# 协作式整体和局部的分类机

张战成 [1] 王士同 [1, 2] 钟富礼 [2]

**摘 要** 提出了一种协作式整体局部分类算法, 即 C$^2$M (Collaborative classification machine with local and global information), 该算法利用两类样本各自的协方差作为整体方向信息, 获得两个带整体和局部信息的分类面, 并通过组合分类器的平均规则将两个分类面组合, 得到最终的最优判决平面. 该算法可用两次 QP (Quadratic programming) 求解, 时间复杂度为 O($2N^3$), 大大小于 M$^4$ (Maxi-min margin machine) 的 O($N^4$), 线性核时的分类精度高于只利用了局部信息的支持向量机 (Support vector machine, SVM). 理论上证明了在交遇区较多时, C$^2$M 可以比 M$^4$ 更有效地利用全局信息, 并提出了判断整体信息对分类是否有贡献的 4 个判别指标. 模拟数据和标准数据集上与 M$^4$ 和 SVM 的对比实验证明了该算法的有效性.

**关键词** 分类, 整体局部学习, 协作学习, 支持向量机, 最大边界

**DOI** 10.3724/SP.J.1004.2011.01256

## A Collaborative Classifier for Local and Global Learning

ZHANG Zhan-Cheng[1] WANG Shi-Tong[1, 2] CHUNG Fu-Lai[2]

**Abstract** Inspired by covariance matrix stating data direction globally, we construct a novel large margin classifier called collaborative classification machine with local and global information (C$^2$M). By the median rule of combining classifiers, this model collaboratively learns the decision boundary from two hyperplanes with global information. The proposed C$^2$M algorithm can be individually solved as a quadratic programming (QP) problem, and has O($2N^3$) time complexity that is faster than O($N^4$) of existing maxi-min margin machine (M$^4$). We describe the C$^2$M model definition, provide the geometrical interpretation, and present theoretical justifications. As a major contribution, we show that C$^2$M can robustly utilize the global information when M$^4$ loses the global information on those data sets with confused classes margin. We also exploit kernelization trick and extend C$^2$M to nonlinear classification. Moreover, we show that C$^2$M can be transformed into standard support vector machine (SVM) model and can be solved by other quick algorithms widely used by SVM. Furthermore, we propose four indicators to evaluate the global impact of covariance matrix on classification. Experiments on toy and real-world data sets demonstrate that the C$^2$M has comparable performance with SVM that utilizes only local information, while the C$^2$M is more robust and time saving than M$^4$.

**Key words** Classification, learning locally and globally, collaborative learning, support vector machine (SVM), large-margin

借鉴人脑学习的方式, 模式识别的研究也有局部和整体的学习算法. 线性降维方法 (Linear discriminant analysis, LDA)[1] 使用类内散度矩阵 $S_w$ 和类间散度矩阵 $S_b$ 表示模式的整体信息. Lanckriet 等[2] 提出的最小最大概率机 (Minimax probability machine, MPM), 使用样本均值和协方差作为分类的整体信息. LDA 和 MPM 的学习过程仅仅利用了样本的整体信息而忽略了样本的局部信息, 而支持向量机 (Support vector machine, SVM) 仅仅利用了样本局部信息而忽略了样本的整体信息.

针对以上学习算法在局部和整体信息处理上的不足, Huang 等[3] 提出了一种新的分类器: M$^4$ (Maxi-min margin machine). M$^4$ 的学习过程综合了样本的局部信息 (每个样本点自身的特征信息) 和整体信息 (每类样本的协方差矩阵), 表现出了较好的识别率, Huang 还分析了 M$^4$ 与 LDA、MPM 以及 SVM 的联系, 并进一步证明了 LDA、MPM 和 SVM 可以看成是 M$^4$ 的特例. M$^4$ 的求解过程是一维最优逼近优化问题, 每次逼近是二次锥优化问题 (Second order cone programming, SOCP), 对于 $N$ 个 $n$ 维样本的训练, 其算法时间复杂度为 O($Nn^3$), 核化后的时间复杂度为 O($N^4$), 相对于原始的用二次规划 (Quadratic programming, QP) 求解 SVM 的时间复杂度 O($N^3$)[4], M$^4$ 的时间复杂度增长 $N$ 倍, 在文献 [3] 中, 虽然 Huang 也讨论了从样本点中抽取部分数据训练的方法来减少 $N$ 的大小以减少 M$^4$ 的运行时间, 但是抽取的规则和抽取的终止条件的选取不仅增加了训练的参数, 而且造成样本局部

收稿日期 2011-01-17 录用日期 2011-04-09
Manuscript received January 17, 2011; accepted April 9, 2011
国家自然科学基金 (60903100, 61103128), 中央高校基本科研业务费专项资金 (JUDCF09034, JUSRP211A34) 资助
Supported by National Natural Science Foundation of China (60903100, 61103128) and Fundamental Research Funds for the Central Universities (JUDCF09034, JUSRP211A34)
1. 江南大学数字媒体学院 无锡 214122 2. 香港理工大学电子计算学系 香港
1. School of Digital Media, Jiangnan University, Wuxi 214122 2. Department of Computing, Hong Kong Polytechnic University, Hong Kong

信息的损失, 可能引入误差.

本文提出一种新的学习算法, 即整体和局部的协作式分类器 (Collaborative classification machine with local and global information, C$^2$M), 该算法采用分治协作的思想把 M$^4$ 中提出的整体信息引入到两个局部学习过程中, 每个包含整体信息的局部学习过程可以用 QP (Quadratic programming) 求解, 整个学习过程的算法复杂度为 O($2N^3$), 相对 M$^4$ 可以大大减少时间复杂度. 本文还进一步证明, 在线性不可分的情况下, 两类数据交遇区较多时, M$^4$ 无法有效利用整体信息, 而 C$^2$M 依然可以有效利用整体信息. 本文还分析 C$^2$M 和 SVM 的关系, 讨论了 C$^2$M 的快速解法和核化方法.

本文结构如下: 第 1 节详细介绍本文新的协作式算法 C$^2$M, 给出了该算法的数学定义和直观的几何解释, 并分析了该算法与 M$^4$、SVM 的关系; 第 2 节讨论了 C$^2$M 的核化方法; 第 3 节对比了 C$^2$M, M$^4$ 和 SVM 在模拟数据集和 UCI (University of California at Irvine) 标准数据集[5] 上的分类性能, 并提出了评价全局信息对分类贡献程度的 4 个指标; 第 4 节总结全文, 并提出了以后期待解决的问题.

## 1 整体和局部的协作式学习

首先, 假设样本是线性可分的, 在此前提下建立 C$^2$M 的问题模型, 并用 QP 求解最优分类面. 然后, 讨论 C$^2$M 与 M$^4$、SVM 的关系并进一步推广到线性不可分的情况.

### 1.1 线性可分的情况

借鉴 M$^4$ 的方法, 可以用样本的协方差矩阵表示样本分布方向的整体信息. 先建立样本在线性可分情况下的数学模型, 并给出这种情况下的几何解释.

问题定义: 设两类线性可分的样本 $X=\{\boldsymbol{x}_i\}_{i=1}^{N_x}$ 和 $Y=\{\boldsymbol{y}_j\}_{j=1}^{N_y}$ 分别代表正类和负类, $\boldsymbol{x}_i,\boldsymbol{y}_j\in\mathbf{R}^n$, 训练样本的总数为 $N=N_x+N_y$, 求一个分类面 $f(\boldsymbol{z})=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{z}+b$, $\boldsymbol{w},\boldsymbol{z}\in\mathbf{R}^n$, $b\in\mathbf{R}$, 使得未知样本 $\boldsymbol{z}$ 属于 $X$ 类, 当且仅当 $f(\boldsymbol{z})>0$, 否则 $\boldsymbol{z}$ 属于 $Y$ 类.

C$^2$M 的数学模型为: 考虑 $X$ 的整体信息有

$$\max_{\rho_x,\boldsymbol{w}_x\neq\mathbf{0},b_x}\quad \rho_x \tag{1}$$

$$\text{s.t.}\quad \frac{(\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{x}_i+b_x)}{\sqrt{\boldsymbol{w}_x^{\mathrm{T}}\Sigma_x\boldsymbol{w}_x}}\geq\rho_x,\quad i=1,2,\cdots,N_x \tag{2}$$

$$\frac{-(\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{y}_j+b_x)}{\sqrt{\boldsymbol{w}_x^{\mathrm{T}}\Sigma_x\boldsymbol{w}_x}}\geq\rho_x,\quad j=1,2,\cdots,N_y \tag{3}$$

考虑 $Y$ 的整体信息有

$$\max_{\rho_y,\boldsymbol{w}_y\neq\mathbf{0},b_y}\quad \rho_y \tag{4}$$

$$\text{s.t.}\quad \frac{(\boldsymbol{w}_y^{\mathrm{T}}\boldsymbol{x}_i+b_y)}{\sqrt{\boldsymbol{w}_y^{\mathrm{T}}\Sigma_y\boldsymbol{w}_y}}\geq\rho_y,\quad i=1,2,\cdots,N_x \tag{5}$$

$$\frac{-(\boldsymbol{w}_y^{\mathrm{T}}\boldsymbol{y}_j+b_y)}{\sqrt{\boldsymbol{w}_y^{\mathrm{T}}\Sigma_y\boldsymbol{w}_y}}\geq\rho_y,\quad j=1,2,\cdots,N_y \tag{6}$$

其中, $\rho_x,\rho_y$ 为两类样本之间的间隔, $\Sigma_x,\Sigma_y\in\mathbf{R}^{n\times n}$ 分别表示 $X$ 和 $Y$ 的协方差, 一般都是正定的. 注意到式 (1)$\sim$(3) 中, $\|\boldsymbol{w}_x\|$ 的大小并不影响最优解, 不失一般性, 使得 $\rho_x\sqrt{\boldsymbol{w}_x^{\mathrm{T}}\Sigma_x\boldsymbol{w}_x}=1$, 式 (1)$\sim$(3) 等价为

$$\min_{\boldsymbol{w}_x\neq\mathbf{0},b_x}\quad \boldsymbol{w}_x^{\mathrm{T}}\Sigma_x\boldsymbol{w}_x \tag{7}$$

$$\text{s.t.}\quad (\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{x}_i+b_x)\geq 1,\quad i=1,2,\cdots,N_x \tag{8}$$

$$-(\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{y}_j+b_x)\geq 1,\quad j=1,2,\cdots,N_y \tag{9}$$

式 (7)$\sim$(9) 是一个典型的 QP 问题, 同理, 式 (4)$\sim$(6) 也是一个 QP 问题. 求解两个独立的 QP 问题, 分别得到最优解 $\{\boldsymbol{w}_x^*,b_x^*\}$ 和 $\{\boldsymbol{w}_y^*,b_y^*\}$, 以及分别包含 $X$ 和 $Y$ 的整体的信息的两个分类面 $f_x(\boldsymbol{z})=\boldsymbol{w}_x^{*\mathrm{T}}\boldsymbol{z}+b_x^*$ (图 1 (a) 中灰色虚线) 和 $f_y(\boldsymbol{z})=\boldsymbol{w}_y^{*\mathrm{T}}\boldsymbol{z}+b_y^*$ (图 1 (b) 中灰色虚线).

获得两个独立的判别平面后, 如何协作地做出最终的判别实际是一个分类器组合[6] 的问题. 考虑距两平面的马氏距离相等的点集, 有

$$\begin{cases}\dfrac{|f_x(\boldsymbol{z})|}{\sqrt{\boldsymbol{w}_x^{*\mathrm{T}}\Sigma_x\boldsymbol{w}_x^*}}=\dfrac{|f_y(\boldsymbol{z})|}{\sqrt{\boldsymbol{w}_y^{*\mathrm{T}}\Sigma_y\boldsymbol{w}_y^*}}\\ \mathrm{sgn}(f_x(\boldsymbol{z})f_y(\boldsymbol{z}))<0\end{cases} \tag{10}$$

最终的分类面 $f(\boldsymbol{z})$ 应该是满足方程 (10) 的根集, 令 $s_x=\sqrt{\boldsymbol{w}_x^{*\mathrm{T}}\Sigma_x\boldsymbol{w}_x^*}$, $s_y=\sqrt{\boldsymbol{w}_y^{*\mathrm{T}}\Sigma_y\boldsymbol{w}_y^*}$, 最终的分类面为

$$f(\boldsymbol{z})=(s_y\boldsymbol{w}_x^*+s_x\boldsymbol{w}_y^*)^{\mathrm{T}}\boldsymbol{z}+(s_yb_x^*+s_xb_y^*) \tag{11}$$

式 (10) 的组合规则相当于分类器组合[6] 中的平均规则 (Median rule) 的一种变形, 原始规则为算术平均, 式 (10) 为马氏距离的平均. 式 (10) 的几何意义相当于图 1 (c) 中一系列共心的椭圆组 (灰色点划线的小椭圆) 的共心点的集合, 小椭圆组中, 代表 $X$ 方向信息的小椭圆与 $f_x(\boldsymbol{z})$ 相切, 代表 $Y$ 方向信息的小椭圆与 $f_y(\boldsymbol{z})$ 相切, 其共心点距离 $f_x(\boldsymbol{z}),f_y(\boldsymbol{z})$ 的马氏距离相等, 共心点的集合为 $f(\boldsymbol{z})$.

与 M$^4$ 相同, C$^2$M 通过最大化马氏距离间隔求得带有整体信息的分类面, 但是 C$^2$M 把整体信息分两次计算, 在保留整体信息的同时, 把一维搜索的多次 SOCP 求解化简为两次 QP, 从而把时间复杂度从 O($N^4$) 减低为 O($2N^3$), 并且, 在后面的第 1.3 节中可以看到, 基于 SVM 的各种快速算法[7] 同样适用 C$^2$M.

### 1.2 几何解释

图 1 中, $X$ 分布在左边以 $x_0$ 为中心的椭圆区域内, $Y$ 分布在右边以 $y_0$ 为中心的椭圆区域内, 椭圆的大小和旋转情况代表了各自的整体信息. 图 1 (a) 的判决平面 (灰色虚线) 反映了 $X$ 的整体信息; 图 1 (b) 的判决平面 (黑色虚线) 反映了 $Y$ 的整体信息, 计算距离这两个判决平面马氏距离相等点集得到图 1 (c) 的判决平面 (灰色实线). C$^2$M 与图 1 (d) 中的 M$^4$ 相当, 通过协作的方式综合考虑两类的整体信息, 而 SVM 仅仅考虑了最大化支持向量 (图 1 (d) 中灰色实心圆点) 间的平面间隔, 忽略了整体的方向信息.

### 1.3 C$^2$M 和 SVM 的关系

式 (7) ∼ (9) 中, 令 $\Sigma_x = U^{\mathrm{T}}U$, $\hat{\boldsymbol{w}} = U\boldsymbol{w}$, $\hat{\boldsymbol{x}}_i = (U^{\mathrm{T}})^{-1}\boldsymbol{x}_i$, $i = 1, 2, \cdots, N$, 则原模型可以表示为

$$\min_{\hat{\boldsymbol{w}}\neq\boldsymbol{0},\hat{b}} \quad \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{w}} \tag{12}$$

$$\text{s.t.} \quad (\hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_i + \hat{b}) \geq 1, \quad i = 1, 2, \cdots, N_x \tag{13}$$

$$-(\hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_j + \hat{b}) \geq 1, \quad j = 1, 2, \cdots, N_y \tag{14}$$

式 (12) ∼ (14) 相当于把原来的样本 $X$ 和 $Y$ 作了一个以 $(U^{\mathrm{T}})^{-1}$ 为变换矩阵的线性变换, 变换后的形式是标准的 SVM 形式, 可以用 SVM 的快速算法[7] 求解.

和 M$^4$ 相同, 当 $\Sigma_x = \Sigma_y = I$ 时, $X$ 和 $Y$ 的整体信息相同, 即式 (1) ∼ (3) 和式 (4) ∼ (6) 相同, 只需要求解一个就可以了, 此时 C$^2$M 退化为 SVM 模型.

### 1.4 线性不可分的情况

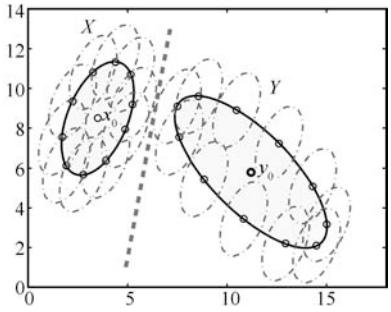
两类之间线形不可分的情况下, 通过引入非负松弛变量表示和理想可分条件下的偏差. 原来线性可分的形式可以改写为如下形式:

考虑 $X$ 的整体信息

$$\min_{\boldsymbol{w}_x\neq\boldsymbol{0},b_x,\xi} \boldsymbol{w}_x^{\mathrm{T}}\Sigma_x\boldsymbol{w}_x + C_x\sum_{k=1}^{N}\xi_k \tag{15}$$

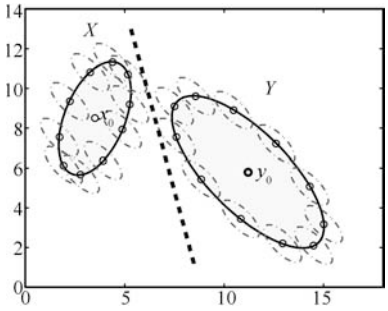
$$\text{s.t.} \quad (\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{x}_i + b_x) \geq 1 - \xi_i, \qquad i = 1, 2, \cdots, N_x \tag{16}$$

$$-(\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{y}_j + b_x) \geq 1 - \xi_{j+N_x}, \quad j = 1, 2, \cdots, N_y \tag{17}$$



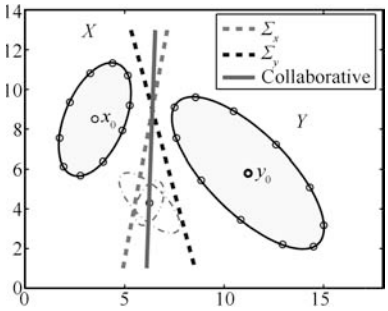

(a) $X$ 的整体信息

(a) Globally considering $X$



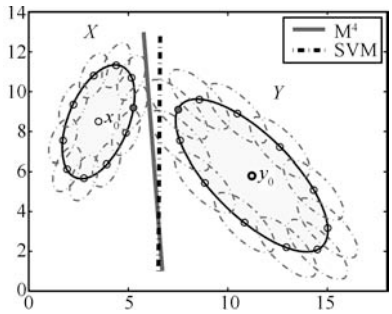

(b) $Y$ 的整体信息

(b) Globally considering $Y$




(c) $X$ 和 $Y$ 的整体信息协作

(c) Collaboratively considering $X$ and $Y$




(d) M$^4$ 和 SVM

(d) M$^4$ and SVM

图 1 几何解释

Fig. 1 Geometric interpretation

考虑 $Y$ 的整体信息

$$\min_{\boldsymbol{w}_y\neq\boldsymbol{0},b_y,\varepsilon} \boldsymbol{w}_y^{\mathrm{T}}\Sigma_y\boldsymbol{w}_y + C_y\sum_{k=1}^{N}\varepsilon_k \tag{18}$$

$$\text{s.t.}\,(\boldsymbol{w}_y^{\mathrm{T}}\boldsymbol{x}_i + b_y) \geq 1-\varepsilon_i, \qquad i=1,\cdots,N_x \tag{19}$$

$$-(\boldsymbol{w}_y^{\mathrm{T}}\boldsymbol{y}_j + b_y) \geq 1-\varepsilon_{j+N_y}, \quad j=1,\cdots,N_y \tag{20}$$

其中, $\xi_k\geq 0$, $\varepsilon_k\geq 0$, $k=1,2,\cdots,N$, $C_x, C_y$ 是可调的正的常数, 控制对错分样本的惩罚程度. 式 (15) ∼ (17) 和式 (18) ∼ (20) 的求解方式与线性可分的情况一样, 都可以用 QP 求解, 分类面的构造方式相同.

### 1.5 交遇区较多的情况

Huang 等已经证明[3], 在线性可分的情况下, 当 $\Sigma_x=\Sigma_y=I$ ($I$ 为 $n\times n$ 的单位阵) 时, M$^4$ 等价于标准的 SVM. 本节进一步讨论线性不可分的情况下 M$^4$ 和其他模型的关系, 并证明在两类数据交遇区较多时, M$^4$ 将丢失全局信息, 而 C$^2$M 依然可以利用全局信息.

为方便讨论, 先列出 M$^4$ 在线性不可分情况下的模型[3]:

$$\max_{\rho,\boldsymbol{w}\neq\boldsymbol{0},b,\xi} \quad \rho - C\sum_{k=1}^{N}\xi_k \tag{21}$$

$$\text{s.t.} \quad (\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\geq\rho\sqrt{\boldsymbol{w}^{\mathrm{T}}\Sigma_x\boldsymbol{w}}-\xi_i,\ i=1,\cdots,N_x \tag{22}$$

$$-(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{y}_j+b)\geq\rho\sqrt{\boldsymbol{w}^{\mathrm{T}}\Sigma_x\boldsymbol{w}}-\xi_{j+N_x},\ j=1,\cdots,N_y \tag{23}$$

两类数据交遇区越多, 边界越模糊, 其分类间隔 $\rho$ 越趋于 0, 而现实数据中, 这种交迭的情况比较多, 第 3 节中基于 UCI 的标准数据集的实验结果表明, 一些情况下, 其最优值 $\rho^*=0$. 若 $\rho=0$, 则式 (21) ∼ (23) 变为

$$\min_{\rho,\boldsymbol{w}\neq\boldsymbol{0},b,\xi} \quad C\sum_{k=1}^{N}\xi_k \tag{24}$$

$$\text{s.t.} \quad (\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\geq-\xi_i, \qquad i=1,2,\cdots,N_x \tag{25}$$

$$-(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{y}_j+b)\geq-\xi_{j+N_x}, \quad j=1,2,\cdots,N_y \tag{26}$$

式 (24) ∼ (26) 是感知器准则的线性分类器, 此时 M$^4$ 的全局信息失效, 而 C$^2$M 模型中, 式 (15) 和式 (18) 中的全局信息 $\Sigma_x$ 和 $\Sigma_y$ 依然可以发挥作用, 第 3 节中基于模拟数据的实验图 2 (b) 可以很明显地看到这点.

## 2 核化

为提高分类器的非线性识别能力, 利用核技巧[8], 通过映射 $\varphi:\mathbf{R}^n\mapsto\mathbf{R}^f$, 将原问题空间 $\mathbf{R}^n$ 映射到一个高维的特征空间 $\mathbf{R}^f$, 有 $\boldsymbol{x}_i\mapsto\varphi(\boldsymbol{x}_i)$, $\boldsymbol{y}_i\mapsto\varphi(\boldsymbol{y}_i)$, $i=1,\cdots,N_x$, $j=1,\cdots,N_y$, 在高维特征空间的线性分类面 $f_k(\boldsymbol{z})=\boldsymbol{\alpha}^{\mathrm{T}}\varphi(\boldsymbol{z})+b_k$, 即为原有问题空间的的非线性分类面, 其中, $\boldsymbol{\alpha},\varphi(\boldsymbol{z})\in\mathbf{R}^f$, $\boldsymbol{z}\in\mathbf{R}^n$, $b\in\mathbf{R}$.

考虑 $\varphi(\boldsymbol{x}_i)$ 的整体信息和原问题空间的模型, 式 (7) ∼ (9) 可写为

$$\min_{\boldsymbol{\alpha}\neq\boldsymbol{0},b_x} \boldsymbol{\alpha}^{\mathrm{T}}\Sigma_{\varphi(x)}\boldsymbol{\alpha} \tag{27}$$

$$\text{s.t.} \quad (\boldsymbol{\alpha}^{\mathrm{T}}\varphi(\boldsymbol{x}_i)+b_x)\geq 1, \quad i=1,2,\cdots,N_x \tag{28}$$

$$-(\boldsymbol{\alpha}^{\mathrm{T}}\varphi(\boldsymbol{y}_j)+b_x)\geq 1, \quad j=1,2,\cdots,N_y \tag{29}$$

以及包含 $Y$ 的整体信息的模型:

$$\min_{\boldsymbol{\beta}\neq\boldsymbol{0},b_y} \boldsymbol{\beta}^{\mathrm{T}}\Sigma_{\varphi(y)}\boldsymbol{\beta} \tag{30}$$

$$\text{s.t.} \quad (\boldsymbol{\beta}^{\mathrm{T}}\varphi(\boldsymbol{x}_i)+b_y)\geq 1, \quad i=1,2,\cdots,N_x \tag{31}$$

$$-(\boldsymbol{\beta}^{\mathrm{T}}\varphi(\boldsymbol{y}_j)+b_y)\geq 1, \quad j=1,2,\cdots,N_y \tag{32}$$

为方便地使用内积形式表示原空间的非线性分类面, 需要把式 (27) ∼ (29) 和式 (30) ∼ (32) 表达为 $K(\boldsymbol{x}_i,\boldsymbol{x}_j)=\varphi(\boldsymbol{x}_i)^{\mathrm{T}}\varphi(\boldsymbol{x}_i)$ 的内积形式. 下面讨论在特征空间 $\mathbf{R}^f$ 的求解方式.

**定理 1.** 设特征空间中的两类训练样本 $\{\varphi(\boldsymbol{x}_i)\}_{i=1}^{N_x}$, $\{\varphi(\boldsymbol{y}_j)\}_{j=1}^{N_y}$, 设定如下估计量:

$$\overline{\varphi(\boldsymbol{x})}=\sum_{i=1}^{N_x}\lambda_i\,\varphi(\boldsymbol{x}_i)$$

$$\overline{\varphi(\boldsymbol{y})}=\sum_{j=1}^{N_y}\theta_j\,\varphi(\boldsymbol{y}_j)$$

$$\Sigma_{\varphi(x)}=\tau_x I_f+\sum_{i=1}^{N_x}\Lambda_i\,(\varphi(\boldsymbol{x}_i)-\overline{\varphi(\boldsymbol{x})})(\varphi(\boldsymbol{x}_i)-\overline{\varphi(\boldsymbol{x})})^{\mathrm{T}}$$

$$\Sigma_{\varphi(y)}=\tau_y I_f+\sum_{j=1}^{N_y}\Theta_j\,(\varphi(\boldsymbol{y}_j)-\overline{\varphi(\boldsymbol{y})})(\varphi(\boldsymbol{y}_j)-\overline{\varphi(\boldsymbol{y})})^{\mathrm{T}}$$

其中, $I_f$ 是 $f\times f$ 的单位矩阵, $\lambda_i$, $\theta_j$, $\Lambda_i$, $\Theta_j$ 是样本点 $\{\varphi(\boldsymbol{x}_i)\}_{i=1}^{N_x}$, $\{\varphi(\boldsymbol{y}_j)\}_{j=1}^{N_y}$ 对应的权重系数, $\tau_x$, $\tau_y$ 可以看成是协方差矩阵

的误差调整项. 核化模型式 (27)∼(29) 和式 (30)∼(32) 的解分别为 $\{\boldsymbol{\alpha}^*, b_x^*\}, \{\boldsymbol{\beta}^*, b_y^*\}$, 则 $\boldsymbol{\alpha}^*, \boldsymbol{\beta}^* \in \mathrm{span}(\{\varphi(\boldsymbol{x}_i)\}_{i=1}^{N_x}, \{\varphi(\boldsymbol{y}_j)\}_{j=1}^{N_y})$.

**证明.** 设 $S = \mathrm{span}(\{\varphi(\boldsymbol{x}_i)\}_{i=1}^{N_x}, \{\varphi(\boldsymbol{y}_j)\}_{j=1}^{N_y})$, $S \subset \mathbf{R}^f$, 显然 $\mathbf{R}^f$ 是 Hilbert 空间. 根据正交分解定理, $\mathbf{R}^f = S \oplus S^{\perp}$, $\forall a \in \mathbf{R}^f$ 可唯一地表示为: $\boldsymbol{\alpha} = \boldsymbol{\alpha}_d + \boldsymbol{\alpha}_p$, 其中, $\boldsymbol{\alpha}_d \in S$, $\boldsymbol{\alpha}_p \in S^{\perp}$. 有 $\boldsymbol{\alpha}_p^{\mathrm{T}}\varphi(\boldsymbol{x}_i) = 0$, $\boldsymbol{\alpha}_p^{\mathrm{T}}\varphi(\boldsymbol{y}_j) = 0$, 即正交分量 $\boldsymbol{\alpha}_p$ 对式 (28) 和式 (29) 没有影响. 式 (27) 的目标函数可以写为

$$\boldsymbol{\alpha}^{\mathrm{T}}\Sigma_{\varphi(x)}\boldsymbol{\alpha} = \boldsymbol{\alpha}_d^{\mathrm{T}}\left(\sum_{i=1}^{N_x}\Lambda_i\left(\varphi(\boldsymbol{x}_i)-\overline{\varphi(\boldsymbol{x})}\right)\left(\varphi(\boldsymbol{x}_i)-\overline{\varphi(\boldsymbol{x})}\right)^{\mathrm{T}}\right)\boldsymbol{\alpha}_d + \tau_x\left(\boldsymbol{\alpha}_d^{\mathrm{T}}\boldsymbol{\alpha}_d + \boldsymbol{\alpha}_p^{\mathrm{T}}\boldsymbol{\alpha}_p\right) \tag{33}$$

目标函数是求式 (33) 的最小值, 所以取得最优解 $\boldsymbol{\alpha}^*$ 时, $\boldsymbol{\alpha}_p^* = 0$, 有 $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}_d^*$, 即 $\boldsymbol{\alpha}^* \in \mathrm{span}(\{\varphi(\boldsymbol{x}_i)\}_{i=1}^{N_x}, \{\varphi(\boldsymbol{y}_j)\}_{j=1}^{N_y})$, 同理, $\boldsymbol{\beta}^* \in \mathrm{span}(\{\varphi(\boldsymbol{x}_i)\}_{i=1}^{N_x}, \{\varphi(\boldsymbol{y}_j)\}_{j=1}^{N_y})$.

由定理 1, $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 分别可以表示为训练样本点的线性组合形式:

$$\boldsymbol{\alpha} = \sum_{i=1}^{N_x}\mu_i\,\varphi(\boldsymbol{x}_i) + \sum_{j=1}^{N_y}\nu_j\,\varphi(\boldsymbol{y}_j) \tag{34}$$

$$\boldsymbol{\beta} = \sum_{i=1}^{N_x}\omega_i\,\varphi(\boldsymbol{x}_i) + \sum_{j=1}^{N_y}\upsilon_j\,\varphi(\boldsymbol{y}_j) \tag{35}$$

其中, $\mu_i, \nu_j, \omega_i, \upsilon_j \in \mathbf{R}$. 将这些系数定义为向量形式有:

$$\boldsymbol{\eta}_x = (\mu_1, \mu_2, \cdots, \mu_{N_x}, \nu_1, \nu_2, \cdots, \nu_{N_y})^{\mathrm{T}} \tag{36}$$

$$\boldsymbol{\eta}_y = (\omega_1, \omega_2, \cdots, \omega_{N_x}, \upsilon_1, \upsilon_2, \cdots, \upsilon_{N_y})^{\mathrm{T}} \tag{37}$$

为方便讨论约定如下标记: 全体训练样本 $\{\boldsymbol{z}_i\}_{i=1}^{N}$, 其中:

$$\boldsymbol{z}_i = \boldsymbol{x}_i, \quad i = 1, 2, \cdots, N_x$$
$$\boldsymbol{z}_{j+N_x} = \boldsymbol{y}_j, \quad j = 1, 2, \cdots, N_y$$

Gram 矩阵是 $N \times N$ 的对称矩阵, 每个元素 $K_{i,j} = \varphi(\boldsymbol{z}_i)^{\mathrm{T}}\varphi(\boldsymbol{z}_j)$, $\boldsymbol{K}_i \in \mathbf{R}^N$, $i, j = 1, 2, \cdots, N$, 是 Gram 矩阵 $K$ 的第 $i$ 行的向量, $K$ 的前面 $N_x$ 行定义为矩阵 $K_x$, 后面 $N_y$ 行定义为 $K_y$, 即

$$K = [K_{i,j}] = \begin{bmatrix} K_x \\ K_y \end{bmatrix}$$

参照协方差的点估计公式:

$$\hat{\Sigma}_{\varphi(x)} = \frac{1}{N_x}\sum_{i=1}^{N_x}\left(\varphi(\boldsymbol{x}_i) - \overline{\varphi(\boldsymbol{x})}\right)\left(\varphi(\boldsymbol{x}_i) - \overline{\varphi(\boldsymbol{x})}\right)^{\mathrm{T}}$$

令

$$\boldsymbol{M} = \begin{pmatrix} \sqrt{N_x}\boldsymbol{M}_x \\ \sqrt{N_y}\boldsymbol{M}_y \end{pmatrix} = \begin{pmatrix} K_x - \mathbf{1}_{N_x}\boldsymbol{m}_x^{\mathrm{T}} \\ K_y - \mathbf{1}_{N_y}\boldsymbol{m}_y^{\mathrm{T}} \end{pmatrix}$$

其中, $\boldsymbol{m}_x, \boldsymbol{m}_y \in \mathbf{R}^N$, 其每个元素定义为

$$(\boldsymbol{m}_x)_i = \frac{1}{N_x}\sum_{i=1}^{N_x}\boldsymbol{K}_i, \qquad i = 1, 2, \cdots, N_x$$

$$(\boldsymbol{m}_y)_j = \frac{1}{N_y}\sum_{i=N_x+1}^{N}\boldsymbol{K}_i, \quad j = 1, 2, \cdots, N_y$$

单位向量 $\mathbf{1}_{N_x} \in \mathbf{R}^{N_x}$, $\mathbf{1}_{N_y} \in \mathbf{R}^{N_y}$ 的所有元素均为 1. 协方差的点估计公式可以表示为

$$\hat{\Sigma}_{\varphi(x)} = \boldsymbol{M}_x{}^{\mathrm{T}}\boldsymbol{M}_x \tag{38}$$

$$\hat{\Sigma}_{\varphi(y)} = \boldsymbol{M}_y{}^{\mathrm{T}}\boldsymbol{M}_y \tag{39}$$

因为 $\tau_x$ 对式 (33) 的解没有影响, 所以可以令 $\tau_x = 0$, 同理也可令 $\tau_y = 0$, 则可以使用协方差的点估计式 (38) 和式 (39) 代替式 (27) 和式 (30) 的协方差.

最终的核化模型可以表达为内积形式:

$$\min_{\boldsymbol{\eta}_x \neq \mathbf{0}, b_x} \quad \boldsymbol{\eta}_x{}^{\mathrm{T}}\boldsymbol{M}_x{}^{\mathrm{T}}\boldsymbol{M}_x\boldsymbol{\eta}_x \tag{40}$$

$$\text{s.t.} \quad (\boldsymbol{\eta}_x{}^{\mathrm{T}}\boldsymbol{K}_i + b_x) \geq 1, \quad i = 1, 2, \cdots, N_x \tag{41}$$

$$-(\boldsymbol{\eta}_x{}^{\mathrm{T}}\boldsymbol{K}_{j+N_x} + b_x) \geq 1, \quad j = 1, 2, \cdots, N_y \tag{42}$$

同理考虑 $Y$ 的全局信息有:

$$\min_{\boldsymbol{\eta}_y \neq \mathbf{0}, b_y} \quad \boldsymbol{\eta}_y{}^{\mathrm{T}}\boldsymbol{M}_y{}^{\mathrm{T}}\boldsymbol{M}_y\boldsymbol{\eta}_y \tag{43}$$

$$\text{s.t.} \quad (\boldsymbol{\eta}_y{}^{\mathrm{T}}\boldsymbol{K}_i + b_y) \geq 1, \quad i = 1, 2, \cdots, N_x \tag{44}$$

$$-(\boldsymbol{\eta}_y{}^{\mathrm{T}}\boldsymbol{K}_{j+N_x} + b_y) \geq 1, \quad j = 1, 2, \cdots, N_y \tag{45}$$

式 (40)∼(42) 和式 (43)∼(45) 的问题同样可以用 QP 求解.

原来 $\mathbf{R}^n$ 空间的分类判别平面式 (11) 在 $\mathbf{R}^f$ 高维特征空间可以表达为

$$f_k(\boldsymbol{z}) = (s_y\boldsymbol{\eta}_x^* + s_x\boldsymbol{\eta}_y^*)^{\mathrm{T}}\boldsymbol{K}_z + (s_y b_x^* + s_x b_y^*) \tag{46}$$

其中, $s_x = \sqrt{\boldsymbol{\eta}_x^{*\mathrm{T}}\hat{\Sigma}_{\varphi(x)}\boldsymbol{\eta}_x^*}$, $s_y = \sqrt{\boldsymbol{\eta}_y^{*\mathrm{T}}\hat{\Sigma}_{\varphi(y)}\boldsymbol{\eta}_y^*}$, $\boldsymbol{K}_z \in \mathbf{R}^N$, $(\boldsymbol{K}_z)_i = K(\boldsymbol{z}_i, \boldsymbol{z})$, $i = 1, 2, \cdots, N$. □

## 3　实验结果分析

本节通过模拟数据集和标准数据集上的实验进一步分析验证 C$^2$M、M$^4$ 和 SVM 之间的关系, 对比它们的测试精度和算法效率.

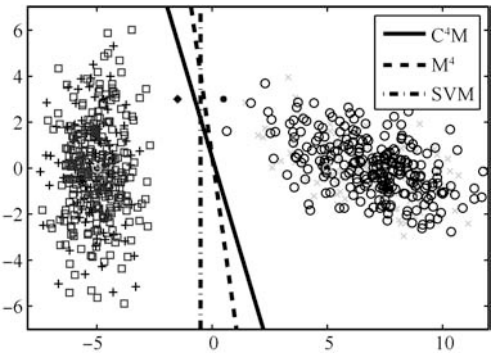
### 3.1　模拟数据集实验分析

为了说明全局信息所起的作用, 我们构造如图 2 (a) 所示的两类正态分布的数据, 左边的数据是 $X$ 类, 其中心为 $[-5, 0]$, 协方差为 $[1, 0; 0, 5]$, 右边的数据为 $Y$ 类, 由以 $[7, 0]$ 为中心, $[1, 0; 0, 5]$ 为协方差的数据逆时针旋转 $3/8\pi$ 组成. 每类数据随机划分为 120 个训练数据, 250 个测试数据, "+" 和 "□" 分别表 $X$ 的训练数据和测试数据, "×" 和 "○" 分别表示 $Y$ 的训练数据和测试数据, "•" 表示 SVM 训练得到的支持向量.

在两类测试数据中分别加入 $[-1.5, 3]$ 和 $[0.5, 3]$ 两个噪音点, 因为两个噪音点在边界上, 每次试验中, SVM 都将这两个噪音点训练为支持向量. 从图 2 (a) 中可以看出, C$^2$M 和 M$^4$ 都可以很好地反映 $Y$ 的旋转情况, 而 SVM 被噪音干扰, 仅仅依靠两个支持向量 (实际上是噪音点) 来确定分类面, C$^2$M 和 M$^4$ 的分类线更能准确地反映数据的旋转情况.

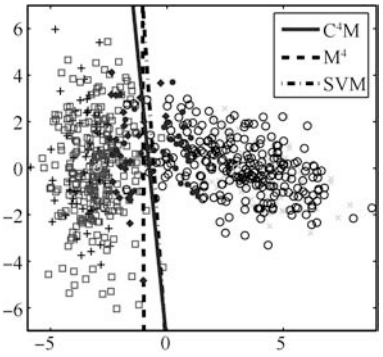
为构造交遇区较多的情况, 将图 2 (a) 中 $X$ 类数据向右移动 2 个单位, $Y$ 类向左移动 4 个单位, 方差保持不变. 按照 120 个训练数据 250 个测试数据的比例随机划分运行 10 次, 得到平均精度从大到小依次为: C$^2$M 为 98.6 %, M$^4$ 为 98.4 %, SVM 为 97.4 %. 图 2 (b) 为其中一次随机试验的结果, 可以看到 M$^4$ 的分界线已经不能反映整体的方向信息, 而 C$^2$M 依然可以很好地反映方向整体信息, 从而进一步验证了第 1.5 节中的分析.

图 2 (c) 显示了 M$^4$ 使用式 (21) ∼ (23) 的模型在图 2 (b) 数据集上的分类精度和目标函数随 $\rho$ 的变化情况, 惩罚系数 $C$ 的取值分别为 $[10^7, 10^8, 10^9]$, 为了图形直观方便, 三个参数对应的目标函数的曲线分别向上移动 $[70, 60, 50]$, 这样的平移保持曲线的形状不变, 不影响后面的分析. 从图 2 (c) 可以看到, M$^4$ 在 $\rho = 0$ 时获得最好的测试精度, $C = 10^9$ 对应的曲线也体现了这点, 从 $C = 10^9$ 的曲线上还可以看到 M$^4$ 的目标函数存在局部最优点, 所以使用 M$^4$ 的一维搜索算法, 很有可能陷入局部最优, 因此, M$^4$ 搜索最优 $\rho$ 的范围需要小心地调整, 而 C$^2$M 是求解 QP, 一定存在全局最优, 只需要调整惩罚系数就可以了, 从而 C$^2$M 具有更好的鲁棒性.



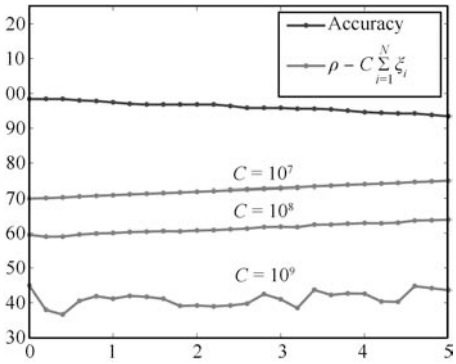

(a) 对比方向信息

(a) Demonstration of direction




(b) 两类样本交遇较多

(b) Demonstration of bad margin




(c) M$^4$ 的精度曲线

(c) Curves of accuracy and $\rho$ in M$^4$

图 2　模拟数据集分类结果

Fig. 2　Demonstration of toy data set

### 3.2　标准数据集实验分析

为了进一步验证 C$^2$M 的分类性能, 选取 UCI 标准数据集[5] 上的 10 套数据集, 各个数据集的规模列于表 1 中. 在这些数据集上, 使用线性核和高斯核分别与 M$^4$ 和 SVM 对比测试, 试验中, 惩罚系数 $C$ 和高斯核参数 $\gamma$ 通过网格搜索[9] 逐步寻找 $K$ 折交叉验证 ($K = 10$) 的最优参数, $\gamma$ 的搜索间隔为 $[2^{-15}, 2^{-13}, \cdots, 2^3]$, $C$ 的搜索间隔为 $[2^{-5}, 2^{-3}, \cdots, 2^{15}]$. 表 2 列出的最终的分类精度

是最优参数时 10 折测试正确率的均值以及标准差, 为了比较三个分类器的差异程度, 我们将 10 次测得的结果在 0.05 的显著性水平下作成对 $t$ 检验. 统计上看, 线性核时 C$^2$M 和 M$^4$ 分别有 7 个胜出, 稍好于 SVM 的 6 个; 从平均精度上看, 在 Glass、Ionosphere、Sonar 和 Vote 这 4 个数据集上的分类的正确率好于 M$^4$ 和 SVM, M$^4$ 和 SVM 各有 3 个数据获得较高的正确率. 采用高斯核时, C$^2$M 和 M$^4$ 的统计结果并不比 SVM 优, 分别有 5 个、3 个和 7 个胜出, 导致这一问题的原因很有可能是核化后稀疏的高维特征空间对方向信息不敏感.

表 1 UCI 数据集规模和全局信息指标

Table 1 Sizes and indexes of UCI data sets

| Data set | $N_x$ | $N_y$ | $n$ | $R_x$ | $R_y$ | $\triangle R$ | $\overline{\triangle S}$ | std($\triangle S$) |
|---|---|---|---|---|---|---|---|---|
| Breast | 357 | 212 | 30 | 0.26 | 0.24 | 0.02 | 0.10 | 0.07 |
| Glass | 70 | 76 | 9 | 0.46 | 0.38 | 0.08 | 0.15 | 0.14 |
| Heart-disease | 37 | 188 | 13 | 0.89 | 0.82 | 0.07 | 0.13 | 0.11 |
| Ionosphere | 225 | 126 | 34 | 0.64 | 0.36 | 0.28 | 0.24 | 0.17 |
| Musk | 207 | 209 | 167 | 0.61 | 0.56 | 0.05 | 0.07 | 0.06 |
| Parkinsons | 48 | 147 | 22 | 0.20 | 0.62 | 0.42 | 0.18 | 0.09 |
| Pima | 268 | 500 | 8 | 0.22 | 0.16 | 0.06 | 0.06 | 0.03 |
| Sonar | 97 | 111 | 60 | 0.45 | 0.30 | 0.15 | 0.07 | 0.05 |
| Vote | 168 | 267 | 16 | 0.50 | 0.30 | 0.20 | 0.13 | 0.10 |
| Yeast | 463 | 429 | 6 | 0.10 | 0.15 | 0.05 | 0.03 | 0.03 |

接下来, 我们考虑一些量化指标来衡量协方差矩阵对分类器的贡献程度. 为了进一步分析各个数据集上样本的整体信息对测试结果的影响, 考虑如下统计量, 设样本 $Z = \{\boldsymbol{z}_i\}_{i=1}^N, \boldsymbol{z}_i \in \mathbf{R}^n$, 则 $Z$ 的各个特征的标准差 $S = \{s_i\}_{i=1}^n, s_i \in \mathbf{R}$ 反映了样本在每个特征方向上分布的离散程度. $S$ 的极差 range($S$) 又可以进一步反映不同方向之间的差异程度, 即 range($S$) 越小, 说明各向同性度大; 反之, 各向异性度大. 参考图 1 中二维的直观情况, 对于各向同性的情况, 即椭圆趋于圆时 (高维空间是超椭球趋于超球), 整体信息对分类贡献小. 各向异性度大时, 整体信息的贡献度不一定大, 特殊地, $\Sigma_x = \Sigma_y$ 时, 即相当于图 1 中两个椭圆旋转方向相同的情况, 这种情况下整体信息对分类的贡献也不大. 用如下指标来衡量两类样本数据分布旋转的差异程度, 对于 $X$ 和 $Y$ 两类样本的标准差 $S_x, S_y$ 和标准差的极差 $R_x = \text{range}(S_x), R_y = \text{range}(S_y)$, 两个极差的差异程度 $\triangle R = |\text{range}(S_x) - \text{range}(S_y)|$, 两个标准差差异的均值 $\overline{\triangle S} = \overline{(S_x - S_y)}$ 和标准差 $\text{std}(\triangle S) = \text{std}(S_x - S_y)$ 都可以反映两个样本是否向一个方向旋转的情况, 指标值越大, 说明旋转的方向差异大, 整体信息对分类的贡献大; 反之, 指标越小, 说明旋转的方向相同, 整体信息对分类的贡献小. 各个数据集在样本空间的指标数据列于表 2 中, 可以看到, Breast、Musk 和 Pima 三个数据集的 $\triangle$range, $\overline{(S_x - S_y)}$, $\text{std}(S_x - S_y)$ 指标值是最小, 所以整体信息对分类的贡献程度不大, 除 *Pima* 上三种分类器统计意义上无明显差异外, 其他两个数据集上 C$^2$M 和 M$^4$ 的分类精度都比 SVM 差.

M$^4$ 核化后, 除 Glass 数据集外, 其他数据集取得最优解时 $\rho = 0$, 退化为感知器准则的线性分类器, 此时, M$^4$ 无法利用协方差整体信息, 所以, M$^4$ 核化后的分类精度不高, 仅有 3 个胜出, 而 C$^2$M 此时依然可以很好地利用整体信息, 有 5 个胜出, 取得了较好的分类精度. 核化后的方向信息受核参数的影响, 并不一定能保证和线性核时一样的方向性, 所以在某个数据集上, 线性核时精度高, 核化后并不一定还能保持一样好. 可以看到, 线性核时 C$^2$M 在 Glass 和 Ionosphere, M$^4$ 在 Parkinsons 和 Yeast 数据集取得了好分类精度, 而核化后分类精度都比 SVM 差; 线性核 C$^2$M 仅在 Parkinsons 数据集上通过调整核参数, 取得了比 M$^4$ 和 SVM 好的分类精度.

表 2 UCI 数据集上 C$^2$M, M$^4$ 和 SVM 分类精度对比 (%)

Table 2 Comparison of classification accuracies among C$^2$M, M$^4$ and SVM on UCI data sets (%)

| | Linear kernel | | | Gaussian kernel | | |
|---|---|---|---|---|---|---|
| Data set | C$^2$M | M$^4$ | SVM | C$^2$M | M$^4$ | SVM |
| Breast | 96.51 ± 1.35 | 96.10 ± 2.24 | **97.94 ± 1.24** | 95.31 ± 3.13 | 94.15 ± 3.46 | **98.11 ± 2.54** |
| Glass | **73.95 ± 9.10** | **72.72 ± 8.92** | **72.00 ± 7.24** | 76.83 ± 10.97 | 73.32 ± 7.49 | **79.63 ± 1.13** |
| Heart-disease | 85.42 ± 1.49 | **88.14 ± 2.14** | 85.63 ± 2.22 | 82.79 ± 2.55 | **87.67 ± 2.65** | **87.33 ± 2.92** |
| Ionosphere | **90.56 ± 5.14** | **90.49 ± 3.04** | 85.90 ± 5.32 | 90.86 ± 3.88 | 79.18 ± 12.65 | **95.19 ± 4.20** |
| Musk | 84.15 ± 2.28 | 83.04 ± 3.58 | **87.25 ± 4.17** | 93.34 ± 2.85 | **95.59 ± 2.72** | 93.29 ± 3.01 |
| Prkinsons | **89.64 ± 6.05** | **90.73 ± 5.53** | **89.49 ± 3.95** | **97.29 ± 5.05** | 95.16 ± 3.48 | 94.38 ± 3.70 |
| Pima | **77.61 ± 5.18** | **77.03 ± 3.59** | **77.68 ± 3.33** | **77.02 ± 4.88** | 74.11 ± 2.05 | **76.93 ± 2.31** |
| Sonar | **77.58 ± 3.65** | **75.46 ± 7.83** | **76.00 ± 4.46** | **92.35 ± 3.45** | 89.44 ± 5.47 | 88.62 ± 2.24 |
| Vote | **96.85 ± 2.63** | 95.06 ± 2.19 | 94.31 ± 3.07 | **95.85 ± 3.36** | 91.33 ± 2.77 | **94.52 ± 3.19** |
| Yeast | **65.15 ± 2.55** | **66.43 ± 1.67** | 63.14 ± 2.87 | **64.03 ± 3.70** | **63.14 ± 2.45** | **66.04 ± 6.92** |

为了对比 $C^2M$、$M^4$ 和 SVM 平均运行时间, 考虑需要消除不同编程语言和编程风格之间的差异, 三个模型的求解统一使用 SeDuMi[10] 求解, 并且只计算 IPM (Interior point method) 的迭代时间, 忽略组装矩阵和格式化输出的时间, 这样可以更准确地反映算法求解的时间. 图 3 中的运行时间是各个数据集使用线性核 10 折交叉验证的平均 IPM 迭代时间. 考虑到运行时间数值的差距比较大, 图 3 中 $Y$ 轴使用了对数坐标刻度. 图 3 中可以直观地看出 $C^2M$ 的运行时间远远小于 $M^4$ 的运行时间, 从而进一步说明了 $C^2M$ 考虑整体信息时, 算法的运行效率比 $M^4$ 高.

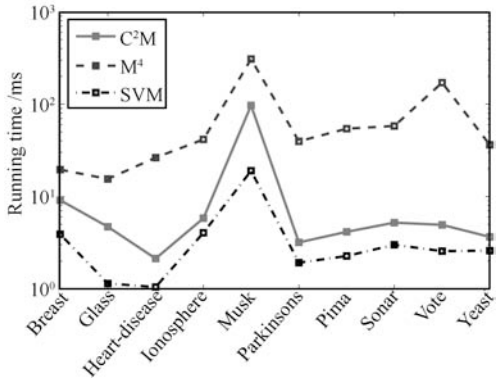



图 3 运行时间对比

Fig. 3 Comparisons of running time

## 4 结论

本文通过分析 $M^4$ 算法的不足, 提出了一种新的协作式局部整体分类机, 即 $C^2M$. 该算法和 $M^4$ 同样都利用了协方差作为整体信息, 但是可以在保持精度不低于 $M^4$ 的前提下, 将时间复杂度从 $M^4$ 的 $O(N^4)$ 降为 $O(2N^3)$. 通过分析 $M^4$ 的最大边界 $\rho$ 的物理意义, 指出 $M^4$ 在两类数据交遇区较多的情况下, 全局信息失效并退化为感知器准则的线性分类器, 而 $C^2M$ 依然可以有效利用整体信息. 本文还提出了 4 个指标衡量整体信息对分类的贡献度. UCI 数据集上的试验表明线性核时 $C^2M$ 和 $M^4$ 分类精度好于 SVM. 在稀疏的高维核空间, 如何有效地利用全局信息, 以及是否还有其他更好的整体信息可以被用来提高分类器的精度是我们下一步将要研究的方向.

## References


1 Fisher R A. The use of multiple measures in taxonomic problems. *Annals of Eugenics*, 1936, **7**: 179−188

2 Lanckriet G R G, Ghaoui L E, Bhattacharyya C, Jordan M I. A robust minimax approach to classification. *The Journal of Machine Learning Research*, 2003, **3**: 555−582

3 Huang K, Yang H, King I, Lyu M R. Maxi-min margin machine: learning large margin classifiers locally and globally. *IEEE Transactions on Neural Networks*, 2008, **19**(2): 260−272

4 Collobert R, Bengio S, Bengio Y. A parallel mixture of SVMs for very large scale problems. *Neural Computation*, 2002, **14**(5): 1105−1114

5 Frank A, Asuncion A. UCI Machine learning repository [Online], available: http://archive.ics.uci.edu/ml, July 20, 2011

6 Kittler J, Hatef M, Duin R P W, Matas J. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, **20**(3): 226−239

7 Platt J C. Fast training of support vector machines using sequential minimal optimization. *Advances in Kernel Methods*. Cambridge: The MIT Press, 1999. 185−208

8 Scholkopf B, Smola A J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge: The MIT Press, 2002

9 Chang C C, Lin C J. LIBSVM: a library for support vector machines [Online], available: http://www.csie.ntu.edu.tw/∼cjlin/libsvm/, July 20, 2011

10 Sturm J F. Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 1999, **11**(1): 625−653



**张战成** 江南大学博士研究生. 主要研究方向为模式识别与网络安全技术. 本文通信作者.
E-mail: cimszhang@163.com
(**ZHANG Zhan-Cheng** Ph. D. candidate at Jiangnan University. His research interest covers pattern recognition and network security. Corresponding author of this paper.)



**王士同** 江南大学教授. 主要研究方向为人工智能、模式识别、数据挖掘、神经网络、模糊系统、医学图像处理和生物信息学. E-mail: wxwangst@yahoo.com.cn
(**WANG Shi-Tong** Professor at Jiangnan University. His research interest covers artificial intelligence, pattern recognition, data mining, neural network, fuzzy system, medical image processing, and bioinformation.)



**钟富礼** 香港理工大学计算学系副教授. 主要研究方向为数据挖掘、机器智能、计算智能、多媒体技术和模式识别.
E-mail: cskchung@comp.polyu.edu.hk
(**CHUNG Fu-Lai** Associate professor in the Department of Computing, Hong Kong Polytechnic University. His research interest covers data mining, machine learning, computational intelligence, multimedia processing, and pattern recognization.)