

文章编号: 1001-0920(2006)01-0001-06

经典线性算法的非线性核形式

许建华¹, 张学工²

(1. 南京师范大学 数学与计算机学院, 南京 210097; 2. 清华大学 自动化系, 北京 100084)

摘要: 经典线性算法的非线性核形式是近10年发展起来的一类非线性机器学习技术。它们最显著的特点是利用满足Mercer条件的核函数巧妙地推导出线性算法的非线性形式, 并表述为与样本数目有关、与维数无关的优化问题。为了提高数值计算的稳定性、控制算法的推广能力以及改善迭代过程的收敛性, 部分算法还采用了正则化技术。在概述核思想与核函数、正则化技术的基础上, 系统地介绍了经典线性算法的非线性核形式, 同时分析它们的优缺点, 并讨论了进一步发展的方向。

关键词: 机器学习; 核函数; 核形式; 支持向量机

中图分类号: TP18 **文献标识码:** A

Nonlinear Kernel Forms of Classical Linear Algorithms

XU Jian-hua¹, ZHANG Xue-gong²

(1. School of Mathematical and Computer Sciences, Nanjing Normal University, Nanjing 210097, China; 2. Department of Automation, Tsinghua University, Beijing 100084, China. Correspondent: XU Jian-hua, E-mail: xujianhua@njnu.edu.cn)

Abstract: In machine learning the nonlinear kernel forms of classical linear algorithms are a class of nonlinear techniques developed in the last ten years. The most attractive idea is that by using kernel functions satisfying Mercer condition the classical linear algorithms are skillfully extended to construct their nonlinear kernel forms. These nonlinear algorithms are described by optimization problems that depend on the size of training sets rather than the dimension of sample vectors. In order to improve numerical stability, control generalization ability and improve convergence of iterative procedures, the regularization technique is utilized in some algorithms. On the basis of summarizing the kernel idea and kernel functions and regularization technique, the nonlinear kernel forms of linear algorithms are surveyed. Related properties are discussed and further research directions are pointed out.

Key words: Machine learning; Kernel function; Kernel forms; Support vector machine

1 引言

以支持向量机(SVM)分类算法^[1-4]为代表的核算法(机器)(Kernel Machines)是近10年来机器学习、模式识别以及神经网络领域最有影响力的成果之一。支持向量机分类算法包含如下4大技术: 1) 大间隔思想或正则化技术, 可以实现结构风险最小化原则, 以获得较好的推广能力; 2) 核思想与核函数, 利用满足Mercer条件的核函数实现线性算法的非线性化; 3) 凸二次规划, 使优化问题有唯一解; 4)

出现支持向量或稀疏性, 可以减少核形式判别式的计算量。这4大技术中最先得到研究人员青睐的是核思想与核函数, 这是一种非常有效的设计非线性算法的数学手段。随后, 很多研究人员利用核思想改造经典的线性算法, 得到相应的基于核函数的非线性形式(简称为核算法、核形式或核机器), 并形象地称这一过程为线性算法的核化。此外, 对于其中的一些核形式, 研究人员还利用正则化技术来提高核算法数值计算的稳定性、控制算法的推广能力、改善迭

收稿日期: 2004-12-15; 修回日期: 2005-04-18

基金项目: 国家自然科学基金项目(60275007); 江苏省自然科学基金项目(BK2004142)

作者简介: 许建华(1962—), 男, 浙江长兴人, 教授, 博士, 从事模式识别、机器学习等研究; 张学工(1965—), 山东青州人, 教授, 博士, 从事模式识别、机器学习等研究

代过程的收敛性 尽管目前公认的第1个核算法是60年代提出来的势函数法^[5,6],但正是支持向量机的成功促使了核算法得到进一步的发展与完善

本文首先简要地介绍支持向量机分类算法、核思想与核函数、正则化技术 在此基础上,详细介绍经典线性算法的非线性核形式,分析它们的优缺点,并指出进一步研究的方向以及应注意的一些问题

2 支持向量机分类算法

支持向量机分类算法是建立在统计学习理论中结构风险最小化准则上的实用算法^[1~4]. 设有一个两类的分类问题,对应的训练样本集为 $(x_1, y_1), \dots, (x_l, y_l)$,其中 $x_i \in R^n, y_i \in \{+1, -1\}$, l 为样本总数 又设有某一个非线性映射 $\Phi: R^n \rightarrow F$ (高维特征空间或Hilbert空间),将向量 x 非线性地映射成 F 中的向量 $\Phi(x)$,则支持向量机分类算法表述为

$$\begin{aligned} \min & \frac{1}{2} w^\Phi \cdot w^\Phi + C \sum_{i=1}^l \xi_i, \\ \text{s.t.} & y_i (w^\Phi \cdot \Phi(x_i) + \beta) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, \dots, l \end{aligned} \quad (1)$$

其中: \cdot 为内积运算; w^Φ, β 为 F 中线性判别函数的权向量与阈值; $\xi_i \geq 0$ 为松弛变量(大于零对应错分样本); C 为一个正常数,用于控制对错分样本的惩罚程度 对于线性可分问题,最小化目标函数的第1项等价于最大化两类样本之间的间隔 然而,对于不可分问题,这一项缺乏一个合理的物理解释,现在许多学者直接将它看作一个正则项^[4]. 通过定义优化问题(1)的Lagrange函数,再令Lagrange函数关于 w^Φ, β, ξ_i 的偏导数为零,可以将(1)转化成相应的对偶形式,即

$$\begin{aligned} \max & \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \Phi(x_i) \cdot \Phi(x_j), \\ \text{s.t.} & \sum_{i=1}^l \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i = 1, 2, \dots, l, \end{aligned} \quad (2)$$

其中 α_i 为Lagrange乘子. 利用Hilbert空间中的内积表达式 $\Phi(x_i) \cdot \Phi(x_j) = k(x_i, x_j)$,可将式(2)的目标函数变为一个凸二次函数

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j k(x_i, x_j),$$

其中: k 是满足Mercer定理的核函数;欲求变量个数等于训练样本数 并且称 $\alpha_i > 0$ 的样本为支持向量 根据Kuhn-Tucker条件,可计算出对应的 β 值 此时,非线性判别函数的形式为

$$f(x) = \sum_{\alpha_i > 0} \alpha_i y_i k(x, x_i) + \beta$$

关于支持向量机分类算法更详细的推导请参见文献

[1~4]以及这些书中的原始文献

3 核思想与核函数

根据支持向量机分类算法的推导过程,可将核思想概括为利用满足Mercer条件的核函数实现线性算法非线性化的数学手段,其基本思路是:1)假设有一个非线性变换,可将原始空间中的模式向量非线性地映射到某一个高维的特征空间中;2)在特征空间中设计一个线性算法或者利用一个已有线性算法,并且将它转化为仅涉及内积运算的优化问题;3)为了回避非线性映射的设计与具体形式,用满足Mercer条件的核函数来代替内积运算,从而推导出一个与样本数有关、与样本维数无关的优化问题;4)求解优化问题,得到原始空间中的一个非线性判别函数或者回归函数

从泛函分析角度来看,核函数是满足Mercer定理^[1,2]的双变量非线性对称函数(简称Mercer核),它可实现某一个高维特征空间中两向量之间的内积运算 从半群上的调和与分析角度来看,核函数是一个正定核^[4,7]. 给定一个双变量对称函数 $k: R^n \times R^n \rightarrow R$ 和样本向量 $x_1, \dots, x_l \in R^n$,则称元素为 $k_{ij} = k(x_i, x_j)$ 的 $l \times l$ 矩阵 K 是关于 x_1, \dots, x_l 的核矩阵 如果对所有 $l \in N$ (自然数域),对应的核矩阵均为正定阵,则称 k 为正定核 文献[4]证明了Mercer核也是正定的 目前常用的核函数是多项式、径向基核^[1,2].

对于核函数的研究,目前主要有两条思路 一条思路是从纯数学角度研究核函数 文献[3]指出了核之间运算的法则和构造核函数的若干法则 文献[8]收集到十几种核函数并依据统计学的观点加以分析归类,且给出核的谱表达式以便于新核的设计 核函数的理论研究成果,既有利于构造新的核函数,又有利于分析核算法的性质 但是大量核函数的出现,并没有给实际应用带来太多的好处,反而使大家感到有些迷茫 这样就有了第二条研究思路,即从实际问题出发直接构造特殊的核函数,而且有的核函数还不包含需要人为调节的参数 主要研究工作是针对生物信息学的非数值样本数据进行的,先后提出了局部性改善核^[9]、Fisher核^[10,11]、Top核^[12]等 文献[13]对特殊核函数给予了高度评价,认为有可能克服核算法在非数值样本应用上的不足

4 核算法中的正则化技术

正则化技术^[14]是为了专门处理不适定问题而提出来的数学方法 Hadamard意义下的适定问题必须满足3个条件:解是存在的、唯一的和稳定的 不能满足所有3个条件的问题称为不适定问题 所以,模式识别、神经网络中的很多问题也是不适定的 例如,反向传播训练算法、感知器算法等,因为它

们的解不是唯一的

为了处理不适定问题, 正则化方法定义一个光滑泛函或正则化泛函作为欲最小化的目标函数, 即

$$R_{\text{reg}}(w) = R_{\text{emp}}(w) + \lambda\Omega(w). \quad (3)$$

其中: w 表示与模型有关的参数向量 (有时含阈值项), $R_{\text{emp}}(w)$ 是衡量数据拟合程度的经验风险泛函, $\Omega(w)$ 是反映模型复杂性的正则项, λ 是正则化参数

对于式(3)中的第 1 项, 在模式识别、神经网络和函数回归(逼近)中, 常用的形式有: 错分样本数、误差平方和、误差绝对值之和^[4, 15, 16], 在支持向量机回归算法中增加了 ϵ 不敏感函数等^[1, 2].

对于式(3)中的第 2 项, 主要分为两种类型: 平方或 Gauss 型、绝对值或 Laplace 型^[4, 17], 它们分别对应于原始空间和特征空间中的欲求权向量或系数向量 α (有时含阈值项) 的 2 范数和 1 范数. 在核算法中, 正则项的作用是控制算法的推广能力、提高数值计算的稳定性、改善迭代算法的收敛性以及剪接网络拓扑结构. 所以, 正则化技术已成为核算法设计的一个重要数学工具

5 线性算法的非线性核形式

利用核思想与核函数、正则化技术, 研究人员推广了十几种常用的线性算法, 得到了对应的非线性核形式. 本文将详细地介绍与评述每一类核形式, 并指出它们随后的发展与应用

5.1 特征提取算法的核形式

线性特征提取是利用线性变换降低样本维数的数学手段^[15], 如主成分分析 (PCA)、K-L 变换等. Scholkopf 等人^[18, 19] 1997 年利用核思想改造了经典的线性主成分分析, 提出了基于核函数的非线性主成分分析, 即核主成分分析 (KPCA). 其目的是找到一组非线性的投影方向, 使得各个分量之间不相关. 它的计算过程是: 先求解由训练样本构成的核矩阵的本征值问题, 得到一组本征向量, 然后将样本向量与每一个本征向量作运算, 得到一个新的特征分量. 值得注意的是, 在样本数大于维数时, KPCA 可实现特征的扩张, 即能够得到比原始特征数目更多的特征. 经典的主成分分析是一个线性的坐标旋转变换, 所以它可以进行精确重建, 即反变换到原来的特征. 但是, 核主成分分析是一种非线性变换, 它无法实现精确的重建, 只能进行近似的迭代重建^[20].

Fisher 线性判别分析 (FLD)^[15, 16] 通过找到一个最有利于分类的线性投影方向, 将两类的高维问题降低到一维问题来处理. 1999 年 Mika 等提出了基于核函数的非线性形式, 称为核 Fisher 判别分析 (KFD)^[21], 试图找到一个最佳的非线性投影方向, 使得同类样本尽可能地聚集在一起而两类样本分离

最远. 在计算上采用正则化技术来提高数值计算的稳定性和改善算法的推广能力. 作为两类分类器时, Mika 等人采用一维 SVM 来确定其中的分类阈值, 在 13 个中小规模的基准数据集上取得了比 SVM 更好的结果^[21]. 为了使 KFD 算法具有稀疏性 (即出现大量的零系数), 他们还提出了用二次规划来求解 KFD 的方法及其相应的稀疏训练算法^[4, 22, 23].

线性判别分析是 Fisher 判别分析的推广, 属于考虑多类信息的降维技术. Baudat 等人利用核思想推广了线性判别分析, 并称之为广义判别分析 (GDA)^[24]. 它试图找到一组最佳的非线性投影方向使多类样本数据在少数几个分量上就能很好的分开 (希望是可视的两维). 文献[25]以更简洁的方式推导出多类线性判别分析的核形式, 并提出了利用特征值估计正则化参数的有效方法

KPCA 是企图找出数据结构信息的降维技术 (也可以是升维), 对于分类问题来说不一定是最佳的. 而 KFD 和 GDA 则考虑了类内与类间的信息, 这种降维技术对于分类问题来说是最好的

5.2 最小化误差平方和算法的核形式

最小二乘回归分析是最典型的最小化误差平方和的线性算法, 岭回归分析则在目标函数中增加了一个平方型正则项, 从而可以得到更为稳定的、方差更小的解^[26]. 1998 年 Saunders 等学者将岭回归方法进行核化, 提出了非线性的核岭回归分析方法^[27], 其解满足线性方程组 $(K + \lambda I)\alpha = y$. 应该注意的是, 这一形式不包含阈值项, 当采用线性核 (即 $x^T y$) 时无法退化为线性形式. 在推导核形式时, 建议单独考虑阈值项, 使得核形式能够退化到线性形式. Rospal 等还提出了大家不太熟悉的偏最小二乘回归算法的核形式^[28].

在模式识别中, 最小平方误差分类器 (MSE) 由于能够直接处理线性不可分问题而得到研究人员的重视, 而且它还包含两个特例 (Fisher 线性判别分析和 Bayes 分类器)^[15, 16]. Ruiz 等^[29] 根据矩阵论推导出 MSE 算法的核形式 (不含正则项), 其解满足一线性方程组, 但系数矩阵是奇异的. 他们采用广义逆矩阵来解决系数矩阵的奇异性问题, 算法的缺点是解的性质完全依赖于核函数的合理选取. 在文献[30, 31] 中提出了基于正定核和正则化技术的 3 种非线性 MSE 算法, 称为 MSE 的正则化核形式 (RKMSE), 它们的目标函数包含经验风险 (基于核函数的非线性函数的输出与期望输出之间的误差平方和) 和一合适的平方型正则项 (取 $w^\phi, w^\phi, \alpha, \alpha, \alpha + \beta^2$ 之一). 其解不仅满足一个线性方程组, 而且对于适当的正则化参数, 其系数矩阵为对称

正定阵,存在唯一解 对于不同的正则项和不同的输出编码方式,RKMSSE有3个特例:最小二乘支持向量机(LS-SVM)^[32],核Fisher判别分析(KFD)^[21](即Fisher线性判别分析与MSE关系的推广),核岭回归分析(KRR)^[27].这说明RKMSSE可以作为这些核算法的一个统一框架,以统一的方式解释了这些核算法的本质

这一类算法的缺点是其解不具有稀疏性,即有可能所有样本的系数均不等于零(或不接近于零),所以基于核函数的判别函数或回归函数的计算量较大

5.3 感知器算法的核形式

针对自适应感知器(Adatron)算法^[33],Friess等人^[34]提出了非线性的核Adatron算法,试图通过迭代过程寻找出间隔最大的分类面 但是,这一迭代算法只能处理特征空间中的线性可分问题 在假定推广能力的上界是RBF核宽度参数的连续函数的条件下,文献[35]通过多次训练核Adatron算法来确定RBF核的最佳宽度参数,这是利用核算法确定核参数的一次非常成功的尝试

关于Rosenblatt的感知器算法,已有多种不同形式的核形式 文献[3,36]直接给出了感知器算法的一种对偶形式或核形式,其迭代更新公式是给错分样本的系数加或减1,即:如果样本 j 是错分样本,则 $\alpha_j \leftarrow \alpha_j + y_j$,与势函数方法^[5,6]的思想(加减势函数)相同 所以确切地说它们是势函数方法的另一种表示,同时也只能处理特征空间的线性可分问题 应该注意的是,这一迭代过程可以产生稀疏判别函数,即只有少量样本的系数 α_j 不等于零

严格按照Rosenblatt感知器算法的推导过程^[15,16],文献[37]提出了另一种基于核函数的非线性感知器算法(简称为核感知器),其迭代更新公式是:如果样本 j 是错分样本,则 $\alpha_j \leftarrow \alpha_j + y_j y_{jk}(x_i, x_j)$, $\beta_j \leftarrow \beta_j + 1$.借助于线性感知器的收敛性定理^[2],可以直接证明核感知器的收敛性 由于这一核感知器也只能处理特征空间的线性可分问题,文献[38]利用口袋算法^[39]的思想改造核感知器算法,提出了基于核函数的非线性口袋算法(即核口袋算法),其目标是找到一组系数使错分样本数为最少 对于核口袋算法,其收敛性可以借助于口袋算法的收敛性^[40]来证明 为了进一步改善核口袋算法的推广能力,文献[41]引入支持向量机算法中最大化两类之间间隔的思想,定义了反映两类样本之间距离大小的准则:间隔准则 在核口袋算法的迭代过程中最大化这一准则,使迭代过程同时最小化错分样本数和最大化间隔准则,称之为大间隔核口袋算法 这3种

核算法的不足之处也是缺乏稀疏性,即几乎所有样本的系数均不等于零(或者不接近于零).

5.4 神经元的核形式

对于McCulloch-Pitts神经元,其输入输出关系为^[15]

$$g(x) = \text{sign}\left[\sum_i w_i x_i + b \right].$$

对于支持向量机分类算法,其判别函数为

$$g(x) = \text{sign}\left[\sum_i \alpha_i y_{ik}(x, x_i) + \beta \right].$$

对于支持向量机回归算法,其回归函数的形式为

$$g(x) = \sum_i \alpha_i k(x, x_i) + \beta$$

比较这些函数关系后,文献[42]定义了一种基于核函数的非线性神经元(核神经元),其输入输出函数关系满足

$$g(x) = f\left[\sum_i \alpha_i k(x, x_i) + \beta \right],$$

其中 f 是线性函数或Sigmoid函数等可微函数

类似于经典前向神经网络的BP训练算法,文献[42]在定义了经验风险(即神经元的实际输出与期望输出的误差平方和)的基础上构造出基于梯度法的、带冲量项的单样本迭代训练算法 为了改善算法的推广能力和获得稀疏的核表达形式,借助于文献[43]中剪辑多层前向网络的思路,文献[44,45]分别定义出包含Laplace或Gauss正则项、经验风险的目标函数,然后构造出相应的迭代训练过程 这两算法以简单的迭代过程实现了支持向量机中的核思想、正则化思想和稀疏性 从有限的实验结果来看,Laplace正则项有利于实现稀疏性,而Gauss正则项改善推广性的效果更好

5.5 距离的核形式及其应用

衡量两个样本之间相似性的距离度量(如欧氏距离、Mahalanobis距离等)是很多模式识别算法的基础 距离的核形式可以衡量高维特征空间中两个模式向量之间的相似性 Vapnik^[2]介绍了欧氏距离的核形式(核距离,即 $d^2(x, y) = k(x, x) + k(y, y) - 2k(x, y)$),Yu等人^[46]提出了基于核距离的近邻法,并在手写数字识别中取得较好的应用效果 Ruiz等人^[29]推导出Mahalanobis距离的核形式 文献[19]在早期未正式发表的版本中(<http://www.kernelmachines.org/publications>)依据核距离提出了 K 均值(即 C 均值)聚类算法的核形式 这些核距离形式同样可以应用于其它涉及到距离的模式识别、神经网络算法,例如 K 近邻法、动态聚类算法、自组织特征映射算法等

6 讨论与展望

线性算法的核形式是一类新的非线性机器学习

技术, 它们最显著的特点是利用满足 Mercer 条件的核函数(或者正定核)巧妙地将经典线性算法推广到非线性形式, 并且使欲求解优化问题的规模仅与样本数目有关, 而与样本维数无关, 从而避开了可能出现的“维数灾难”。为了提高数值计算的稳定性和控制算法的推广能力以及改善迭代过程的收敛性, 其中的多数算法还采用了正则化技术。这些核形式显著提高了线性算法处理能力, 并在人工仿真与基准数据集上取得了较好的应用效果。下面讨论进一步的研究方向与应注意的一些问题:

1) 只要能够将线性算法转化成仅包含样本向量之间内积运算的形式, 就可能将其核化, 得到对应的核形式, 这是推导线性算法的非线性核形式的基本思路。本文认为仍可能有少数不被大家熟知的线性算法可以进行核化研究。

2) Mercer 核或正定核是核化研究中最重要理论基础, 一方面它有可能避开多解性问题, 另一方面它对应着一个特征空间或再生核 Hilbert 空间, 使得两个向量(元素)之间的内积等于核函数。如果要放松正定性条件(例如, 采用一般的对称函数), 则必须考虑理论上和数值计算上可能出现的问题(如, 解的不唯一性、矩阵的奇异性等)。

3) 利用核思想与核函数将线性算法转化成核形式时, 通常会遇到不适定问题(主要表现为多解性), 其主要原因是核算法中的所求参数个数往往要大于样本个数。正则化技术是目前处理不适定问题最有效的数学工具。针对具体问题, 需要从计算的简单性以及不同的目的来选择合适的正则项。

4) 在核算法中使用的核函数是 Mercer 核或者正定核, 但是要判断一个对称函数是否满足 Mercer 定理^[1,2]或者是否满足正定性^[4,7]是很困难的, 因此应重视核函数的理论研究成果(如, 核之间运算的法则与构造核函数的规则)^[3,8], 用它们来指导如何构造出满足 Mercer 条件的核函数。

5) 核算法比较成功的应用领域是手写数字识别^[2]、文本分类^[47]与生物信息学^[9,10,48]。有学者认识到, 要想进一步提高核算法的应用效果, 必须考虑问题的背景知识, 采取的措施是在核函数选择与构造中反映出问题的先验知识^[4]。针对样本自然表达形式(非数值样本)和问题背景知识, 构造特殊的核函数在生物信息学领域已做了一些有益的尝试, 并取得较好的应用效果^[4,9~12]。由于面临的实际问题千变万化, 特殊核函数的构造将需进行大量的研究工作, 在构造特殊核函数的过程中同样需要考虑核函数的正定性。

6) 由于核算法的规模与样本数有关, 上面介绍

的核算法目前主要处理中小规模的模式分类和回归分析问题(一般样本数小于 2000)。另外核算法基本上都是离线算法, 无法动态地加入新的训练样本。因此, 研究和设计这些核算法的在线算法和快速算法是一个有实用价值的研究方向。

7) 上面介绍的多数核算法与支持向量机算法一样, 最初形式只能处理两类模式分类问题。对于多类问题, 最简单的策略是将多类问题转化为两类问题来处理。目前, 整体最优多类算法的研究主要是针对支持向量机进行的, 除了理论分析上的困难外, 由于计算能力的限制, 只进行了有限的实验比较^[49]。所以, 将线性算法的核形式从两类算法推广到整体最优的多类算法, 并研究它们的快速计算方法, 是另一个有实用价值的研究方向。

在过去 10 年中, 线性算法的核化研究已取得了丰硕的成果, 但仍存在一些理论与实际问题需要研究解决, 这将进一步推动这类非线性机器学习技术的发展和完善。

致谢 非常感谢清华大学自动系李衍达教授、阎平凡教授和陆文凯副教授对本项研究工作的指导和大力支持。

参考文献(References)

- [1] Vapnik V N. *The Nature of Statistical Learning Theory* [M]. Second Edition. New York: Springer-Verlag, 1999.
- [2] Vapnik V N. *Statistical Learning Theory* [M]. New York: John Wiley & Sons, 1998.
- [3] Cristianini N, Taylor J S. *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods* [M]. Cambridge UK: Cambridge University Press, 2000.
- [4] Scholkopf B, Smola A J. *Learning with Kernels - Support Vector Machines, Regularization, Optimization and Beyond* [M]. Cambridge MA: MIT Press, 2001.
- [5] Aizeman M A, Braveman E M, Rozonoer L I. Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning [J]. *Automation and Remote Control*, 1964, 25 (6): 821-837.
- [6] Tou J T, Gonzalez R C. *Pattern Recognition Principles* [M]. New York: Addison-Wesley, 1974.
- [7] Berg C, Christensen J P R, Ressel P. *Harmonic Analysis on Semigroup* [M]. New York: Springer-Verlag, 1984.
- [8] Genton M G. Classes of Kernels for Machine Learning: A Statistics Perspective [J]. *J of Machine Learning Research*, 2001, 2: 299-312.
- [9] Zein A, Raetsch G, Mika S, et al. Engineering SVM

- Kernels That Recognize Translation Onitiation Sites [J] *Bioinformatics*, 2000, 16(9): 799-807.
- [10] Jaakkola T, Diekhaus M, Haussler D. A Discriminative Framework for Detecting Remote Protein Homologies[J]. *J of Computational Biology*, 2000, 7(1-2): 95-114
- [11] Tsuda K, Akaho S, Kawanabe M, et al. Asymptotic Properties of the Fisher Kernel [J]. *Neural Computation*, 2004, 16(1): 115-137.
- [12] Tsuda K, Kawanabe M, Ratsch G, et al. A New Discriminative Kernel From Probabilistic Models[J]. *Neural Computation*, 2002, 14(10): 2397-2414
- [13] Mjølness E, DeCoste D. Machine Learning for Science: State of the Art and Future Prospects[J]. *Science*, 2001, 293(5537): 2051-2055
- [14] Tikhonov A N, Arsenin V Y. *Solution of Ill Posed Problem* [M]. Washington DC: Winston V H and Sons, 1977.
- [15] 边肇祺, 张学工, 等. *模式识别* [M]. 第二版. 北京: 清华大学出版社, 2000
(Bian Z Q, Zhang X G, et al. *Pattern Recognition* [M]. Second Edition. Beijing: Tsinghua University Press, 2000)
- [16] Duda R O, Hart P E. *Pattern Classification and Scene Analysis* [M]. New York: John Wiley & Sons, 1973
- [17] Saito K, Nakano R. Second Order Learning Algorithm with Squared Penalty Term [J]. *Neural Computation*, 2000, 12(3): 709-729
- [18] Scholkopf B, Smola A, Muller K-R. Kernel Principal Component Analysis [A]. *Lecture Notes in Computer Science 1327 - International Conference on Artificial Neural Networks* [C]. Berlin: Springer, 1997: 583-589
- [19] Scholkopf B, Smola A, Muller K-R. Nonlinear Component Analysis as a Kernel Eigenvalue Problem [J]. *Neural Computation*, 1998, 10(5): 1299-1319
- [20] Mika S, Scholkopf B, Smola A, et al. Kernel PCA and De-noising in Feature Spaces [A]. *Advances in Neural Information Processing Systems 11* [C]. Cambridge MA: MIT Press, 1999: 536-542
- [21] Mika S, Ratsch G, Weston J, et al. Fisher Discriminant Analysis With Kernels [A]. *Neural Networks for Signal Processing 9* [C]. New York: IEEE Press, 1999: 41-48
- [22] Mika S, Ratsch G, Muller K-R. A Mathematical Programming Approach to the Kernel Fisher Algorithm [A]. *Advances in Neural Information Processing System 13* [C]. Cambridge MA: MIT Press, 2001: 591-597.
- [23] Mika S, Smola A J, Scholkopf B. An Improving Training Algorithm for Kernel Fisher Discriminants [A]. *Proc of Artificial Intelligence and Statistics* [C]. San Francisco CA: Morgan Kaufmann, 2001: 98-104
- [24] Baudat G, Anouar F. Generalized Discriminant Analysis Using a Kernel Approach [J]. *Neural Computation*, 2000, 12(10): 2385-2404
- [25] Ma J, Sancho-Gomez J L, Ahal S C. Nonlinear Multiclass Discriminant Analysis [J]. *IEEE Signal Processing Letters*, 2003, 10(7): 196-200
- [26] 王惠文. *偏最小二乘回归方法及其应用* [M]. 北京: 国防工业出版社, 1999
(Wang H W. *Partial Least Squares Regression Method and Its Applications* [M]. Beijing: Defense Industry Press, 1999)
- [27] Saunders C, Gammeman A, Vovk V. Ridge Regression Learning Algorithm in Dual Variables [A]. *Proc of the Fifteenth Int Conf on Machine Learning* [C]. San Francisco CA: Morgan Kaufmann, 1998: 515-521.
- [28] Rosipal R, Trejo L J. Kernel Partial Least Squares Regression in RKHS [J]. *J of Machine Learning Research*, 2001, 2: 97-123
- [29] Ruiz A, Lopez-de-Teruel P E. Nonlinear Kernel-based Statistical Pattern Analysis [J]. *IEEE Transactions on Neural Networks*, 2001, 12(1): 16-31
- [30] Xu J, Zhang X, Li Y. Kernel MSE Algorithm: A Unified Framework for KFD, LS-SVM and KRR [A]. *Proc of 2001 Int Joint Conf on Neural Networks* [C]. New York: IEEE Press, 2001: 1486-1491.
- [31] 许建华, 张学工, 李衍达. 最小平方误差算法的正则化核形式 [J]. *自动化学报*, 2004, 30(1): 27-36
(Xu J H, Zhang X G, Li Y D. Regularized Kernel Forms of Minimum Squared Error Methods [J]. *Acta Automatica Sinica*, 2004, 30(1): 27-36)
- [32] Suykens J A K, Vandewalle J. Least Squares Support Vector Machines [J]. *Neural Processing Letters*, 1999, 9(3): 293-300
- [33] Anlauf J K, Biehl M. The Adatron: An Adaptive Perceptron [J]. *Europhysics Letter*, 1989, 10(7): 687-692
- [34] Friess T, Cristianini N, Campbell C. The Kernel Adatron: A Fast and Simple Learning Procedure for Support Vector Machines [A]. *Proc of the Fifteenth Int Conf of Machine Learning* [C]. San Francisco CA: Morgan Kaufmann, 1998: 188-196
- [35] Cristianini N, Campbell C, Shawe-Taylor J. Dynamically Adapting Kernels in Support Vector Machines [A]. *Advances in Neural Information Processing Systems 11* [C]. Cambridge MA: MIT Press, 1999: 204-210

(下转第12页)

参考文献(References)

- [1] Rifkin R M, Yeo G, Poggio T. Regularized Least Squares Classification [A]. *Advances in Learning Theory: Methods, Model and Applications*[C]. NATO Science Series III: Computer and Systems Sciences, Amsterdam: DS Press, 2003: 131-153
- [2] Poggio T, Smale S. The Mathematics of Learning: Dealing With Data [J]. *Notice of American Mathematical Society*, 2003, 50(5): 537-544
- [3] Cucker F, Smale S. On the Mathematical Foundations of Learning [J]. *Bulletin of American Mathematical Society*, 2001, 39(1): 1-49
- [4] Suykens J A K, Gestel V, De Brabanter J, et al. *Least Squares Support Vector Machine* [M]. Singapore: World Scientific Press, 2002
- [5] Fung G, Mangasarian O L. Proximal Support Vector Machine Classifiers [A]. *Seventh ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining* [C]. San Francisco: Association for Computing Machinery Press, 2001: 77-86
- [6] Cristianini N, Kandola J, Elisseeff A, et al. *On Kernel Target Alignment* [EB/OL]. <http://www.support-vector.net/papers/alignment>. JMLR. ps 2005-01-01.
- [7] Chapelle O, Vapnik V, Mukherjee S. Choosing Multiple Parameters for Support Vector Machines [J]. *Machine Learning*, 2002, 46(1): 131-159
- [8] Jaakkola T, Haussler D. Probabilistic Kernel Regression Models [A]. *Proc of the 1999 Conf on AI and Statistics* [C]. San Mateo: Morgan Kaufmann, 1999
- [9] Ratsch G, Onoda T, Müller K-R. Soft Margins for AdaBoost [J]. *Machine Learning*, 2001, 42(3): 287-320
- [10] Keerthi S S. Efficient Tuning of SVM Hyperparameters Using Radius/margin Bound and Iterative Algorithms [J]. *IEEE Trans on Neural Networks*, 2002, 13(5): 1225-1229
- [11] Staelin C. *Parameter Selection for Support Vector Machines* [R]. HP Laboratories, 2003
- [12] Guyon I, Stork D G. Linear Discriminant and Support Vector Classifiers [A]. *Advances in Large Margin Classifiers* [C]. Cambridge MA: MIT Press, 2000: 147-170
- [13] 许建华, 张学工, 李衍达. 一种基于核函数的非线性感知器算法 [J]. *计算机学报*, 2002, 25(7): 689-695 (Xu J H, Zhang X G, Li Y D. A Nonlinear Perceptron Algorithm Based on Kernel Functions [J]. *Chinese J of Computers*, 2002, 25(7): 689-695)
- [14] 许建华, 张学工, 李衍达. 基于核函数的非线性口袋算法 [J]. *电子学报*, 2003, 31(4): 612-615 (Xu J H, Zhang X G, Li Y D. Nonlinear Pocket Algorithm with Kernels [J]. *Acta Electronica Sinica*, 2003, 31(4): 612-615)
- [15] Gallant S I. Optimal Linear Discriminant [A]. *Proc of the Eighth Int Conf on Pattern Recognition* [C]. New York: IEEE Press, 1986: 849-853
- [16] Muselli M. On Convergence Properties of Pocket Algorithm [J]. *IEEE Trans on Neural Networks*, 1997, 8(3): 623-629
- [17] Xu J, Zhang X, Li Y. Large Margin Kernel Pocket Algorithm [A]. *Proc of 2001 Int Joint Conf on Neural Networks* [C]. New York: IEEE Press, 2001: 1480-1485
- [18] Xu J, Zhang X, Li Y. Kernel Neuron and its Training Algorithm [A]. *Proc of the Eighth Int Conf on Neural Information Processing* [C]. Shanghai: Fudan University Press, 2001, 2: 861-866
- [19] Ishikawa M. Structural Learning with Forgetting [J]. *Neural Networks*, 1996, 9(3): 509-521
- [20] Xu J, Zhang X, Li Y. Sparse Training Procedure for Kernel Neuron [A]. *Proc of Int Conf on Neural Networks and Signal Processing* [C]. New York: IEEE Press, 2003, 1: 49-53
- [21] Xu J, Zhang X. A Learning Algorithm with Gaussian Regularizer for Kernel Neuron [A]. *2004 Int Symposium on Neural Networks* [C]. Berlin: Springer, 2004: 252-257.
- [22] Yu K, Ji L, Zhang X. Kernel Nearest Neighbor Algorithm [J]. *Neural Processing Letters*, 2002, 15(2): 147-156
- [23] Joachims T. Making Large-scale SVM Learning Practical [A]. *Advances in Kernel Methods - Support Vector Learning* [C]. Cambridge MA: MIT Press, 1999: 169-184
- [24] Baldi P, Brunak S. *Bioinformatics: The Machine Learning Approach* [M]. 2nd ed. Cambridge MA: MIT Press, 2001.
- [25] Hsu C W, Lin C J. A Comparison of Methods for Multiclass Support Vector Machines [J]. *IEEE Trans on Neural Networks*, 2002, 13(2): 415-425

(上接第6页)

- [26] 许建华, 张学工, 李衍达. 一种基于核函数的非线性感知器算法 [J]. *计算机学报*, 2002, 25(7): 689-695 (Xu J H, Zhang X G, Li Y D. A Nonlinear Perceptron Algorithm Based on Kernel Functions [J]. *Chinese J of Computers*, 2002, 25(7): 689-695)
- [27] 许建华, 张学工, 李衍达. 基于核函数的非线性口袋算法 [J]. *电子学报*, 2003, 31(4): 612-615 (Xu J H, Zhang X G, Li Y D. Nonlinear Pocket Algorithm with Kernels [J]. *Acta Electronica Sinica*, 2003, 31(4): 612-615)
- [28] Gallant S I. Optimal Linear Discriminant [A]. *Proc of the Eighth Int Conf on Pattern Recognition* [C]. New York: IEEE Press, 1986: 849-853
- [29] Muselli M. On Convergence Properties of Pocket Algorithm [J]. *IEEE Trans on Neural Networks*, 1997, 8(3): 623-629
- [30] Xu J, Zhang X, Li Y. Large Margin Kernel Pocket Algorithm [A]. *Proc of 2001 Int Joint Conf on Neural Networks* [C]. New York: IEEE Press, 2001: 1480-1485
- [31] Xu J, Zhang X, Li Y. Kernel Neuron and its Training Algorithm [A]. *Proc of the Eighth Int Conf on Neural Information Processing* [C]. Shanghai: Fudan University Press, 2001, 2: 861-866
- [32] Ishikawa M. Structural Learning with Forgetting [J]. *Neural Networks*, 1996, 9(3): 509-521
- [33] Xu J, Zhang X, Li Y. Sparse Training Procedure for Kernel Neuron [A]. *Proc of Int Conf on Neural Networks and Signal Processing* [C]. New York: IEEE Press, 2003, 1: 49-53
- [34] Xu J, Zhang X. A Learning Algorithm with Gaussian Regularizer for Kernel Neuron [A]. *2004 Int Symposium on Neural Networks* [C]. Berlin: Springer, 2004: 252-257.
- [35] Yu K, Ji L, Zhang X. Kernel Nearest Neighbor Algorithm [J]. *Neural Processing Letters*, 2002, 15(2): 147-156
- [36] Joachims T. Making Large-scale SVM Learning Practical [A]. *Advances in Kernel Methods - Support Vector Learning* [C]. Cambridge MA: MIT Press, 1999: 169-184
- [37] Baldi P, Brunak S. *Bioinformatics: The Machine Learning Approach* [M]. 2nd ed. Cambridge MA: MIT Press, 2001.
- [38] Hsu C W, Lin C J. A Comparison of Methods for Multiclass Support Vector Machines [J]. *IEEE Trans on Neural Networks*, 2002, 13(2): 415-425