

平均邻近间隔支撑向量机

王晓明, 王士同

(江南大学 信息工程学院, 江苏 无锡 214122)

摘要:近年来,支撑向量机(SVMs)作为一种学习方法已经在机器学习和模式识别的研究领域中得到了成功的运用,然而传统SVMs缺少考虑数据的局部信息.文中将邻近间隔的基本思想引入到SVMs中,提出了平均邻近间隔支撑向量机(ANMSVMs).ANMSVMs继承了传统SVMs的优点,同时又充分利用了数据的局部信息,从而实现了泛化能力的进一步提高.人造数据和真实数据集的实验结果验证了该方法有效性,并且相对于传统SVMs体现出了更强的泛化能力.

关键词:监督学习;支撑向量机;邻近间隔

中图分类号: TP181 **文献标识码:** A **文章编号:** 1673-4785(2010)04-0313-07

Using average neighborhood margin with support vector

WANF Xiao-ming, WANG Shi-tong

(School of Information, Jiangnan University, Wuxi 214122, China)

Abstract: The support vector machine (SVM) is a learning method used successfully in the fields of machine learning and pattern recognition. However, traditional SVMs have not taken local information, or the environment of the data, into full consideration. To provide this context, an average neighborhood margin support vector machine (ANMSVM) was formulated by introducing the basic theories of neighborhood margins into SVMs. This method inherits the characteristics of traditional SVMs, yet also fully considers the local information surrounding the data, and thus shows better learning performance. Experimental results on artificial and real datasets indicated the effectiveness of an ANMSVM.

Keywords: supervised learning; support vector machine; neighborhood margin

近年来,支撑向量机(support vector machines, SVMs)^[1-2]已经在机器学习和模式识别领域的研究中引起了广泛的关注,成为了广大学者的研究热点之一.由于SVMs具有结构简单、泛化能力(Generalization)强等优点,并且有着坚实的统计学习理论基础,能处理线性和非线性的情况,目标优化问题可以通过偶形式化为一个较易处理的优化问题,所以在众多的研究和应用(如人脸识别^[3]、图像检索^[4]、语音识别^[5]等)都中引起了广泛的关注,并取得了显著的成果.

然而,传统的SVMs建立决策超平面是通过使

样本到超平面的间隔距离最大化来实现的,仅仅考虑了每类数据边界上的样本点,而缺少考虑样本数据的局部信息,从而在一定程度上制约了其泛化能力的进一步提高.但是在一些特征排序和特征提取的算法中,数据的局部信息得到了充分的利用.这些算法的共同点就是通过采用邻近间隔(neighborhood margin)的基本思想来充分利用数据的局部特征.在这些算法中,邻近间隔的基本思想主要体现在对每一个样本点都首先建立2个邻域集合:一个是由样本点的同类数据中的 k 个最邻近样本组成的集合,另一个则是由样本点的异类数据中的 K 个最邻近样本组成的集合.然后再利用所有样本点的这2个邻域集合来构建特征排序或特征提取的目标函数.在这些方法中,典型的是RELIEF^[6]、I-RELIEF(Iter-

收稿日期:2009-09-13.

基金项目:国家自然科学基金资助项目(60704047);国家“863”计划资助项目(2007AA1Z158、2006AA10Z313).

通信作者:王士同. E-mail: wxwangst@yahoo.com.cn.

active RELIEF)^[7]、SNMMC (stepwise nonparametric margin maximum criterion)^[8] 和 ANMM (average neighborhood margin maximization)^[9] 算法. 这些算法充分利用了数据的局部性和整体性, 简单而且有效.

在本文中, 为了克服传统的 SVMs 缺少考虑数据局部信息的缺点, 把邻近间隔的基本思想引入到 SVMs 中, 提出了平均邻近间隔支撑向量机 (average neighborhood margin support vector machines, ANMSVMs). 该方法显著的特征是即保留传统 SVMs 的优点, 同时又充分考虑了数据的局部特征, 而且还可以转化为标准的 SVMs 优化问题, 从而能够利用 SVMs 现有的各种软件来解决 ANMSVMs 问题. 首先利用每一个训练样本点的 2 个邻域定义了平均邻近间隔向量, 然后利用所有样本点的平均邻近间隔向量建立了 ANMSVMs 的优化问题. 从距离定义的方式讲, ANMSVMs 采用了马氏距离, 而 SVMs 则直接采用了欧式距离. 文中分析该方法和 SVMs 的联系, 也讨论其非线性化的问题. 实验结果验证该方法的有效性, 并且相对于传统的 SVMs, 其表现出了更强的泛化能力.

1 基于邻近间隔的特征排序和特征提取

文中假定有一个由 N 个样本组成的数据集 $X = [x_1 \cdots x_N]$, $\forall x_i \in \mathbf{R}^M$, $X \in \mathbf{R}^{M \times N}$, 所有样本分属于 2 个不同的类别, 用 X_L 表示第 L ($L = 1, 2$) 类数据子集, 其中共含有 N_L 个样本, y_i 表示样本 x_i 的类别标号, $\forall y_i \in \{1, -1\}$.

1.1 基于邻近间隔的特征排序

RELIEF^[6] 算法是一种经典的特征排序算法. 该算法对每一维特征的鉴别能力进行估计时, 首先找到每一个样本 x 的 2 个邻域集合: 一个是样本点 x 同类 (Hit) 数据集中距离样本 x 最近的样本点 $NH(x)$, 另一个是样本点 x 异类 (Miss) 样本集中距离样本 x 最近的样本点 $NM(x)$; 然后对每一维特征的权重采用了如下的方式进行更新:

$$w_i = w_i + |x^i - NM(x)| - |x^i - NH(x)|. \quad (1)$$

式中: x^i 表示样本 x 点的第 i 维特征, w_i 表示第 i 维特征的权重, $|\cdot|$ 表示求绝对值, 最后利用 w_i 来对特征进行排序.

I-RELIEF^[7] 算法是 RELIEF 算法的改进版本. 在 I-RELIEF 算法中, 首先对于样本点 x_i 定义了下

面的向量:

$$z_i = |x_i - NM(x_i)| - |x_i - NH(x_i)|. \quad (2)$$

式中: $NH(x_i)$ 表示样本点 x_i 同类中最近的样本点, $NM(x_i)$ 表示异类中最近的样本点. 然后 I-RELIEF 算法定义如下的优化问题:

$$\max_w w^T \sum_i z_i, \quad \text{s. t. } w^T w = 1, w \geq 0. \quad (3)$$

I-RELIEF 算法采用了迭代的方法来求解上面优化问题中的 w , 最后也根据 w 来排序特征的重要性. 在文献[10]中, 作者把 I-RELIEF 推广到了特征提取方法中.

1.2 基于邻近间隔的特征提取算法

在特征提取方法中, SNMMC^[8] 算法也采用了邻近间隔的基本思想. 当提取一维特征时, SNMMC 算法定义了下面的优化目标函数:

$$\max_w w^T \left(\sum_{i=1}^N \omega_i (\|\delta_i^E\|^2 - \|\delta_i^I\|^2) \right) w, \quad \text{s. t. } w^T w = 1. \quad (4)$$

式中: ω_i 表示的是样本权重, $\delta_i^E = x_i - x^E$ 表示样本点 x_i 与其异类样本集中最近样本点的差值向量, $\delta_i^I = x_i - x^I$ 则表示样本点 x_i 与其同类样本集中最近样本点的差值向量, $\|\cdot\|$ 表示向量的 L_2 范数. 值得注意的是, 在 SNMMC 算法中采用了差值向量的 L_2 范数, 而在 RELIEF 算法中则采用了差值向量的绝对值.

由于 SNMMC 算法易受例外样本点的影响, 在文[9]中作者针对这个问题提出了 ANMM 算法. 在 ANMM 算法中, 首先定义了:

$$\gamma_i = \sum_{k: x_k \in N_i^e} \frac{\|x_i - x_k\|^2}{N_i^e} - \sum_{k: x_k \in N_i^o} \frac{\|x_i - x_k\|^2}{N_i^o}. \quad (5)$$

式中: N_i^e 表示样本点 x_i 异类的近邻样本组成的样本集, 该样本集的样本数为 N_i^e ; N_i^o 表示样本点 x_i 同类的近邻样本组成的样本集, 该样本集的样本数为 N_i^o . 然后 ANMM 算法定义了下面的优化问题:

$$\max_w w^T \left(\sum_i \gamma_i \right) w, \quad \text{s. t. } w^T w = 1. \quad (6)$$

通过上面的介绍可以看出, 这些算法共同的一点就是利用了样本点的同类近邻样本和异类近邻样本, 依据这 2 个近邻集之间的差异来判断特征的鉴别能力, 充分考虑了数据的局部特征. 不同的是在特征排序中采用了差值向量的绝对值, 在特征提取中

则采用了差值向量的 L_2 范数.

2 线性平均邻近间隔支撑向量机

2.1 线性平均邻近间隔支撑向量机

首先基于样本的邻近间隔思想,对于样本点 x_i 定义如下的向量:

$$z_i = \sum_{k: x_k \in N^H(x_i)} \frac{|x_i - x_k|}{K} - \sum_{k: x_k \in N^M(x_i)} \frac{|x_i - x_k|}{K}. \quad (7)$$

式中: $N^H(x_i)$ 表示样本点 x_i 的同类样本集中 K 个最近样本点组成的样本集, $N^M(x_i)$ 则表示 x_i 的异类样本集中 K 个最近样本点组成的样本集, K 表示近邻数. 注意,这里的向量 z_i 与前面 RELIEF 和 ANMM 算法中的定义不同. 它是同类邻域中的 K 个最近样本点的平均差向量减去了异类邻域中的 K 个最近样本点的平均差向量,称其为样本点 x_i 的平均邻近间隔向量. 利用所有样本点 x_i 的平均邻近间隔向量 z_i , 定义线性可分平均邻近间隔支撑向量机 (ANMSVMs) 的优化问题为

$$\begin{aligned} \min_{w, b} \quad & \sum_{i=1}^N w^T z_i z_i^T w, \\ \text{s. t.} \quad & w^T x_i + b \geq 1, i = 1, 2, \dots, N, \end{aligned} \quad (8)$$

令

$$Z = \sum_{i=1}^N z_i z_i^T. \quad (9)$$

注意,矩阵 Z 是对称的,并且具有半正定性. 为了统一数据线性可分和不可分的情况,通过引入了松弛变量 ξ_i (当 ξ_i 都为 0 时对应于线性可分的情况),则上面的优化问题(8)可以改写为

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} w^T Z w + C \sum_{i=1}^N \xi_i, \\ \text{s. t.} \quad & y_i (w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, \dots, N. \end{aligned} \quad (10)$$

式中: C 为惩罚参数. 假定 $\{w^*, b^*, \xi^*\}$ 是上述优化问题的最优解,则 ANMSVMs 建立的决策超平面为 $[w^*]^T x + b^* = 0$, 决策函数是:

$$f(x) = \text{sgn}([w^*]^T x + b^*). \quad (11)$$

图 1 给出了在一个人造数据情况下 ANMSVMs 和 SVMs 所对应的决策超平面. 从中可以看出,由于 ANMSVMs 算法考虑了数据的局部信息,其得到的决策面超平面充分反映数据的特点,然而 SVMs 的

决策超平面却不能反映数据的特点.

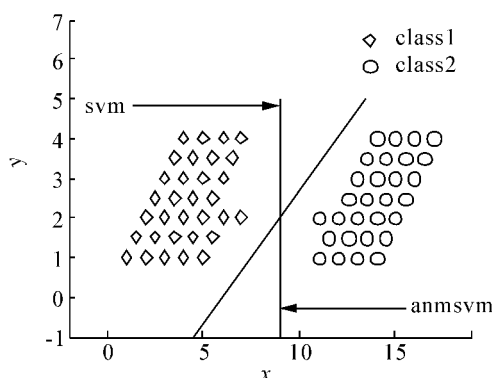


图 1 人造数据情况下 ANMSVMs 和 SVMs 所对应的决策超平面

Fig. 1 The decision superplanes constructed by ANMSVMs and SVMson the artificial dataset

类似于 SVMs, 可以把 (ANMSVMs) 的优化问题 (10) 转化为其对偶优化问题来求解. 利用拉格朗日法^[11], 假定矩阵 Z 是非奇异 (对于 Z 奇异的情况, 将在后面进行介绍), 可以得的优化问题 (10) 的对偶优化问题为

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \alpha^T H \alpha + e^T \alpha, \\ \text{s. t.} \quad & \sum_{i=1}^N \alpha_i z_i = 0, 0 \leq \alpha_i \leq C, i = 1, 2, \dots, N. \end{aligned} \quad (12)$$

式中: $H_{ij} = y_i y_j x_i^T Z^{-1} x_j$, $\alpha^* = [\alpha_1 \alpha_2 \dots \alpha_N]^T \in \mathbb{R}^N$. 假定式 (12) 的最优解为 $\alpha^* = [\alpha_1^* \alpha_2^* \dots \alpha_N^*]^T$, 利用 KKT 条件^[11]可以得到原始优化问题 (10) 的 w 的最优解为

$$w^* = Z^{-1} \sum_{i=1}^N y_i \alpha_i^* x_i. \quad (13)$$

对于 $i, 0 < \alpha_i < C$, 类似于 SVMs, 称 x_i 为支撑向量. 用 D_{SV} 来表示支撑向量集, N_{SV} 表示支撑向量个数. 此时, 根据 KKT 条件^[11]有 $y_i ([w^*]^T x_i + b^*) = 1$, 可以据此得到 b 的最优解. 但是为了使计算更可靠, 通常取所有支撑向量的均值, 即:

$$b^* = \frac{1}{N_{SV}} \sum_{x_i \in D_{SV}} (y_i - y_i \sum_{j=1}^N \alpha_j^* H_{ij}). \quad (14)$$

把最优解 $\{w^*, b^*\}$ 代入式 (11), 可以得到线性情况下 ANMSVMs 的决策函数为

$$f(x) = \text{sgn}(\sum_{i=1}^N z_i y_i^* (x_i^T Z^{-1} x) + b^*). \quad (15)$$

通过前面的讨论, 可以得到线性 ANMSVMs 算法.

算法 1: 线性 ANMSVMs

- 1) 利用式(7)和(9)建立矩阵 Z ;
- 2) 求解出优化问题(12), 得到其最优解为 $\alpha^* = [\alpha_1^* \alpha_2^* \cdots \alpha_N^*]^T$;
- 3) 利用式(12)和(14)得到;
- 4) 利用上一步得到的最优解 $\{w^*, b^*\}$ 并根据式(15)建立 ANMSVMs 的决策函数.

2.2 小样本的问题

在前面都假定了矩阵 Z (其是 $M \times M$ 矩阵) 是非奇异的, 然而在实际应用中则可能出现奇异的情况, 尤其是在小样本数据中. 在这种情况下可以采用 Z 的伪逆^[12]来代替其逆矩阵. 为了计算矩阵 Z 的伪逆, 首先对矩阵进行奇异值分解

$$Z = P\Lambda P^T. \quad (16)$$

式中: $\Lambda = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_m)$ 包含了矩阵 Z 的大于 0 的特征值, m 是矩阵 Z 的秩, P 为矩阵 Z 的大于 0 的特征值所对应的特征向量组成矩阵. 从而, 矩阵 Z 的伪逆是

$$Z^+ = P\Lambda^{-1}P^T. \quad (17)$$

3 非线性平均邻近间隔支撑向量机

由于矩阵 Z 是由每一个样本点 x^i 的平均邻近间隔向量 z_i 构成, 然而邻近间隔向量 z_i 中却含有对向量的元素求绝对值对于, 因而 ANMSVMs 不能直接非线性化. 这里将首先对非线性 SVMs 进行分析, 然后依据相同的原理引出了非线性的 ANMSVMs.

首先来分析一下线性 SVMs 和 PCA^[13] 相结合的一个性质. 在 PCA 算法中, 定义总体散度矩阵: $S_i = \sum_{i=1}^N (x_i - m)(x_i - m)^T$, 这里 $m = \frac{1}{N} \sum_{i=1}^N x_i$. 文献[14]指出, 对于 S_i 的特征向量, 用 Ψ 表示其非零特征值对应的特征向量张成的空间, Π 表示零特征值对应的特征向量所张成的空间, 则空间 Ψ 和空间 Π 是互补空间并一起组成了 S_i 的特征向量空间. 对于 SVMs, 有以下定理:

引理 1: 对于任意 $w \in R^M$, w 都可以表示成 $w = v + \gamma$, 这里 $v \in \Psi$, $\gamma \in \Pi$. 从而定义的 SVMs 的优化问题

$$\min_{w, b} \frac{1}{2} w^T w, \quad \text{s. t. } y_i (w^T x_i + b) \geq 1, i = 1, 2, \cdots, N. \quad (18)$$

可以等价表示为

$$\min_{v, b} \frac{1}{2} v^T v,$$

$$\text{s. t. } y_i (v^T x_i + b) \geq 1, i = 1, 2, \cdots, N. \quad (19)$$

证明见附录.

由于 Ψ 空间是 S_i 非零特征值对应的特征向量张成的空间, 假定 S_i 非零特征值有 m 个, 则 Ψ 空间就是 m 维, 根据线性代数理论, 其必然和 m 维的 R^m 空间同构. 所以存在标准正交的 $P \in R^{M \times m}$ 使得 $v = P\eta$ 成立, 这里 $v \in \Psi$, $\eta \in R^m$. 一般情况下矩阵 P 可以用 S_i 非零特征值所对应的特征向量组成, 这样式(19)就转换为

$$\min_{\mu, b} \frac{1}{2} \eta^T \eta,$$

$$\text{s. t. } y_i (\eta^T \tilde{x}_i + b) \geq 1, i = 1, 2, \cdots, N \quad (20)$$

式中: $\eta \in R^m$, $\tilde{x}_i = P^T x_i$. 根据引理 1 和式(20), 可以认为在空间 R^M 中应用 SVMs 等价于在与 Ψ 空间同构的 R^m 空间中应用 SVMs. 因此, 可以得到下面的引理.

引理 2: 非线性 SVMs KPCA 线性 SVMs.

这里仅对引理 2 进行简单的分析. 由于非线性的 SVMs 是首先采用核函数把样本映射到特征空间, 然后再在这个特征空间中应用线性 SVMs 算法; 然而由引理 1 可知, 在特征空间中应用 SVMs 等价于在与 Ψ 空间同构的欧氏空间中应用 SVMs. 这个过程中先采用核函数把样本映射到特征空间, 然后再采用 PCA 算法把特征空间中的数据投影到与 Ψ 空间同构的欧氏空间中, 实际上相当于直接采用 KPCA^[15] 算法对原始数据进行投影变换. 所以, 非线性的 SVMs 等价与先采用 KPCA 把原始数据投影到一个新的数据空间, 然后再在这个空间中应用线性 SVMs.

所以, 对于非线性的 ANMSVMs, 也可以采用相同的方式, 即首先采用 KPCA 把数据投影到一个新的数据空间(称其为变换数据空间), 然后再在这个数据空间中应用线性 ANMSVMs 算法. 因此对于非线性的 ANMSVMs, 可以在变换数据空间中建立如下的优化问题:

$$\min_{w, b, \xi} \frac{1}{2} w^T \tilde{Z} w + C \sum_{i=1}^N \xi_i,$$

$$\text{s. t. } y_i (w^T \tilde{x}_i + b) \geq 1 - \xi, \xi_i \geq 0, i = 1, 2, \cdots, N.$$

(21)

式中: \tilde{x}_i 为在采用 KPCA 投影后的变换数据空间中样本点 x_i 所对应的新样本, \tilde{Z} 为采用 KPCA 投影后的投影数据空间中根据式(7)和(9)建立的矩阵. 假定优化问题(21)的最优解为 $\{w^*, b^*, \xi^*\}$, 则可以得到非线性的 ANMSVMs 的决策函数为

$$f(x) = \text{sgn}([w^*]^T \tilde{x} + b^*). \quad (22)$$

这里, \tilde{x} 表示未知类别信息的样本点 x 采用 KPCA 投影后在变换数据空间中所对应的新样本.

由上,可以得到非线性 ANMSVMs 算法:

算法2:非线性 ANMSVMs

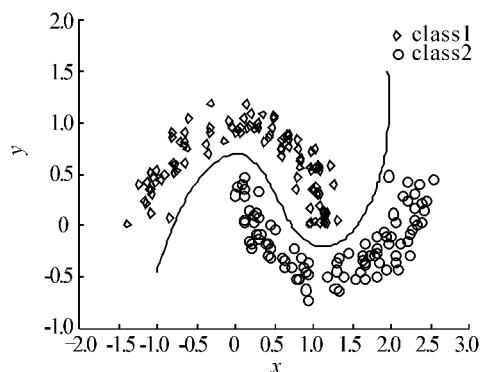
- 1) 利用 KPCA 把数据投影到一个新的数据空间中, 利用式(7)和(9)建立矩阵;
- 2) 求解出优化问题(21), 得到其最优解为;
- 3) 利用上一步得到的最优解 并根据式(22) 建立非线性 ANMSVMs 的决策函数.

4 实验结果及分析

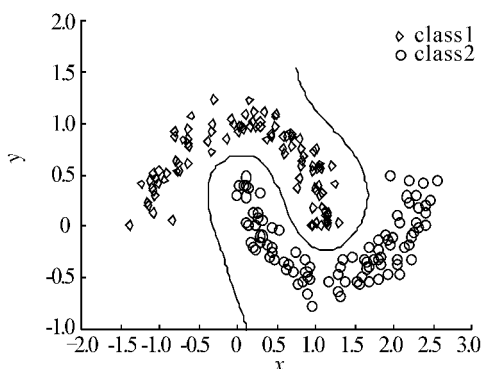
为了验证本文提出的 ANMSVMs 算法的可行性以及泛化能力,在这一部分内容中将进行实验研究,把它和传统的 SVMs 进行了比较. 实验分为2部分. 在第一部分实验中通过人工拟合数据来演示了 ANMSVMs 和 SVMs 在非线情况下构造的决策超平面的不同;在第2部分实验中,采用了10个UCI^[16]数据来测试和比较了 ANMSVMs 和 SVMs 在线性情况和非线性情况下的泛化能力. 在实验中,对于 SVMs 采用了 LIBSVM^[17].

4.1 拟合数据

前面图1演示了 ANMSVMs 算法和 SVMs 算法在线性情况下构造数据决策超平面时的不同. 在这一个实验中,将通过2个拟合数据进一步揭示在非线情况下它们之间的区别. 图2采用的数据是 two-moons 数据集. 这里采用了高斯核,核参数为5. 图2的(a)是直接采用了非线性 SVMs 结果,(b)给出了非线性的 ANMSVMs 结果,也即是采用了 KPCA + 线性 ANMSVMs 的方法. 从这些实验结果中可以看出,文中提供的 ANMSVMs 非线性化方式是有效的,并且由于 ANMSVMs 算法进一步利用数据的局部信息,相对于 SVMs,其构造出的决策超平面更加反映了数据的特征.



(a) 非线性 SVMs



(b) 非线性 ANMSVMs

图2 two-moons 数据中采用高斯核的情况下 ANMSVMs 和 SVMs 所对应的决策超平面

Fig. 2 The decision superplanes constructed by ANMSVMs and SVMs in the rbf. kernel case on the two-moons dataset

4.2 UCI 数据

在这一部分的实验中采用了10个UCI的数据来进一步测试 ANMSVMs 的泛化能力. 表4给出这些数据的基本信息. 其中的数据既有2类的情况,也有多类的情况. 对于多类的情况,采用了一对一的方式^[18]. 在进行实验前,所有数据都把其每一维特征处理为均值为0,方差为1. 实验中,采用了5重交叉测试(Cross alidation)^[18]精度作为泛化性能的度量标准. 5重交叉测试就是把数据样本平均划分为5部分,进行5次训练测试过程,然后计算出这5次测试的平均测试精度. 在第1次训练测试过程中,把数据的第1部分作为测试数据集,剩下的数据集作为训练数据集;第2次训练测试过程是把数据的第2部分数据作为测试数据集,其它的数据集作为训练数据集. 依照这样的过程,直到所有5次训练测试过程做完. 在本文中,就是把这5次训练测试的平均测试精度作为泛化性能的评估准则.

实验中在 $\{0.001, 0.01, 0.1, 1, 10, 100\}$ 范围内变化惩罚系数 C , 对于 ANMSVMs 算法的近邻参数 K 也在 2 到 20 范围变化. 表 5 给出了 SVMs 和 ANMSVMs 在线性和非线性情况下的测试结果. 这里给出的是这些参数下最好的测试结果. 对于非线性采用了 2 次多项式核, 即核函数为 $(\mathbf{u}^T \mathbf{v} + 1)^2$. 对于非线性 SVMs 算法, 直接采用了非线性的方法; 然而由于对于 ANMSVMs 算法, 正如前面文中的第 4 部分内容所述, 由于其不能直接非线性化, 采用了 KPCA 加线性算法的方法.

表 1 实验中 UCI 数据基本信息描述

Table 1 Characteristics of the selected datasets from the UCI repository

Dataset	No. of Patterns	No. of Features	No. of Classes
Heart	270	13	2
Beast	699	9	2
Pima	768	8	2
Wdbc	569	30	2
Ionosphere	351	34	2
Newthyroid	215	5	3
Wine	178	13	3
Iris	150	4	3
Vehicle	846	8	4
Glass	214	9	6

表 2 在 UCI 数据上的 5 重交叉测试的平均精度和方差

Table 2 Mean accuracy and standard deviation of the 5-fold cross validation on the selected UCI datasets / %

Dataset	Linear		Nonlinear (Poly Kernel)	
	SVMs	ANMSVMs	SVMs	ANMSVMs
Heart	84.8(0.04)	85.5(0.02)	68.8(0.02)	71.8(0.05)
Breast	96.8(0.02)	96.7(0.02)	93.9(0.03)	94.2(0.02)
Pima	77.4(0.01)	77.3(0.01)	69.5(0.01)	77.2(0.02)
Wdbc	97.5(0.00)	96.1(0.01)	88.4(0.01)	94.3(0.01)
Ionosphere	87.1(0.03)	88.0(0.05)	88.8(0.03)	91.1(0.00)
Newthyroid	96.2(0.02)	98.6(0.01)	92.5(0.06)	95.8(0.03)
Wine	93.7(0.03)	93.8(0.02)	86.0(0.06)	93.8(0.03)
Iris	96.6(0.04)	98.0(0.02)	90.0(0.00)	98.0(0.01)
Vehicle	79.7(0.01)	80.4(0.01)	79.4(0.02)	79.6(0.02)
Glass	60.2(0.06)	59.3(0.05)	60.7(0.08)	64.0(0.07)

从这些实验结果可以看出, 在多数情况下 AN-

MSVMs 都表现出了比传统 SVMs 算法更好的泛化能力了, 尤其是多项式核的情况. 这说明, 充分利用数据的局部信息, 有利于提高分类的精度. 而且还可以看出, ANMSVMs 的最优惩罚系数 C 相对 SVMs 而言, 总体上趋于更大.

5 结束语

在本文中, 结合邻近间隔的基本思想提出了 ANMSVMs 算法. 相对于传统的 SVMs, ANMSVMs 进一步利用了数据的局部特征, 表现出了更强的泛化能力. ANMSVMs 可以转换为标准的 SVMs, 从而能够利用现有的各种 SVMs 软件来求解 ANMSVMs. 但是在非线性情况下, ANMSVMs 不能直接非线性化. 文中采用了 KPCA 加线性算法来处理这种情况, 实验也验证这种方法的有效性. 但是, 是否可以用更好、更有效的方法来处理这种情况, 这也是将来进一步研究的方向之一.

附录: 引理 1 的证明

证明: 由于 $\gamma \in \mathbf{H}$, 其对应的特征值为 0, 所以 $\gamma^T \mathbf{S}_i \gamma = 0$, 也就是说所有样本点向 γ 方向投影后总体散度为 0. 因而当 $\mathbf{x}_i \neq \mathbf{x}_j$ 时有 $\gamma^T \mathbf{x}_j$, 这样所有的样本点向 γ 方向投影后是同一个点, 为一个常量. 令 $\gamma^T \mathbf{x}_i = r$, 则有:

$$\mathbf{w}^T \mathbf{x}_i = (\mathbf{v} + \gamma)^T \mathbf{x}_i = \mathbf{v}^T \mathbf{x}_i + r. \quad (23)$$

所以有:

$$\mathbf{w}^T \mathbf{w} = (\mathbf{v} + \gamma)^T (\mathbf{v} + \gamma) = \mathbf{v}^T \mathbf{v} + \gamma^T \gamma. \quad (24)$$

由 (18) 的 KKT 条件 $\mathbf{a}^T \mathbf{z} = \sum_{i=1}^N \alpha_i z_i = 0$, 所以有 $\sum_{i=1}^N \alpha_i z_i r = r \sum_{i=1}^N \alpha_i z_i = 0$. 再结合式 (23) 和 (24), 则对于优化问题 (18) 的拉格朗日式子有:

$$\begin{aligned} L(\mathbf{w}, \mathbf{b}, \mathbf{a}) = & \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N \alpha_i [y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1] = \\ & \frac{1}{2} \mathbf{v}^T \mathbf{v} + \frac{1}{2} \gamma^T \gamma - \sum_{i=1}^N \alpha_i [y_i (\mathbf{v}^T \mathbf{x}_i + b) - 1]. \end{aligned} \quad (25)$$

注意, 则 \mathbf{w} 的最优解可以表示成 $\mathbf{w}^* = \mathbf{v}^* + \gamma^*$. 由式 (25), 有:

$$\frac{\partial L}{\partial \mathbf{v}} \Big|_{\mathbf{v}=\mathbf{v}^*} = 0 \Leftrightarrow \mathbf{Z} \mathbf{v}^* - \sum_{i=1}^N \alpha_i z_i \mathbf{x}_i = 0,$$

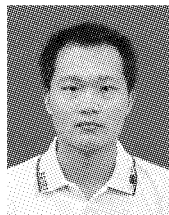
$$\frac{\partial L}{\partial \gamma^*} = 0 \Leftrightarrow \gamma^* = 0. \quad (26)$$

注意,式(26)与式(19)所提出的优化问题的拉格朗日式子是一样的,假定 $\nu^* \in \Psi$ 是它的最优解,又由于 $\gamma^* = 0$,则 w 的最优解可以表示成 $w^* = \nu^*$. 证毕.

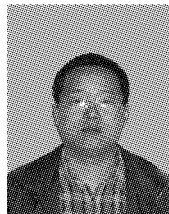
参考文献:

- [1] VAPNIK V. Statistical learning theory [M]. New York: Wiley, 1998:13-18.
- [2] SCHOLKOPF B, SMOLA A. Learning with kernels [M]. Cambridge, MA: MIT Press, 2002:25-32.
- [3] ER M J, CHEN W, WU S. High-speed face recognition based on discrete cosine transform and RBF neural networks [J]. IEEE Transactions on Neural Networks, 2005, 16 (3): 679-691.
- [4] TONG S, CHANG E. Support vector machine active learning for image retrieval [C]//Proceedings of ACM International Conference on Multimedia, 2001:107-118.
- [5] GANAPATHIRAJU A, HAMAKER J E, PICONE J. Applications of support vector machines to speech recognition [J]. IEEE Transactions on Signal Processing, 2004, 52 (8): 2348-2355.
- [6] KIRA K, RENDELL L A. A practical approach to feature selection [C]//Proc Ninth Int'l Conf Machine Learning. [s.l.], 1992: 249-256.
- [7] SUN Y J. Iterative RELIEF for feature weighting: algorithms, theories, and applications [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29 (6): 1035-1051.
- [8] QIU X, WU L. Face Recognition by stepwise nonparametric margin maximum criterion [C]//Proc ICCV. 2005: 1567-1572.
- [9] WANG F, ZHANG C S. Feature extraction by maximizing the average neighborhood margin [C]//Proc ICCV. Beijing, 2007:1-8.
- [10] SUN Y J, WU Dapeng: A RELIEF based feature extraction algorithm [C]//SDM. [s.l.], 2008: 188-195.
- [11] FLETCHER R. Practical methods of optimization [M]. 2nd ed. New York: Wiley, 1987:81-82.
- [12] LIU J, CHEN S C, TAN X Y. Efficient pseudo-inverse linear discriminant analysis and its nonlinear form for face recognition [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2008.
- [13] LIPOVETSKY S. PCA and SVD with nonnegative loadings. Pattern Recognition, 2009, 42(1): 68-76.
- [14] 杨键, 杨静宇, 叶晖, 等. Fisher 线性鉴别分析的理论研究及其应用 [J]. 自动化学报, 2003, 29(4): 482-493.
- [15] WANG L, PANG Y W, SHEN D Y, et al. An iterative algorithm for robust kernel principal component analysis [C]//2007 International Conference on Machine Learning and Cybernetics, 2007, (6): 3484-3489.
- [16] ASUNCION A, NEWMAN D. J. UCI machine learning repository [EB/OL]. <http://www.ics.uci.edu/~mllearn/MLRepository.html>, 2007.
- [17] CHANG C C, LIN C J. LIBSVM: a library for support vector machines [EB/OL]. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.
- [18] ABRIL L G, ANGULO C, VELASCO F, et al. A note on the bias in SVMs for multi-classification [J]. IEEE Transactions on Neural Networks, 2008, 19(4): 723-725.

作者简介:



王晓明,男,1977年生,博士研究生,主要研究方向为机器学习,模式识别,图像处理.



王士同,男,1964年生,教授,博导,主要研究方向为模式识别,模糊系统,生物信息学,先后十多次留学英国、日本和香港地区,在国内外重要杂志上发表学术论文数十篇.