

一种新的构造 SVM 分类器的几何最近点法

刘振丙¹ 陈忠¹ 刘建国¹

摘要 引入了尺度化凸壳 (Scaled convex hull, SCH) 的概念, 证明了与之相关的性质, 通过这些性质可以把求解线性不可分支持向量机 (Support vector machine, SVM) 的问题转化为计算两类训练样本分别生成的尺度化凸壳间的最近点对的问题. 然后可以用几何最近点法计算尺度化凸壳间的最近点对, 把垂直平分连接最近点对线段的超平面作为线性不可分问题的分类超平面. 此外, 还把这种方法推广到非线性情形, 并给出了解决非线性问题的一种简化算法. 理论分析和实验均表明, 与已有的方法相比, 尺度化凸壳法在取得相同分类成功率的同时, 训练时间大大减少, 特别适用于样本较多的大规模分类问题.

关键词 最大间隔, 尺度化凸壳, 最近点对, Mitchell-Dem'yanov-Malozemov (MDM) 算法

DOI 10.3724/SP.J.1004.2010.00791

A Novel Geometric Nearest Point Algorithm for Constructing SVM Classifiers

LIU Zhen-Bing¹ CHEN Zhong¹ LIU Jian-Guo¹

Abstract In this paper, the notion of “scaled convex hull (SCH)” is employed and a set of theoretical results are exploited to support it, through which the nonseparable support vector machine (SVM) classification problems can be transformed to the problems of computing the pair of nearest points between SCHs. As a practical application of the SCH framework, a popular nearest point algorithm has been adopted to find the pair of nearest points between SCHs (each is generated by training patterns of each class), and the separating hyperplane bisects, and is normal to the line segment joining these two nearest points. Then, the proposed method is generalised to solve nonlinear problems and a simplified version for the nonlinear case is presented. The theoretical analysis and experiments show that the proposed method may achieve better performance than the state-of-the-art methods in terms of the kernel evaluations and execution time, making it suitable for large scale classification.

Key words Maximal margin, scaled convex hull (SCH), the pair of nearest points, Mitchell-Dem'yanov-Malozemov (MDM) algorithm

支持向量机 (Support vector machine, SVM)^[1-2] 以统计学习理论为基础, 具有简洁的数学形式、直观的几何解释和良好的泛化能力, 是一种解决分类、回归和概率密度估计等问题的有效工具.

对 SVM 几何解释的研究最早可追溯到文献 [3], 该文首先研究了感知机的几何解释, 并把这种几何思想推广到 SVM 情形中. 此后, SVM 的几何解释吸引了越来越多学者的关注. 如在文献 [4] 中, 作者通过凸壳 (Convex hull) 的概念给出了线性可分 SVM 的几何解释, 通过压缩凸壳 (Reduced convex hull, RCH) 的概念给出了线性不可分 SVM 的几何解释, 证明了求解线性可分 SVM 最优 (最大间隔) 分类器等价于求解两类训练样本分别生成的凸壳间的最近点对. 在文献 [5] 中, 作者证明了求解线性不可分 SVM 最优分类器等价于求解压缩

凸壳间的最近点对. 这促使人们提出求解大规模 SVM 分类问题的几何化算法. 比较早的几何方法有 Keerthi 引入的最近点算法^[6] 及其他一些方法^[7]. 后来, Mavroforakis 等^[8-10] 研究了压缩凸壳的一些性质, 在此基础上给出了基于压缩凸壳的 SVM 分类器设计方法, 同样的思想也出现在 Tao 等^[11] 的文章中. 这些方法都把 SVM 分类器的构造问题转化为求解压缩凸壳间的最近点对的问题, 然后用已有的方法快速求解最近点对, 而 SVM 最大间隔分类超平面 (即最大间隔分类器对应的超平面) 就是垂直平分连接最近点对线段的超平面.

凸壳间的最近点对是直接依赖于凸壳的顶点的. 在线性可分情形下, 凸壳的顶点就是原来的训练点的一部分; 而对于线性不可分问题, 必须采用压缩凸壳定义中的线性组合形式进行穷举以遍历所有顶点, 这大大增加了计算复杂度. 为此, 本文采用了一种新的缩小凸壳的方法——尺度化凸壳法. 通过尺度化凸壳, 可以把线性不可分 SVM 求解问题转化为计算最近点对的问题. 此外, 我们把这种方法推广到非线性情形, 并给出了求解非线性问题的简化算法.

1 SVM 的几何解释

本文只考虑二分类问题. 设 $X^+ = \{\mathbf{x}_i : i \in$

收稿日期 2009-04-20 录用日期 2009-06-24
Manuscript received April 20, 2009; accepted June 24, 2009
国家自然科学基金 (40801164) 资助
Supported by National Natural Science Foundation of China for Young Scholars (40801164)
1. 华中科技大学图像识别与人工智能研究所多谱信息处理技术国家重点实验室 武汉 430074
1. National Key Laboratory of Science and Technology on Multi-spectral Information Processing, Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan 430074

I^+ , $X^- = \{\mathbf{x}_i : i \in I^-\}$ 是两个训练样本集, X^+ 中每个样本的标签为 +1, X^- 中每个样本的标签为 -1, $n^+ = |I^+|$, $n^- = |I^-|$, $n = n^+ + n^-$, $I = I^+ \cup I^-$, 假设 X^+ 和 X^- 的样本独立同分布. 当 X^+ 和 X^- 线性可分时, SVM 通过构造一个线性最优分类超平面 $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + t = 0$, 将这两类样本点正确分开, 并使得它们之间的间隔 (Margin) $2/\|\mathbf{w}\|^2$ 最大. 在这种情形下, 最大间隔分类超平面 (即最优分类超平面) 就是垂直平分连接两个凸壳间的最近点对的线段的超平面, 其对应的分类器则称为 SVM 最大间隔分类器. 其中样本集 X^+ 所生成的凸壳定义为

$$Cv(X^+) = \left\{ \sum_{i \in I^+} a_i \mathbf{x}_i : \sum_{i \in I^+} a_i = 1, 0 \leq a_i \leq 1 \right\} \quad (1)$$

同样地, 可以定义样本集 X^- 所生成的凸壳.

对于线性不可分问题, 样本点所生成的两个凸壳是相交的, 分类器未能把所有样本点正确分开. 为此, 需要引入压缩凸壳的概念^[5], 把线性不可分问题转化为线性可分问题. X^+ 所生成的压缩凸壳定义为

$$R(X^+, \mu) = \left\{ \sum_{i \in I^+} a_i \mathbf{x}_i : \sum_{i \in I^+} a_i = 1, 0 \leq a_i \leq \mu \right\} \quad (2)$$

同样地, 可以定义样本集 X^- 所生成的压缩凸壳.

式 (2) 中参数 μ 是压缩系数, 可以控制压缩凸壳的大小, 由于该参数受到了限制, 因此压缩凸壳比原凸壳 $Cv(X^+)$ 小. 通过不断地缩小两类样本点所生成的压缩凸壳, 可使压缩凸壳变得线性可分, 从而就可以用最近点算法寻找这两个压缩凸壳间的最近点对. 这种方法还可以推广到非线性情形. 文献 [5] 中证明, C -SVM 分类器的求解等价于压缩凸壳间的最近点对的求解, 这就为求解不可分 SVM 问题的几何方法奠定了理论基础.

前面已经提到, 凸壳间的最近点对是直接依赖于凸壳的顶点的. 在线性可分情形下, 凸壳的顶点就是原来的训练样本集的一部分; 而对于线性不可分问题, 压缩凸壳的每个顶点是 $\lceil 1/\mu \rceil$ 个原样本点的线性组合. 因此要计算所有的顶点, 必须采用压缩凸壳定义中的线性组合形式进行穷举, 以使得算法遍历所有顶点, 这大大增加了计算复杂度. 为了解决这个问题, Mavroforakis 等^[9] 研究了压缩凸壳的一些性质, 指出在实际计算中, 可以通过对原训练点在某个方向上的投影按大小排序, 然后对排序后的前 $\lceil 1/\mu \rceil$ 个最大投影进行特定的组合求出压缩凸壳的顶点, 这就避免了直接用原训练点的各种组合穷举导致的组合问题. 但是使用这种技巧仍然要对原训

练点集进行投影和排序, 这比线性可分问题要复杂得多, 并且复杂度随参数 μ 的改变而变化. 为了克服这一难题, 本文引入一种新的凸壳压缩法——尺度化凸壳方法^[12].

2 尺度化凸壳与 SVM

样本集 X^+ 所生成的尺度化凸壳定义为

$$SR(X^+, \lambda) = \left\{ \lambda \sum_{i \in I^+} a_i \mathbf{x}_i + (1 - \lambda) \mathbf{m}^+ : \sum_{i \in I^+} a_i = 1, 0 \leq a_i \leq 1 \right\} \quad (3)$$

其中, $\mathbf{m}^+ = \frac{1}{n^+} \sum_{i \in I^+} \mathbf{x}_i$ 是样本集 X^+ 的中心. 式 (3) 等价于

$$SR(X^+, \lambda) = \left\{ \sum_{i \in I^+} a_i (\lambda \mathbf{x}_i + (1 - \lambda) \mathbf{m}^+) : \sum_{i \in I^+} a_i = 1, 0 \leq a_i \leq 1 \right\} \quad (4)$$

同样地, 另一类样本集 X^- 所生成的尺度化凸壳定义为

$$SR(X^-, \lambda) = \left\{ \lambda \sum_{i \in I^-} a_i \mathbf{x}_i + (1 - \lambda) \mathbf{m}^- : \sum_{i \in I^-} a_i = 1, 0 \leq a_i \leq 1 \right\} \quad (5)$$

其中, $\mathbf{m}^- = \frac{1}{n^-} \sum_{i \in I^-} \mathbf{x}_i$ 是样本集 X^- 的中心. 而式 (5) 等价于

$$SR(X^-, \lambda) = \left\{ \sum_{i \in I^-} a_i (\lambda \mathbf{x}_i + (1 - \lambda) \mathbf{m}^-) : \sum_{i \in I^-} a_i = 1, 0 \leq a_i \leq 1 \right\} \quad (6)$$

下面以 $SR(X^+, \lambda)$ 为例, 给出尺度化凸壳的几何意义. 由式 (3) 可以看出, 若固定尺度参数 (压缩因子) λ , 则尺度化凸壳 $SR(X^+, \lambda)$ 的每一点是原凸壳 $Cv(X^+)$ 对应点和中心的线性组合. 换句话说, 尺度化凸壳上的点位于原凸壳点与中心点的连接线段上. 从几何直观上说, 尺度化凸壳上的点距离中心更近, 即尺度化凸壳比原凸壳更小. 因此, 尺度化凸壳与原凸壳有相同的几何形状, 而且形状不随尺度参数 λ 的变化而变化, 这也是我们称其为尺度化凸壳的原因.

为书写方便, 记 $\lambda \mathbf{x}_i + (1 - \lambda) \mathbf{m}^+$ 为 $\mathbf{x}'_i$ ($i \in I^+$), 记 $\lambda \mathbf{x}_j + (1 - \lambda) \mathbf{m}^-$ 为 $\mathbf{x}'_j$ ($j \in I^-$), 令 $X^+ = \{\mathbf{x}'_i :$

$i \in I^+\}$, $X^{-'} = \{\mathbf{x}'_j : j \in I^-\}$. 这样, 式 (4) 和式 (6) 可分别写成

$$SR(X^+, \lambda) = \left\{ \mathbf{w} : \mathbf{w} = \sum_{i \in I^+} a_i \mathbf{x}'_i, \sum_{i \in I^+} a_i = 1, 0 \leq a_i \leq 1 \right\} = \text{conv}(X^{+'}) \quad (7)$$

和

$$SR(X^-, \lambda) = \left\{ \mathbf{w} : \mathbf{w} = \sum_{i \in I^-} a_i \mathbf{x}'_i, \sum_{i \in I^-} a_i = 1, 0 \leq a_i \leq 1 \right\} = \text{conv}(X^{-'}) \quad (8)$$

式 (7) 和式 (8) 表明, 尺度化凸壳可以看成是由压缩后的点 $\mathbf{x}'_i$ 所生成的凸壳, 即尺度化凸壳 $SR(X^+, \lambda)$ 和 $SR(X^-, \lambda)$ 的顶点集分别为 $X^{+'}$ 和 $X^{-'}$. 而 $X^{+'}$ 和 $X^{-'}$ 与原凸壳的顶点集 X^+ 和 X^- 有相同数量的元素, 即尺度化凸壳与原凸壳有相同数量的顶点. 这些顶点可以很容易预先算出, 这就为求解最近点对提供了简单的方法, 并且求解最近点对的复杂度和压缩参数无关. 实际上, 由于尺度化凸壳与原凸壳有相同数量的顶点, 一旦顶点被预先算出, 求解这两类凸壳间的最近点对就有相同的复杂度. 压缩凸壳与尺度化凸壳的比较见图 1 所示.



图 1 压缩凸壳与尺度化凸壳的比较

Fig. 1 The comparison of a reduced convex hull and a scaled convex hull

通过不断减小参数 λ 的取值, 尺度化凸壳 $SR(X^+, \lambda)$ 和 $SR(X^-, \lambda)$ 不断变小直至可分. 当它们可分时, 就可以用最近点算法找出其最近点对, 而垂直平分最近点对的超平面就是最大间隔分类超平面, 对应的分类器称为基于尺度化凸壳的分类器. 这种构造分类器的思想和压缩凸壳情形是一致的, 因此该方法可以看成是求解 SVM 分类器的一种变形方法.

下面给出尺度化凸壳的几条性质.

性质 1. 当 $\lambda = 1$ 时, 尺度化凸壳 (3) 就是原凸壳 (1); 当 $\lambda = 0$ 时, 尺度化凸壳变为点 $\mathbf{m}$.

证明. 将 $\lambda = 1$ 和 $\lambda = 0$ 分别带入式 (3), 即得结论. $\square$

性质 2. 式 (4) 等价于

$$SR(X^+, \lambda) = \left\{ \sum_{i \in I^+} b_i \mathbf{x}_i : \sum_{i \in I^+} b_i = 1, \frac{1-\lambda}{n^+} \leq b_i \leq \lambda + \frac{1-\lambda}{n^+} \right\} \quad (9)$$

证明. 由定义, 尺度化凸壳上的每一点可表示为

$$\begin{aligned} \mathbf{w} &= \lambda \sum_{i \in I^+} a_i \mathbf{x}_i + (1-\lambda) \mathbf{m}^+ = \\ &= \lambda \sum_{i \in I^+} a_i \mathbf{x}_i + (1-\lambda) \frac{1}{n^+} \sum_{i \in I^+} \mathbf{x}_i = \\ &= \sum_{i \in I^+} \left(\frac{1-\lambda}{n^+} + \lambda a_i \right) \mathbf{x}_i \end{aligned}$$

令 $b_i = (1-\lambda)/n^+ + \lambda a_i$, 即得结论. $\square$

比较式 (2) 和式 (9) 可知, 当样本数目 n^+ 趋于无穷时, $(1-\lambda)/n^+$ 趋于 0, 尺度化凸壳收敛于压缩凸壳, 因此基于尺度化凸壳的分类器趋于压缩凸壳的分类器. 从这个角度看, 尺度化凸壳方法是压缩凸壳法的一个近似.

下面的定理给出使得两个尺度化凸壳可分的条件. 首先给出一些记号

$$\begin{aligned} r &= \|\mathbf{m}^+ - \mathbf{m}^-\| \\ r^+ &= \max_{i \in I^+} \|\mathbf{x}_i - \mathbf{m}^+\| \\ r^- &= \max_{i \in I^-} \|\mathbf{x}_i - \mathbf{m}^-\| \\ r_s^+ &= \max_{i \in I^+} \|\mathbf{x}'_i - \mathbf{m}^+\| \\ r_s^- &= \max_{i \in I^-} \|\mathbf{x}'_i - \mathbf{m}^-\| \end{aligned}$$

定理 1. 当 $\lambda^+ r^+ + \lambda^- r^- < r$ 时, 尺度化凸壳 $SR(X^+, \lambda^+)$ 和 $SR(X^-, \lambda^-)$ 可分.

证明. 由图 2 可以看出, 尺度化凸壳 $SR(X^+, \lambda^+)$ 被以 $\mathbf{m}^+$ 为圆心、 r_s^+ 为半径的圆所包围; $SR(X^-, \lambda^-)$ 被以 $\mathbf{m}^-$ 为圆心、 r_s^- 为半径的圆所包围. 当两个圆可分时, 两个凸壳是可分的, 而使得两个圆可分的条件是 $r > r_s^+ + r_s^-$.

又

$$\begin{aligned}
 r_s^+ + r_s^- = & \max_{i \in I^+} \|\mathbf{x}_i' - \mathbf{m}^+\| + \max_{i \in I^-} \|\mathbf{x}_i' - \mathbf{m}^-\| = \\
 & \max_{i \in I^+} \|\lambda^+ \mathbf{x}_i + (1 - \lambda^+) \mathbf{m}^+ - \mathbf{m}^+\| + \\
 & \max_{i \in I^-} \|\lambda^- \mathbf{x}_i + (1 - \lambda^-) \mathbf{m}^- - \mathbf{m}^-\| = \\
 & \max_{i \in I^+} \|\lambda^+ \mathbf{x}_i - \lambda^+ \mathbf{m}^+\| + \max_{i \in I^-} \|\lambda^- \mathbf{x}_i - \lambda^- \mathbf{m}^-\| = \\
 & \lambda^+ \max_{i \in I^+} \|\mathbf{x}_i - \mathbf{m}^+\| + \lambda^- \max_{i \in I^-} \|\mathbf{x}_i - \mathbf{m}^-\| = \\
 & \lambda^+ r^+ + \lambda^- r^-
 \end{aligned}$$

即当 $\lambda^+ r^+ + \lambda^- r^- < r$ 时, 尺度化凸壳 $SR(X^+, \lambda^+)$ 和 $SR(X^-, \lambda^-)$ 可分. $\square$

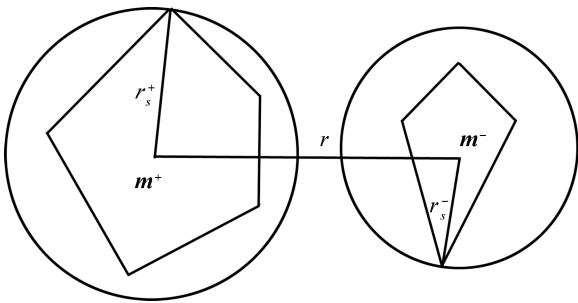


图 2 尺度化凸壳可分的条件

Fig. 2 The condition making the two SCHs separable

定理 2. 通过适当的参数调整, 当训练样本数量变大时, 基于尺度化凸壳的分类器的法向量收敛于 C-SVM 分类器的法向量.

证明. 根据性质 2, 寻找可分的尺度化凸壳 $SR(X^+, \lambda)$ 和 $SR(X^-, \lambda)$ 间的最近点对的优化问题可表示为

$$\min_{\mathbf{b}} \frac{1}{2} \left\| \sum_{i \in I^+} b_i \mathbf{x}_i - \sum_{i \in I^-} b_i \mathbf{x}_i \right\| \quad (10)$$

s.t.

$$\begin{aligned}
 & \sum_{i \in I^+} b_i = 1, \quad \sum_{i \in I^-} b_i = 1 \\
 & \lambda + \frac{1 - \lambda}{n^+} \geq b_i \geq \frac{1 - \lambda}{n^+}, \quad i \in I^+ \\
 & \lambda + \frac{1 - \lambda}{n^-} \geq b_i \geq \frac{1 - \lambda}{n^-}, \quad i \in I^-
 \end{aligned} \quad (11)$$

通过尺度化目标函数, 可以重写为

$$\min \frac{1}{4} \sum_{i, j \in I} b_i b_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \quad (12)$$

s.t.

$$\begin{aligned}
 & \sum_{i \in I} b_i y_i = 0, \quad \sum_{i \in I} b_i = 2 \\
 & \lambda + \frac{1 - \lambda}{n^+} \geq b_i \geq \frac{1 - \lambda}{n^+}, \quad i \in I^+ \\
 & \lambda + \frac{1 - \lambda}{n^-} \geq b_i \geq \frac{1 - \lambda}{n^-}, \quad i \in I^-
 \end{aligned} \quad (13)$$

假设 $(b_1^*, b_2^*, \dots, b_n^*)$ 是上述问题的解, $\tilde{f}(\mathbf{x}) = \langle \tilde{\mathbf{w}}, \mathbf{x} \rangle + \tilde{t}$ 是要求的判别函数, 则有

$$\tilde{\mathbf{w}} = \frac{1}{2} \left(\sum_{i \in I^+} b_i^* \mathbf{x}_i - \sum_{i \in I^-} b_i^* \mathbf{x}_i \right) = \frac{1}{2} \sum_{i \in I} b_i^* y_i \mathbf{x}_i \quad (14)$$

$$\tilde{t} = -\langle \tilde{\mathbf{w}}, \mathbf{p} \rangle \quad (15)$$

其中, $\mathbf{p} = \frac{1}{2} \left(\sum_{i \in I^+} b_i^* \mathbf{x}_i + \sum_{i \in I^-} b_i^* \mathbf{x}_i \right)$, 即分类超平面垂直平分最近点对连接线段.

C-SVM 的求解问题可以表达为

$$\min_{\mathbf{w}, \xi, t} C * \left(\sum_{i \in I} \xi_i \right) + \|\mathbf{w}\|^2$$

s.t.

$$y_i (\langle \mathbf{x}_i, \mathbf{w} \rangle + t) \geq 1 - \xi_i, \quad \xi_i > 0$$

其中, $C > 0$ 是可调参数. 假设 $(\mathbf{w}', t')$ 是上述问题的解, 则判别函数为

$$f'(\mathbf{x}) = \langle \mathbf{w}', \mathbf{x} \rangle + t' \quad (16)$$

上述问题的 wolf 对偶为^[6]

$$\min_{\mathbf{c}} \frac{1}{4} \sum_{i, j \in I} c_i c_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \quad (17)$$

s.t.

$$\sum_{i \in I} c_i y_i = 0, \quad \sum_{i \in I} c_i = 2, \quad C \geq c_i \geq 0 \quad (18)$$

假设 $(c_1^*, c_2^*, \dots, c_n^*)$ 是上述对偶问题的解, 则 $\mathbf{w}' = \frac{1}{2} \sum_{i \in I} c_i^* y_i \mathbf{x}_i$.

可以看出, 上面两个优化问题的不同之处仅在于限制条件 (13) 和 (18) 不同. 因此令 $\lambda = C$, 则当样本数量变大时, $(1 - \lambda)/n^+$ 和 $(1 - \lambda)/n^-$ 收敛于 0, 约束条件 (13) 收敛于式 (18), 优化问题 (12) 的解 $\tilde{\mathbf{w}}$ 趋于式 (17) 的解 $\mathbf{w}'$. 即基于尺度不变压缩的分类器的法向量趋于 C-SVM 分类器的法向量.

C-SVM 的 KKT (Karush-Kuhn-Tucker) 条件为

$$(C - c_i) \xi_i = 0 \quad (19)$$

$$c_i (y_i (\langle \mathbf{x}_i, \mathbf{w}' \rangle + t') - \xi_i) = 0 \quad (20)$$

假设 $\mathbf{w}'$ 和 $\tilde{\mathbf{w}}$ 是基本相同的, 把式 (20) 的两边同乘以 y_i , 累加 i 并结合式 (19), 得到

$$t' = \tilde{t} - \frac{C}{2} \sum_{i \in I} y_i \xi_i \quad (21)$$

□

由定理 2 的证明过程可以看出, 尺度化凸壳分类器的法向量 $\tilde{\mathbf{w}}$ 收敛于 C -SVM 分类器的法向量 $\mathbf{w}'$, 而对于偏置 $\tilde{t}$ 和 t' 则没有类似的结论. 但是文献 [1] 已经指出, 分类器的泛化能力由最大间隔 $2/\|\mathbf{w}\|^2$ 决定, 而最大间隔由法向量 $\mathbf{w}$ 决定, 因此法向量的选择是关键. 尺度化凸壳分类器的法向量趋于 C -SVM 分类器的法向量, 因此可以有比较好的泛化能力. 而对于偏置的选择, 文献 [4] 也指出, 并没有哪种选择偏置的方法是最好的, 而我们选择偏置的方法和压缩凸壳是相同的 (即选择最近点对连接线段的中垂线), 也可以作为一种新的候选. 综上所述, 尺度化凸壳分类器是 C -SVM 分类器的一种近似逼近. 换句话说, 通过尺度化凸壳, 我们可以把线性不可分 SVM 问题转化为计算最近点对问题.

3 构造 SVM 分类器的最近点法

通过上面的论述, 我们可以把求解线性不可分 SVM 问题转化为求解尺度化凸壳间的最近点对的问题. Mitchell-Dem'yanov-Malozemov (MDM) 算法^[6] 是通过求解最近点对构造 SVM 分类器的一个有效的方法, 该方法的优点是计算过程只涉及到凸壳的顶点. 而尺度化凸壳 $SR(X^+, \lambda)$ 和 $SR(X^-, \lambda)$ 的顶点集 $X^{+'}$ 和 $X^{-'}$ 很容易算出, 因此我们用该算法计算其最近点对.

定义 1. 设 Q 为 $\mathbf{R}^d$ 上的一个凸紧集, 称 $\eta(\mathbf{y}) = \max_{\mathbf{x} \in Q} \langle \mathbf{x}, \mathbf{y} \rangle$ 为 Q 上的支持函数.

定义 2. 设 $s(\mathbf{y})$ 是 $\mathbf{R}^d$ 的一个凸函数, 如果有

$$s(\mathbf{y}) \in \{\mathbf{x} : \langle \mathbf{x}, \mathbf{y} \rangle = \eta(\mathbf{y})\} \cap Q$$

$$\mathbf{y} \neq \mathbf{0}, s(\mathbf{0}) \in Q$$

则称 $s(\mathbf{y})$ 为 $\mathbf{R}^d$ 上的关联函数.

显然, $s(\mathbf{y})$ 为 Q 的一个顶点.

求解线性 SVM 分类器的 MDM 算法如下:

步骤 1. 初始化. 任取 $\mathbf{w}_1$ 为 $\mathbf{x} \in X^{+'}$, $\mathbf{w}_2$ 为 $\mathbf{x} \in X^{-'}$, 设定停止精度 ε .

步骤 2. 检验停止条件. 求 $\mathbf{x}'_i$ 使得

$$t \in \arg \min_{i \in I^+ \cup I^-} m(\mathbf{x}'_i)$$

$$\text{其中, } m(\mathbf{x}'_i) = \frac{\langle \mathbf{x}'_i - \mathbf{w}_2, \mathbf{w}_1 - \mathbf{w}_2 \rangle}{\|\mathbf{w}_1 - \mathbf{w}_2\|} \quad (i \in I^+), m(\mathbf{x}'_i) = \frac{\langle \mathbf{x}'_i - \mathbf{w}_1, \mathbf{w}_2 - \mathbf{w}_1 \rangle}{\|\mathbf{w}_1 - \mathbf{w}_2\|} \quad (i \in I^-).$$

如果 $\|\mathbf{w}_1 - \mathbf{w}_2\| - m(\mathbf{x}'_i) < \varepsilon$, 则法向量 $\mathbf{w} = \mathbf{w}_1 - \mathbf{w}_2$ 和偏置 $(\|\mathbf{w}_1\|^2 - \|\mathbf{w}_2\|^2)/2$ 就给出了分类器, 否则取 $\mathbf{z} = \mathbf{w}_1 - \mathbf{w}_2$, 转到步骤 3.

步骤 3. 局部优化与更新. 如果 $\mathbf{x}'_t \in X^{+'}$, 求 $t_{\min} \in I^+$, 使得

$$\langle -\mathbf{z}, \mathbf{x}'_{t_{\min}} \rangle = \min\{\langle -\mathbf{z}, \mathbf{x}'_i \rangle : a_i > 0, i \in I^+\}$$

令

$$\mathbf{d} = s(-\mathbf{z}) - \mathbf{x}'_{t_{\min}}, \mathbf{z}' = \mathbf{z} + a_{t_{\min}} * \mathbf{d}$$

取 $\mathbf{z}^{\text{new}}$ 为 $\mathbf{z}$ 和 $\mathbf{z}'$ 连线上的最小范数点, 令

$$\mathbf{w}_2^{\text{new}} = \mathbf{w}_2, \mathbf{w}_1^{\text{new}} = \mathbf{z}^{\text{new}} + \mathbf{w}_2^{\text{new}}$$

如果 $\mathbf{x}'_t \in X^{-'}$, 求 $t_{\min} \in I^-$, 使得

$$\langle -\mathbf{z}, \mathbf{x}'_{t_{\min}} \rangle = \min\{\langle -\mathbf{z}, \mathbf{x}'_i \rangle : a_i > 0, i \in I^-\}$$

令

$$\mathbf{d} = s(-\mathbf{z}) - \mathbf{x}'_{t_{\min}}, \mathbf{z}' = \mathbf{z} + a_{t_{\min}} * \mathbf{d}$$

取 $\mathbf{z}^{\text{new}}$ 为 $\mathbf{z}$ 和 $\mathbf{z}'$ 连线上的最小范数点, 令

$$\mathbf{w}_1^{\text{new}} = \mathbf{w}_1, \mathbf{w}_2^{\text{new}} = \mathbf{w}_1^{\text{new}} - \mathbf{z}^{\text{new}}$$

返回步骤 2 继续执行.

算法得到的判别函数为 $f(\mathbf{x}) = (\mathbf{w}_1 - \mathbf{w}_2)^T \mathbf{x} + (\|\mathbf{w}_1\|^2 - \|\mathbf{w}_2\|^2)/2$, 分类超平面为 $f(\mathbf{x}) = 0$.

关于算法的直观解释, 文献 [6] 已经给出, 这里不再赘述.

在文献 [11] 中, 作者给出了求解可分问题的 MDM 算法和基于压缩凸壳的求解不可分问题的 MDM 算法. 接下来, 我们将比较新算法与上述两个算法的异同. 我们主要比较步骤 2, 此处是算法最耗时、区别最大的地方. 由算法的求解步骤可以看出, 本文方法与求解可分问题的 MDM 算法有相同的计算复杂度. 这是因为可分与不可分两种情形下凸壳和尺度化凸壳有着相同数量的顶点, 并且尺度化凸壳的顶点很容易计算, 而顶点数目决定了算法的复杂度. 本文方法比基于压缩凸壳的 MDM 算法更简单, 这是因为压缩凸壳的顶点是原训练样本的组合, 需要对原训练样本在某个方向上的投影按大小排序后再进行组合才能算出, 这大大增加了计算复杂度.

上面的算法仅仅涉及到样本点的内积运算, 通过应用核技巧, 可用来求解非线性分类问题. 但由于不知道映射 ϕ 的具体形式, 这给计算中心 $\frac{1}{n^+} \sum_{i \in I^+} \phi(\mathbf{x}_i)$, $\frac{1}{n^-} \sum_{i \in I^-} \phi(\mathbf{x}_i)$ 和顶点带来困难. 为此, 我们采用了一种简化技巧, 不是把原样本点 $\mathbf{x}_i$, 而是把顶点 $\mathbf{x}'_i$ 通过 ϕ 直接映射到特征空间. 这样 $\phi(\mathbf{x}'_i)$ 就是生成特征空间中的凸壳的顶点, 然后就可以使用 MDM 算法通过核技巧求出由 $\phi(\mathbf{x}'_i)$ ($i \in I^+$ 或 I^-) 生成的凸壳间的最近点对, 进而构造最大间隔分类器. 此时只需要预先算出 $\mathbf{x}'_i$ 并存储

起来, 这大大简化了计算. 由于特征空间中的凸壳是 $\phi(\mathbf{x}_i)$ 的线性组合, 而最近点对通常只依赖于少部分顶点, 因此使用这种技巧还可以保持解系数的稀疏性. 这就是求解非线性问题的简化算法.

4 实验结果及分析

我们用 Matlab 实现了新算法 (记为 MDM-S), 作为对照, 我们还实现了另外两个算法: 1) 一种改进的序贯最小优化算法^[13-14] (记为 SMO); 2) 基于压缩凸壳的 MDM 算法^[8,11] (记为 MDM-R). 我们从公共测试数据库 UCI 中选择了若干数据集来比较上述算法的性能. 由于算法的运行时间主要是在核函数的计算上, 所以以核函数的计算次数和程序的执行时间作为评测标准. 选择高斯核函数 $K(\mathbf{x}, \mathbf{y}) = e^{-\|\mathbf{x}-\mathbf{y}\|^2/2\sigma^2}$, 使用交叉验证策略 (为了便于比较, 取和文献 [8] 相同的样本数, 训练样本的具体数目分

别选为 400, 160, 400, 170, 666, 700, 500, 16 000, 见表 1; 然后在其余的数据集中选择测试样本, 数目分别为 368, 500, 4 600, 100, 400, 300, 180, 16 000), 精度 ε 取为 0.05. 参数 C 的选择采用的是交叉验证中常用的方法^[7,14]: 从区间 $[0, 5\ 000]$ 中均匀地选取 100 个数, 即 $50 \times i$ ($i = 1, \dots, 100$), 分别用于分类器的训练; 然后, 用所得分类器分别对测试样本进行分类, 使测试样本的分类正确率最高的数即为 C . 参数 λ 的选择采用类似的方法: 从区间 $[0, 1]$ 中均匀地选取 100 个数, 即 $i/100$ ($i = 1, \dots, 100$), 分别用于分类器的训练, 然后用所得分类器分别对测试样本进行分类, 使测试样本分类正确率最高的数即为 λ . 参数 μ 采用文献 [8] 中的值 (也是用和参数 λ 的选择类似的方法). 在参数确定的同时, 也得到了测试样本的最高分类正确率并用其评价分类性能 (泛化能力), 即表 1 中的正确率. 实验结果如表 1.

表 1 各个方法在不同数据集下的核计算次数和训练时间比较

Table 1 Results achieved for each algorithm in the number of kernel evaluations and run times

数据集	算法	训练样本数	参数	正确率 (%)	核计算次数	时间 (s)
Diabetes	SMO	400	$\sigma = 100, C = 100$	76.70 ± 1.8	1.5×10^6	8.4
Diabetes	MDM-R	400	$\sigma = 100, \mu = 0.0075$	76.30 ± 1.8	3.0×10^5	5.5
Diabetes	MDM-S	400	$\sigma = 100, \lambda = 0.94$	76.80 ± 1.8	1.7×10^5	3.6
Thyroid	SMO	160	$\sigma = 30, C = 1\ 000$	94.6 ± 2.1	8.3×10^4	1.7
Thyroid	MDM-R	160	$\sigma = 30, \mu = 0.05$	94.7 ± 2.2	3.1×10^4	1.0
Thyroid	MDM-S	160	$\sigma = 30, \lambda = 0.84$	94.6 ± 2.0	2.3×10^4	1.0
Waveform	SMO	400	$\sigma = 20, C = 1\ 000$	89.20 ± 0.5	2.2×10^6	65.0
Waveform	MDM-R	400	$\sigma = 20, \mu = 0.02$	88.30 ± 0.8	1.5×10^6	49.0
Waveform	MDM-S	400	$\sigma = 20, \lambda = 0.62$	88.20 ± 0.7	0.8×10^6	32.0
Heart	SMO	170	$\sigma = 120, C = 1\ 000$	83.9 ± 3.3	2.6×10^5	1.5
Heart	MDM-R	170	$\sigma = 120, \mu = 0.017$	84.2 ± 2.7	4.0×10^4	1.0
Heart	MDM-S	170	$\sigma = 120, \lambda = 0.76$	83.9 ± 2.7	2.7×10^4	1.0
Solar Flare	SMO	666	$\sigma = 30, C = 1\ 000$	67.6 ± 1.8	1.0×10^7	30.4
Solar Flare	MDM-R	666	$\sigma = 30, \mu = 0.0039$	67.6 ± 1.8	2.9×10^6	18.7
Solar Flare	MDM-S	666	$\sigma = 30, \lambda = 0.48$	67.5 ± 1.8	1.7×10^6	10.9
German	SMO	700	$\sigma = 10, C = 3\ 100$	76.1 ± 2.2	9.0×10^6	31.0
German	MDM-R	700	$\sigma = 10, \mu = 0.0052$	75.5 ± 0.5	2.1×10^6	3.1
German	MDM-S	700	$\sigma = 10, \lambda = 0.78$	75.9 ± 0.5	1.4×10^6	2.0
Cancer	SMO	500	$\sigma = 100, C = 1\ 000$	95.6 ± 2.0	8.3×10^4	4.4
Cancer	MDM-R	500	$\sigma = 100, \mu = 0.02$	95.3 ± 2.0	3.7×10^4	3.1
Cancer	MDM-S	500	$\sigma = 100, \lambda = 0.05$	95.4 ± 2.0	2.1×10^4	1.9
Adult	SMO	16 000	$\sigma = 1\ 000, C = 1\ 000$	83.3 ± 1.5	2.3×10^8	573.0
Adult	MDM-R	16 000	$\sigma = 1\ 000, \mu = 0.02$	83.2 ± 2.0	1.1×10^7	247.0
Adult	MDM-S	16 000	$\sigma = 1\ 000, \lambda = 0.35$	83.3 ± 2.0	0.7×10^7	168.0

从表 1 可以看出, 与其他的两个算法相比: 1) 本文算法在取得几乎相同的分类正确率的前提下, 核计算的次数大大减少, 运算速度有了明显提高. 分析其原因, 与压缩凸壳相比, 尺度化凸壳与原凸壳有相同数量的顶点数, 我们可以由定义很容易地算出其顶点并存储, 从而加快了寻找最近点对的速度. 2) 本文方法求得的分类器是尺度化凸壳间的最大间隔分类器, 而较大的间隔通常意味着较好的泛化能力. 3) 由于尺度化凸壳与原凸壳形状尺度不变, 尺度化凸壳的形状不会随着压缩因子 λ 的变化而发生变化, 因此更能保持样本的分布特征, 得到的分类超平面也比较稳定. 4) 由于顶点是中心和原训练样本点的组合, 本文算法考虑到了样本的总体信息, 克服了传统的 SVM 算法只依靠少数几个训练样本的缺点, 但同时又能保持解系数的稀疏性. 与 SMO 方法相比, 尺度化凸壳法有明显的几何意义, 即寻找两个凸壳间的最近点对, 任务更简单, 因而收敛速度更快, 更容易找到分类超平面.

5 结论

本文提出了一种新的构造 SVM 分类器的方法——尺度化凸壳法, 该方法可以把线性不可分问题转化为求解几何最近点对问题, 这为求解不可分问题提供了一种新的思路. 同时, 该方法有明确的几何意义, 求解过程更直观、简洁. 最后, 本文给出了求解非线性问题的简化算法. 实验表明, 本文算法能大大缩短 SVM 分类器的训练时间, 特别适用于样本较多的大规模分类问题. 此外, 通过赋予尺度化凸壳以不同的压缩因子, 该方法可以用来解决代价敏感问题.

References

- 1 Vapnik V N [Author], Xu Jian-Hua, Zhang Xue-Gong [Translator]. *Statistical Learning Theory*. Beijing: Publishing House of Electronics Industry, 1998 (Vapnik V N [著], 许建华, 张学工 [译]. 统计学习理论. 北京: 电子工业出版社, 2004)
- 2 Cortes C, Vapnik V N. Support vector networks. *Machine Learning*, 1995, **20**(3): 273–297
- 3 Bennett K P, Bredensteiner E J. Geometry in learning. *Geometry at Work*. Lincoln: Mathematical Association of America, 1998. 132–145
- 4 Bennett K P, Bredensteiner E J. Duality and geometry in SVM classifiers. In: *Proceedings of the 17th International Conference on Machine Learning*. San Francisco, USA: Morgan Kaufmann Publishers, 2000. 57–64
- 5 Crisp D J, Burges C J C. A geometric interpretation of v -SVM classifiers. *Advances in Neural Information Processing System*. Denver: MIT Press, 1999. 244–250
- 6 Keerthi S S, Shevade S K, Bhattacharyya C, Murthy K P K. A fast iterative nearest point algorithm for support vector machine classifier design. *IEEE Transactions on Neural Network*, 2000, **11**(1): 124–136
- 7 Franc V, Hlavac V. An iterative algorithm learning the maximal margin classifier. *Pattern Recognition*, 2003, **36**(9): 1985–1996
- 8 Mavroforakis M E, Sdralis M, Theodoridis S. A geometric nearest point algorithm for the efficient solution of the SVM classification task. *IEEE Transactions on Neural Network*, 2007, **18**(5): 1545–1549
- 9 Mavroforakis M E, Theodoridis S. A geometric approach to support vector machine (SVM) classification. *IEEE Transactions on Neural Network*, 2006, **17**(3): 671–682
- 10 Mavroforakis M E, Sdralis M, Theodoridis S. A novel SVM geometric algorithm based on reduced convex hulls. In: *Proceedings of the 18th International Conference on Pattern Recognition*. Hong Kong, China: IEEE, 2006. 564–568
- 11 Tao Q, Wu G W, Wang J. A general soft method for learning SVM classifiers with L_1 -norm penalty. *Pattern Recognition*, 2008, **41**(3): 939–948
- 12 Tian Li, Liu Zhen-Bing, Liu Xiao-Mao. Modified linear SVM. *Computer Engineering and Applications*, 2007, **43**(20): 173–176 (田立, 刘振丙, 刘小茂. 一种改进的线性 SVM. 计算机工程与应用, 2007, **43**(20): 173–176)
- 13 Chapelle O, Sindhwani V, Keerthi S. Optimization techniques for semi-supervised support vector machines. *The Journal of Machine Learning Research*, 2008, **9**: 203–233
- 14 Keerthi S, Shevade S, Bhattacharyya C, Murthy K. Improvements to Platt's SMO Algorithm for SVM Classifier Design, Technical Report CD-99-03, Department of Mechanical and Production Engineering, National University of Singapore, Singapore, 1999

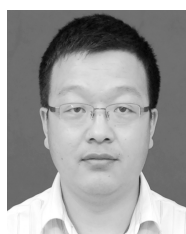


刘振丙 华中科技大学博士研究生. 2006 年获华中科技大学硕士学位. 主要研究方向为机器学习和信号处理. 本文通信作者.

E-mail: zbluhust@gmail.com

(LIU Zhen-Bing Ph.D. candidate at Huazhong University of Science and Technology (HUST). He received his

master degree from HUST in 2006. His research interest covers machine learning and signal processing. Corresponding author of this paper.)



陈忠 华中科技大学副教授. 2008 年华中科技大学博士后出站. 主要研究方向为遥感图像处理和模式分类.

E-mail: henpacked@163.com

(CHEN Zhong Associate professor at HUST. He completed his postdoctor research from HUST in 2008. His research interest covers remote sensing image processing and pattern classification.)



刘建国 华中科技大学教授. 1996 年在香港大学获得博士学位. 主要研究方向为信号处理的快速算法、图像处理、模式识别. E-mail: jgliu@ieee.org

(LIU Jian-Guo Professor at HUST. He received his Ph.D. degree from The University of Hong Kong in 1996. His research interest covers fast algorithms in signal processing, image processing, and pattern recognition.)