脉冲曲线. 随着参数 C 的增大, 识别率也逐渐增大. 当参数 C 增大到 c_0 时, 识别率达到最大值. 另外, 以往的研究成果表明: 当类内散布矩阵奇异时, 最大散度差鉴别准则逐步逼近大间距线性投影准则. 而且, 随着参数 C 的不断增大, 最大散度差分类算法的识别率也单调增大并最终稳定到大间距线性投影分类算法的识别率上. 为此, 我们提出了基于最大散度差鉴别准则的自适应分类算法. 新算法可以根据训练样本的特性 (类内散布矩阵是否奇异) 自动选择恰当的参数 C . 在 UCI 机器学习数据库上的 6 个数据集以及 AR 人脸图像数据库上的测试结果表明, 自适应最大散度差分类算法具有良好的分类性能.

关键词 最大散度差, 大间距线性投影, Fisher 鉴别准则, 自适应算法, 机器学习, 人脸识别中图分类号 TP39

Adaptive Classification Algorithm Based on Maximum Scatter Difference Discriminant Criterion

SONG Feng-Xi¹ ZHANG David² YANG Jing-Yu³ GAO Xiu-Mei⁴

¹(Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055)

²(Hong Kong Polytechnic University)

³(Nanjing University of Science and Technology, Nanjing 210094)

⁴(Huaiyin Normal College, Huaiyin 223001)

(E-mail: fx_song@sina.com)

Abstract In this paper we first prove that the optimal discriminant direction of Maximum scatter difference (MSD) discriminant criterion with a certain value c_0 is equivalent to the optimal Fisher discriminant direction. Second, sample recognition rate curves of MSD are illustrated. The recognition rate curve is usually a pulse curve when the within-class scatter matrix is nonsingular. With the increase of parameter C , the recognition rate of MSD also increases. The recognition rate of MSD achieves its maximum when C is equal to c_0 . In addition, former study showed that, when the within-class scatter matrix is singular, MSD criterion is approaching the large margin linear projection criterion as parameter C increases.

1) 国家自然科学基金 (60472060) 和江苏省高校自然科学研究指导性项目 (04KJD520037) 资助
Supported by National Natural Science Foundation of P. R. China (60472060) and Natural Foundation of Education Bureau of Jiangsu Province of P. R. China (04KJD520037)
收稿日期 2005-6-29 收修改稿日期 2006-3-10
Received June 30, 2005; in revised form March 10, 2006

Moreover, the recognition rate curve of MSD is non-decreasing. Thus, an adaptive classification algorithm based on maximum scatter difference discriminant criterion is proposed based on these facts. The new algorithm can tune parameter C automatically according to the characteristics of training samples. Experiment conducted on 6 datasets from UCI Machine Learning Repository and AR face database demonstrates that the adaptive classification algorithm for maximum scatter difference has good classification property.

Key words Maximum scatter difference, large margin linear projection, Fisher discriminant criterion, adaptive algorithm, machine learning, face recognition

1 引言

最大散度差鉴别准则是一种新的线性鉴别准则^[1], 同 Fisher 鉴别准则^[2] 一样, 最大散度差准则试图寻找一个最优投影方向, 使得不同类别的高维模式样本沿该方向投影后, 能最大限度地分离开来. 与 Fisher 鉴别准则不同的是, 最大散度差准则采用广义散度差 — 类间散度减去 C 倍的类内散度 (C 为模型参数, 通常为非负实数), 作为投影后数据的可分性度量, 而不再沿用广义 Rayleigh 商, 从而有效避免了 Fisher 鉴别准则面临的类内散布矩阵的奇异性问题.

基于最大散度差鉴别准则的线性分类器 — 最大散度差分类器, 在人脸识别领域得到了良好运用^[1]. 实验结果表明, 其分类效果优于各种传统的人脸识别方法如 Eigenfaces^[3] 和 Fisherfaces^[4], 甚至可以与具有很强泛化能力的支持向量机^[5] 相媲美. 而且, 最大散度差分类器在人脸数据集上的分类性能并不明显依赖参数 C 的选择.

避免了奇异性问题、具有优良的分类效果、无须对参数进行烦琐的调试等优点, 使得最大散度差分类器成为一种新的富有成效的人脸识别方法.

遗憾的是, 当我们试图将最大散度差分类器推广运用到类内散布矩阵为非奇异矩阵的一些模式识别问题时, 发现其分类性能对参数 C 的取值非常敏感. 参数 C 的微小变化可能导致识别率的重大改变. 此时, 选择恰当的参数 C 对于取得良好的分类效果至关重要. 虽然人们可以通过在验证集 (Validation set) 上对参数 C 进行调试来弥补最大散度差分类器此时的不足, 但是要彻底解决这个问题必须从最大散度差分类算法本身入手.

本文首先简要介绍了最大散度差鉴别准则和最大散度差分类器, 然后探讨了最大散度差鉴别准则与 Fisher 鉴别准则以及大间距线性投影鉴别准则之间的关系, 最后提出了一种自适应最大散度差分类算法.

在 UCI 机器学习数据库^[6] (UCI machine learning repository) 的 6 个数据集以及 AR 人脸图像数据库^[7] 上的测试结果表明, 自适应最大散度差分类器在类内散布矩阵为奇异矩阵和非奇异矩阵两种情况下, 能分别近似具有大间距线性投影分类器和 Fisher 线性分类器的优良识别性能.

2 最大散度差鉴别准则与最大散度差分类器

设 ω_1 和 ω_2 是两个模式类, 模式 $\mathbf{x} \in R^d$ 为 d 维实向量, 第 i 类训练样本的个数为 N_i , 样本均值 \mathbf{m}_i 、类间散布矩阵 S_b 和类内散布矩阵 S_w 分别定义为

$$\mathbf{m}_i = \frac{1}{N_i} \sum_{\mathbf{x} \in \omega_i} \mathbf{x}, \quad i = 1, 2 \quad (1)$$

$$S_b = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T \quad (2)$$

$$S_w = \sum_{i=1}^2 \sum_{\mathbf{x} \in \omega_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \quad (3)$$

将模式样本沿某个方向 $\mathbf{w} \in R^d (\mathbf{w} \neq 0)$ 进行投影, 投影后样本的类间散度 \tilde{S}_b 和类内散度 \tilde{S}_w 分别为

$$\tilde{S}_b = \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \quad (4)$$

$$\tilde{S}_w = \frac{\mathbf{w}^T S_w \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \quad (5)$$

类间散度和类内散度是衡量数据可分性的两个常用指标: 类间散度愈大, 数据的可分性愈好; 类内散度愈小, 数据的可分性愈好. 因此, 寻找“使得投影后模式样本的可分性达到最大”的最优鉴别方向的问题, 可以归结为求解以下多目标规划问题

$$\max_{\mathbf{w} \neq 0} \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \quad (6)$$

$$\min_{\mathbf{w} \neq 0} \frac{\mathbf{w}^T S_w \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \quad (7)$$

显然, 我们无法直接求解 (6-7) 两式确定的最优化问题, 首先需要将其转化为单目标规划问题. 化多目标规划问题为单目标规划问题的办法有很多, 评价函数法^[7] 是其中一种行之有效的方法. 运用评价函数法中的乘除法, 我们可以得到以下最优化模型

$$\max_{\mathbf{w} \neq 0} \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}} \quad (8)$$

它就是著名的 Fisher 鉴别准则.

Fisher 鉴别准则将类间散度与类内散度的比值作为衡量投影后数据可分性的一个综合性度量, 使得这一综合可分性度量达到最大的向量即为 Fisher 最优投影方向.

如果运用评价函数法中的线性加权法, 我们即可以得到最大散度差 (Maximum scatter difference, 简记为 MSD) 鉴别准则

$$\max_{\mathbf{w} \neq 0} \frac{\mathbf{w}^T S_b \mathbf{w} - C \cdot \mathbf{w}^T S_w \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \quad (9)$$

最大散度差鉴别准则将类间散度与类内散度的广义差作为衡量投影后数据可分性的一个综合性度量, 使得这一综合可分性度量达到最大的向量即为 MSD 最优投影方向.

当类内散布矩阵 S_w 为奇异矩阵时, (8) 式成为一个不适定问题 (Ill-posed problem), 从而限制了 Fisher 鉴别准则的应用范围. 最大散度差鉴别准则却巧妙地回避了困扰 Fisher 鉴别准则的奇异性问题, 同时适用于类内散布矩阵奇异和非奇异两种不同场合.

最大散度差鉴别准则的最优投影方向的确定也非常简单, 可以归结为求矩阵的特征向量问题.

引理 1^[4]. 矩阵 $S_b - C \cdot S_w$ 最大特征值对应的特征向量即为最大散度差鉴别准则的最优投影方向.

记 $\mathbf{w}^*(C)$ 为 (9) 确定的最优投影方向, b 和 b_1 分别为所有训练样本和第一类训练样本沿 $\mathbf{w}^*(C)$ 轴投影后的样本均值, 即

$$b = \frac{1}{N_1 + N_2} \sum_{\mathbf{x} \in \omega_1 \cup \omega_2} \mathbf{w}^*(C)^T \mathbf{x} \quad (10)$$

$$b_1 = \frac{1}{N_1} \sum_{\mathbf{x} \in \omega_1} \mathbf{w}^*(C)^T \mathbf{x} \quad (11)$$

则, 基于最大散度差鉴别准则的最大散度差分类器定义为:

$$f(\mathbf{x}) = \text{sign}(\mathbf{w}^*(C)^T \mathbf{x} - b) \cdot \text{sign}(b_1 - b) \quad (12)$$

其中, sign 为符号函数, $f(\mathbf{x}) = 1$ 表示 \mathbf{x} 属于 ω_1 , $f(\mathbf{x}) = -1$ 表示 \mathbf{x} 属于 ω_2 .

根据 $\mathbf{w}^*(C)$ 的含义, 当训练样本沿该轴投影时, 两类样本被最大限度地分隔开了. 作为训练样本平均投影值的 b , 因而是两类样本的一个理想的分界点, 而系数 $\text{sign}(b_1 - b)$ 则反映了不同类别训练样本与该分界点之间的关系. $\text{sign}(b_1 - b) > 0$ 意味着落在分界点右侧的样本绝大部分来自 ω_1 , 而 $\text{sign}(b_1 - b) < 0$ 则意味着落在分界点右侧的样本绝大部分来自 ω_2 . (12) 式中系数 $\text{sign}(b_1 - b)$ 的作用在于确保来自 ω_1 的样本的函数值始终为 1, 而无论它们落在分界点的左侧还是右侧.

3 最大散度差分类器与其他分类器之间的关系

记

$$F(c) = \frac{\mathbf{w}^*(c)^T (S_b - c \cdot S_w) \mathbf{w}^*(c)}{\mathbf{w}^*(c)^T \mathbf{w}^*(c)} = \max_{\mathbf{w} \neq 0} \frac{\mathbf{w}^T (S_b - c \cdot S_w) \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \quad (13)$$

显然, $F(c)$ 是 c 的一个实值函数. 由广义 Rayleigh 商的极值性知, $F(c)$ 等于矩阵 $S_b - c \cdot S_w$ 的最大特征值.

定理 1. $F(c)$ 是 c 的单调减函数. 特别, 当类内散布矩阵 S_w 为非奇异矩阵时, $F(c)$ 是 c 的严格单调减函数, 且 $\lim_{c \rightarrow \infty} F(c) = -\infty$; 当 S_w 为奇异矩阵时, $\lim_{c \rightarrow \infty} F(c) = \max_{\mathbf{w}^T S_w \mathbf{w} = 0} \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T \mathbf{w}}$.

证明. 设 $c_1 < c_2$, \mathbf{w}_1 和 \mathbf{w}_2 分别是矩阵 $(S_b - c_1 \cdot S_w)$ 和 $(S_b - c_2 \cdot S_w)$ 最大特征值所对应的单位特征向量.

$$\begin{aligned} F(c_1) &= \mathbf{w}_1^T (S_b - c_1 \cdot S_w) \mathbf{w}_1 \geq \mathbf{w}_2^T (S_b - c_1 \cdot S_w) \mathbf{w}_2 = \\ & \mathbf{w}_2^T (S_b - c_2 \cdot S_w) \mathbf{w}_2 + (c_2 - c_1) \mathbf{w}_2^T S_w \mathbf{w}_2 = F(c_2) + (c_2 - c_1) \mathbf{w}_2^T S_w \mathbf{w}_2 \end{aligned}$$

由于 S_w 为半正定矩阵, 因而有 $\mathbf{w}_2^T S_w \mathbf{w}_2 \geq 0$. 故 $F(c_1) \geq F(c_2)$, 即 $F(c)$ 是 c 的单调减函数.

特别的, 若 S_w 非奇异, 则 S_w 为正定矩阵. 因而, 对于任意的单位向量 \mathbf{w} , 恒有

$$\mathbf{w}^T S_w \mathbf{w} \geq \lambda_w > 0 \quad (14)$$

其中, λ_w 为类内散布矩阵 S_w 的最小特征值.

此时, $F(c_1) > F(c_2)$, 即 $F(c)$ 是 c 的严格单调减函数.

记 λ_b 为矩阵 S_b 的最大特征值, 则对于任意的单位向量 \mathbf{w} , 恒有

$$\mathbf{w}^T S_w \mathbf{w} \leq \lambda_b \quad (15)$$

由 (14), (15) 知

$$F(c) = \mathbf{w}^*(c)^T S_b \mathbf{w}^*(c) - c \cdot \mathbf{w}^*(c)^T S_w \mathbf{w}^*(c) \leq \lambda_b - c \cdot \lambda_w \quad (16)$$

不难看出, $\lim_{c \rightarrow \infty} F(c) = -\infty$.

当 S_w 为奇异矩阵时, 文献 [1] 的定理 2 已经证明当参数 C 趋向无穷大时, 由最大散度差鉴别准则确定的最优投影方向等同于最优化模型 $\max_{\mathbf{w}^T S_w \mathbf{w} = 0} \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T \mathbf{w}}$ 确定的最优投影方向. 因而有 $\lim_{c \rightarrow \infty} F(c) = \max_{\mathbf{w}^T S_w \mathbf{w} = 0} \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T \mathbf{w}}$. \square

引理 2. $F(c)$ 是自变量 c 的连续函数.

证明. $F(c)$ 是矩阵 $S_b - c \cdot S_w$ 的最大特征值, 因而是该矩阵各个元素的连续函数^[8]. 而该矩阵的所有元素均是 c 的连续函数.

故 $F(c)$ 是 c 的连续函数. \square

定理 2. 若 S_w 为非奇异矩阵, 则方程 $F(c) = 0$ 存在唯一正实根 c_0 , 且矩阵 $S_b - c_0 \cdot S_w$ 的最大特征值所对应的特征向量即为 Fisher 最优鉴别向量.

证明. 由 (13) 式知, $F(0) = \max_{\mathbf{w} \neq 0} \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T \mathbf{w}} = \lambda_b > 0$.

另外, 由 (16) 式知, 当 $c^* > \frac{\lambda_b}{\lambda_w}$ 时, $F(c^*) < 0$.

由于 $F(c)$ 是自变量 c 的连续函数, 故在区间 $(0, c^*)$ 内必然存在一点 c_0 , 使得 $F(c_0) = 0$.

又由于 $F(c)$ 是自变量 c 的严格单调函数, 故 c_0 是唯一的.

由文献 [2] 知, Fisher 最优投影方向为 $\mathbf{w}^* = S_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$.

记 $c_1 = \frac{\mathbf{w}^{*T} S_b \mathbf{w}^*}{\mathbf{w}^{*T} S_w \mathbf{w}^*}$, 显然有 $c_1 > 0$.

由 Fisher 最优投影方向的含义知, 对于任意一个投影方向 \mathbf{w} , 均有 $\frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}} \leq c_1$.

由此不难证明, 对于任意一个投影方向 \mathbf{w} , 均有

$$\frac{\mathbf{w}^{*T} (S_b - c_1 \cdot S_w) \mathbf{w}^*}{\mathbf{w}^{*T} \mathbf{w}^*} = 0 \geq \frac{\mathbf{w}^T (S_b - c_1 \cdot S_w) \mathbf{w}}{\mathbf{w}^T \mathbf{w}}$$

这意味着, \mathbf{w}^* 同时为参数 c_1 下的 MSD 最优投影方向, 即矩阵 $S_b - c_1 \cdot S_w$ 的最大特征值所对应的特征向量 (至多相差一个常数倍数).

因而有, $F(c_1) = \frac{\mathbf{w}^{*T} (S_b - c_1 \cdot S_w) \mathbf{w}^*}{\mathbf{w}^{*T} \mathbf{w}^*} = 0$.

由于 $F(c)$ 的零根唯一, 故 $c_1 = c_0$. \square

该定理表明, 若 S_w 为非奇异矩阵, 当参数 C 取值为矩阵 $S_w^{-1} S_b$ 的最大特征值时, MSD 的最优投影方向等同于 Fisher 最优投影方向.

$F(c)$ 随自变量 c 变化的函数曲线如图 1 所示.

图 1(a) 是根据 Cancer 数据集 (参见表 2) 上的实验结果绘制的, 它形象地表明了, 当类内散布矩阵为非奇异矩阵时, 函数 $F(c)$ 随自变量 c 变化的情况. 随着参数 c 的不断增大, 函数值 $F(c)$ 单调趋于负无穷大. 图 1(b) 是根据 AR 人脸图像数据集上的实验结果绘制的, 它表明当类内散布矩阵为奇异矩阵时, 函数值 $F(c)$ 随着参数 c 的不断增大, 单调递减到某个极限值.

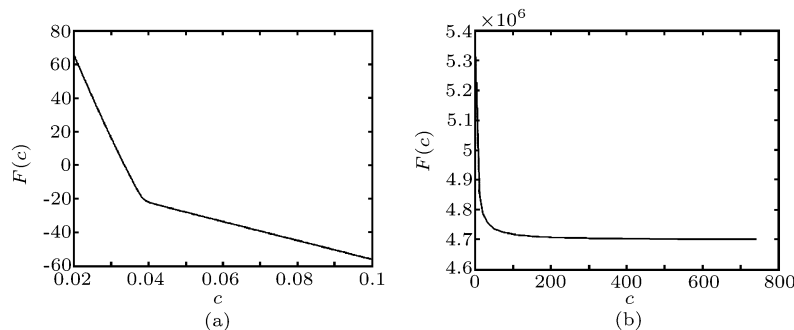


图 1 $F(c)$ 随自变量变化的函数曲线: (a) S_w 非奇异; (b) S_w 奇异
Fig. 1 The curve of $F(c)$: (a) S_w is nonsingular; (b) S_w is singular

定理 3^[1]. 若 S_w 为奇异矩阵, 则当参数 C 趋向无穷大时, MSD 的最优投影方向等价于由以下最优化模型确定的最优投影方向

$$\max \mathbf{w}^T S_b \mathbf{w} \quad (17)$$

$$\begin{cases} \mathbf{w}^T S_w \mathbf{w} = 0 \\ \|\mathbf{w}\| = 1 \end{cases} \quad (18)$$

对应于 (17), (18) 两式的鉴别准则称之为大间距线性投影 (Large margin linear projection, 简记为 LMLP) 准则.

我们知道, 如果类内散布矩阵 S_w 为奇异矩阵, 则其零空间 $N(S_w)$ 的维数不小于 1. 根据零空间的性质, 对于任意一个投影方向 $\mathbf{w} \in N(S_w)$, 均有 $\mathbf{w}^T S_w \mathbf{w} = 0$. 这意味着, 当训练样本沿这些方向投影后, 来自同一类别的训练样本都集中到一个点上. 这是比较理想的一种情况. 大间距线性投影准则试图在这些比较理想的投影方向中再找到一个同时使得类间散度 $\frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T \mathbf{w}}$ 达到最大的方向 (因为 $\max_{\|\mathbf{w}\|=1} \mathbf{w}^T S_b \mathbf{w}$ 等同于 $\max_{\|\mathbf{w}\|=1} \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T \mathbf{w}}$).

在人脸图像数据集上的实验结果表明, 基于大间距线性投影准则的分类算法具有良好的分类性能, 其识别率可以同线性支持向量机相媲美^[9].

另外, 实验表明: 当类内散布矩阵 S_w 为非奇异矩阵时, 最大散度差分类算法的识别率随参数 C 变化的曲线近似为一脉冲曲线. 随着参数 C 的增大, 识别率也逐渐增大. 当参数 C 增大到 c_0 时, 识别率达到最大; 当类内散布矩阵 S_w 为奇异矩阵时, 该识别率曲线近似为一单调递增曲线. 两种场合下, 最大散度差分类算法的识别率随参数 C 变化的曲线分别如图 2 和图 3 所示.

图 2 是根据 Australian 数据集 (参见表 2) 上的实验结果绘制的, 它形象地表明了, 当类内散布矩阵为非奇异矩阵时, 最大散度差分类算法的识别率随自变量 c 变化的情况.

图 3 是根据 ORL 人脸图像^[8] 数据集上的实验结果绘制的, 它近似地表明了, 当类内散布矩阵为奇异矩阵时, 最大散度差分类算法的识别率随着自变量 c 的增大单调增大直到趋近于某个极限值.

因此, 最大散度差鉴别准则实际上将仅适用于 S_w 非奇异场合的 Fisher 鉴别准则和仅适用于 S_w 奇异场合的大间距线性投影鉴别准则有机地统一了起来. 但这种统一的前提是: 当 S_w 非奇异时, 参数 C 应充分接近矩阵 $S_w^{-1} S_b$ 的最大特征值, 使得 MSD 的最优投影方向与 Fisher 最优投影方向相一致; 当 S_w 奇异时, 参数 C 应充分大, 使得 MSD 的最优投影方向与大间距线性投影的最优投影方向相一致.

为此, 我们提出了基于最大散度差鉴别准则的自适应分类算法, 使得 S_w 非奇异时, 参数 C 充分接近矩阵 $S_w^{-1}S_b$ 的最大特征值, 而当 S_w 奇异时, 参数 C 足够大.

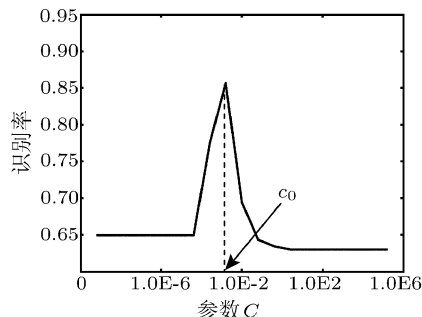


图 2 S_w 非奇异时最大散度差分类器的识别率随参数 C 变化的函数曲线

Fig. 2 The recognition rate of MSD classifier vs. the parameter C when S_w is nonsingular

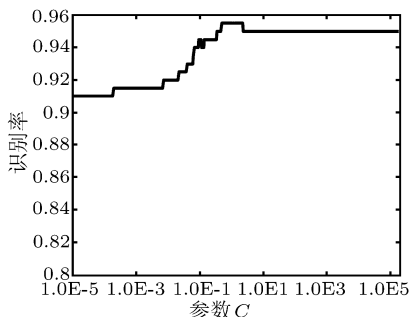


图 3 S_w 奇异时最大散度差分类器的识别率随参数 C 变化的函数曲线

Fig. 3 The recognition rate of MSD classifier vs. the parameter C when S_w is singular

4 自适应最大散度差分类算法

基本思想是迭代求解参数 C , 使得函数值 $F(C)$ 的绝对值小于某个给定的误差 ε (此时, 类内散布矩阵非奇异, 最大散度差准则蜕化为 Fisher 鉴别准则), 或者使得函数值 $F(C)$ 的改变量 $\Delta F(C)$ 的绝对值小于该给定的误差 ε (此时, 类内散布矩阵奇异, 最大散度差准则蜕化为大间距线性投影准则) (参见图 1).

下面给出自适应最大散度差 (Adapting maximum scatter difference, 简记为 AMSD) 分类算法.

- 第 1 步. 置 $C \leftarrow C_0$, $\varepsilon \leftarrow \varepsilon_0$, 并计算矩阵 $S_b - C \cdot S_w$ 的最大特征值, 即 $F(C)$.
- 第 2 步. 若 $F(C) \geq 0$, 置 $C \leftarrow kC$, 并计算 $F(C)$ 和 $\Delta F(C)$.
- 第 3 步. 若 $|\Delta F(C)| \leq \varepsilon$, 则转第 7 步.
- 第 4 步. 置 $C \leftarrow kC$, 并计算 $F(C)$.
- 第 5 步. 若 $F(C) < 0$, 置 $flag \leftarrow -1$, 并转第 8 步.
- 第 6 步. 计算 $\Delta F(C)$, 并转第 3 步.
- 第 7 步. 输出矩阵 $S_b - C \cdot S_w$ 最大特征值对应的特征向量, 算法结束.
- 第 8 步. 置 $a \leftarrow 0$, $b \leftarrow C$, $C \leftarrow (a + b)/2$, 并计算 $F(C)$.
- 第 9 步. 若 $|F(C)| \leq \varepsilon$, 则转第 12 步.
- 第 10 步. 若 $F(C) > 0$, 则置 $a \leftarrow C$, 否则置 $b \leftarrow C$.
- 第 11 步. 置 $C \leftarrow (a + b)/2$, 计算 $F(C)$, 并转第 9 步.
- 第 12 步. 输出矩阵 $S_b - C \cdot S_w$ 最大特征值对应的特征向量, 算法结束.

参数 C 的初值 C_0 可以根据需要灵活设定. 例如, 当应用 AMSD 求解人脸识别问题时, 参数 C 的初值可以取得大一些. 我们的仿真实验结果表明, 对于非奇异的类内散布矩阵 S_w , 取误差界 $\varepsilon_0 = 0.01$; 对于奇异的类内散布矩阵 S_w , 取误差界 $\varepsilon_0 = 1$, 即可取得良好的分类效果. 参数 k 的作用是, 在类内散布矩阵为奇异矩阵的时 (如进行人脸识别时), 使最大散度差准则逼近大间距线性投影准则. k 值越大, 逼近的速度就越快. k 的取值不宜过大或过小: k 值过大, 会增加类内散布矩阵非奇异时, 最大散度差准则逼近 Fisher 鉴别准则所需的时间; 反之, k 值过小, 又会增加类内散布矩阵奇异时, 最大散度差准则逼近大间距线性投影准则所需时间. 在本文的实验中, 参数 k 均取值为 10.

5 仿真实验

实验一

数据集为 AR 人脸图像数据库. 它由 126 人的 4000 多幅彩色图像组成, 其中 120 人每人有 26 张在不同时期、光照、姿势、表情、遮挡等条件下拍摄的人脸图像. 这 26 张图像摄于两个时期, 前 13 张图像为一个时期, 后 13 张图像为另一个时期, 前后相差 14 天. 本实验仅考察无遮挡的人脸图像, 即每个人的第 1-7 张和第 14-20 张图像. 取每个人的第 1-7 张图像作为训练样本, 第 14-20 张图像作为测试样本, 训练样本和测试样本总数均为 840 个. 为消除人脸位置、光照等因素的影响, 我们采用了分辨率为 50×40 的规范化人脸图像, 并进行了灰度均衡化处理.

本实验对自适应最大散度差分类器和大间距线性投影分类器 (Large margin linear projection, 简记为 LMLP)^[9]、支持向量机的分类效果进行比较.

自适应最大散度差分类器、大间距线性投影分类器与支持向量机本质上都是两类分类器, 用于多类分类问题时需要将原问题分解为若干个两类分类问题. 常见的分解方式有以下三种: 一对一 (one-vs-one)、一对多 (one-vs-rest) 和有向树 (Directed acyclic graph, 其缩写为 DAG)^[10]. 实验中我们仅采用一对一分解方式.

自适应最大散度差分类器可以用 Matlab 中求稀疏矩阵的特征值及特征向量的函数 eig 等来实现, 大间距线性投影分类器可以用 Matlab 中的二次规划函数 quadprog 等来实现, 而支持向量机则采用 Junshui Ma^[11] 等开发的基于 Matlab 的 SVM 工具箱中的 LinearSVC.

表 1 列出了大间距线性投影、线性支持向量机和自适应最大散度差分类器在 AR 数据集上的正确识别率.

表 1 各个分类器在 AR 人脸图像数据集上的正确识别率 (%)
Table 1 Recognition rates of various classifiers on AR face image database (%)

分类器	大间距线性投影	线性支持向量机	自适应最大散度差
正确识别率	66.19	66.19	66.31

实验二

数据集来自 UCI 机器学习数据库, 它们分别是 Heart, Australian, Cancer, German, Liver 和 Pima. 本文在这六个数据集上的实验均采用留一法进行精度的测试.

表 2 列出了线性支持向量机、Fisher 分类器和自适应最大散度差分类器在 Heart, Australian, Cancer, German, Liver 和 Pima 等 6 个数据集上的正确识别率.

表 2 各个分类器在 6 个 UCI 数据集上的正确识别率 (%)
Table 2 Recognition rates of various classifiers on 6 UCI datasets (%)

数据集名称	Australian	Cancer	German	Heart	Liver	Pima
数据集维数	690×14	699×9	1000×24	270×13	345×6	768×8
LSVM	80.72	96.57	76.10	84.81	69.57	65.10
正确识别率 Fisher	85.94	97.42	70.30	83.70	62.90	74.22
AMSD	85.94	97.42	70.30	83.70	62.90	74.09

6 结束语

Fisher 鉴别准则和大间距线性投影准则均不带参数, 分别适应于类内散布矩阵非奇异和奇异两种特殊场合. 文献 1 已经证明: 当类内散布矩阵为奇异矩阵时, 大间距线性投影准则为最大散度差鉴别准则的渐进形式. 在本文中, 我们进一步揭示了最大散度差鉴别准

则与 Fisher 鉴别准则的内在联系, 并提出了能综合 Fisher 鉴别准则和大间距线性投影准则优势的基于最大散度差鉴别准则的自适应分类算法, 较好地克服了它们各自存在的不足.

需要指出的是, 自适应最大散度差分类算法虽然将大间距线性投影分类算法和 Fisher 分类算法有机统一起来, 因而同时适用于类内散布矩阵为奇异矩阵和非奇异矩阵等多种场合, 但是由于需要通过迭代来确定合适的参数 C , 其分类效率要低于大间距线性投影分类算法和 Fisher 分类算法. 因此, 有必要找到其快速实现方法. 这是需要我们今后加以解决的一个主要问题.

References

- 1 Song F X, Cheng K, Yang J Y, Liu S H. Maximum scatter difference, large margin linear projection and support vector machines. *Acta Automatica Sinica*, 2004, **30**(6): 890~896
- 2 Duda R O, Hart P E, Stork D G. *Pattern Classification*. New York: John Wiley & Sons, 2001
- 3 Turk M, Pentland A. Human face recognition. *Journal of Cognitive Neuroscience*, 1991, **3**(1): 71~86
- 4 Belhumeur P N, et al. Eigenfaces vs Fisherfaces: Recognition using class specific linear projections. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 711~720
- 5 Vapnik V. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995
- 6 <http://www.ics.uci.edu/mlearn/MLRepository.html>
- 7 Wei Q L, Wang R S, Xu B, Wang J Y, Bai W L. *Mathematical Programming and Optimization Design*. Beijing: National Defense Industry Press, 1984
- 8 Olivetti Research Laboratory face database:<http://www.uk.research.att.com/facedatabase.html>
- 9 Song F X, Yang J Y, Liu S H. Large margin linear projection and face recognition. *Pattern Recognition*, 2004, **37**(9): 1953~1955
- 10 Hsu C, Lin C. A Comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 2002, **13**(2): 415~425
- 11 Ma J S, Zhao Y, Ahalt S. "OSU SVM Classifier Matlab Toolbox (ver 3.00)", http://eewww.eng.ohio-state.edu/maj/osu_svm/

宋枫溪 哈尔滨工业大学深圳研究院教授, 博士后. 研究方向为模式识别理论与应用.

(**SONG Feng-Xi** Professor in Artillery Academy and post doctor fellow in Shenzhen Graduate School at Harbin Institute of Technology. His research interests include pattern recognition theory and its application.)

张大鹏 香港理工大学教授, 博导. 研究方向为生物特征识别.

(**ZHANG David** Professor at Hong Kong Polytechnic University. His research interest includes biometric recognition.)

杨静宇 南京理工大学教授, 博导. 研究方向为模式识别与智能系统.

(**YANG Jing-Yu** Professor at Nanjing University of Science and Technology. His research interests include pattern recognition and intelligent system.)

高秀梅 淮阴师范学院副教授, 博士. 研究方向为模式识别与机器学习.

(**GAO Xiu-Mei** Ph.D. and associate professor at Huaiyin Normal College. Her research interests include pattern recognition and machine learning.)