基于再生核的机器学习方法

孙建成 戴利云 著

江西高校出版社

图书在版编目(CIP)数据

基于再生核的机器学习方法/孙建成,戴利云著.—南昌:江西高校出版社,2015.11
ISBN 978-7-5493-3837-5

Ⅰ.①基… Ⅱ.①孙… ②戴… Ⅲ.①机器学习 Ⅳ.①TP181

中国版本图书馆 CIP 数据核字(2015) 第 282932 号

出版发行	江西高校出版社
社　　址	江西省南昌市洪都北大道 96 号
邮政编码	330046
总编室电话	(0791) 88504319
销售电话	(0791) 88511423
网　　址	www.juacp.com
印　　刷	南昌达胜德数字印刷有限责任公司
照　　排	江西太元科技有限公司照排部
经　　销	各地新华书店
开　　本	890mm×1240mm　1/32
印　　张	4.625
字　　数	180 千字
版　　次	2015 年 11 月第 1 版第 1 次印刷
书　　号	ISBN 978-7-5493-3837-5
定　　价	28.00 元

赣版权登字-07-2015-841

目 录

前　言

机器学习是人工智能的一个重要分支，主要研究如何构建有效的学习方法，使之通过学习获得蕴藏在观测样本中的规律，然后利用这些规律对未来样本进行分析和预测。当前，在机器学习领域，基于再生核的学习和基于贝叶斯推理的学习受到高度关注。前者将样本映射到高维特征空间中，并运用事先定义的再生核取代特征空间中的内积运算，从而简化了学习任务，提供学习方法非线性化的新途径，规避因内积运算而可能引发的维数灾难；后者在学习过程中融入贝叶斯推理，不仅可以使学习方法具有概率背景，而且能在学习过程中导入先验知识，并允许学习结果出现不确定性。

基于再生核的学习和基于贝叶斯推理的学习相结合，使得学习方法可以兼具以上两者的优势。本书正是在上述结合点上展开，通过深入研究核参数的选择、核函数的修改、支持向量

机以及最小最大概率机等方法，获得了以下研究成果：

(1) 对支持向量机中的核函数的参数选择问题进行了深入的研究。在支持向量机的算法中，一个重要的步骤是选择核函数的类型以及核函数所涉及的相关参数，核函数的类型以及核函数参数值对于核方法的性能有着重要的影响。本书所做的相关工作是在特征空间中，通过最大化两类均值之间的距离来选择最佳的核函数参数。基于一个归一化的核函数，将两类均值距离作为可分性度量。通过分析可知，两类均值距离不仅包含了类间距离的信息，也隐含地包括了类内数据几何分布方面的信息。将两类均值距离作为目标函数，采用梯度下降法寻找最优的核函数参数，基于几何方面分析以及仿真实验，所提出的算法的计算复杂度在很大程度上得到降低。利用多个人工和实际数据作为测试数据，仿真实验结果表明所提出算法的性能优于经典的核参数选择算法。

(2) 对核函数进行修改，以期望获得更好的分类性能。核函数的类型和形式对于分类的性能非常重要，所做的主要工作是提出了仅依赖于分类数据，对核函数进行修改和调整以优化核方法分类性能的算法。传统算法一般是使用一定的分类算法训练得出分类面，并在特征空间中寻找支持向量，然后基于支持向量对核函数进行修改。基于分类面附近样本点的特性，本书提出在特征空间中寻找分类面附近的样本点来替代支持向量，然后基于这些样本点修改和调整核函数。所提出的算法不需要采用任何分类算法进行事先的训练以获得分类面，并可以应用于多种核方法的分类算法。仿真实验结果表明，所提出算法的分类性能和稳定性都得到了一定程度的提高。

(3) 基于核方法对非线性时间序列进行建模和预测。非线性时间序列的预测一直是时间序列分析领域中的重要研究内容，非线性时间序列由于其随机性等特性，对其进行预测是一项困难的工作，本

书主要包括两方面的内容:①首先将 RLS-SVM 应用于混沌信号的预测与建模,由于衰落信道具有混沌特性,所以这对于将 RLS-SVM 应用到衰落信道预测具有重要的指导意义。通过重建系统的高维相空间,从而获得比标量时间序列更多的系统信息,进而采用 RLS-SVM 进行回归预测,由于 RLS-SVM 算法本身的缺陷,并没有充分利用到高维空间的有用信息,影响了预测性能,因而通过修改代价函数,使得训练数据集的输入和输出均为矢量,使得 RLS-SVM 能够充分地利用系统重建相空间的信息。仿真结果表明,与传统线性算法相比,一方面提高了预测性能,另一方面发现 RLS-SVM 对噪声具有不敏感性,使得本算法适于噪声环境中。②基于最小最大概率回归机(Minimax Probability Machine Regression,MPMR),对时间序列进行预测。由于正的李雅普诺夫指数(Lyapunov exponents)标示着非线性时间序列的非线性特性以及稳定性,在对时间序列进行预测时,利用李雅普诺夫指数项对代价函数进行加权,以补偿预测误差的扩散。仿真结果表明此方法可以较好地捕获系统的特性,并适用于时间序列的多步预测,为时间序列的长期预测提供了新的思路和方法。

孙建成,戴利云

2015 年秋

第 1 章　绪论

1.1　研究的背景和意义

学习是人类获得知识与探索世界的重要手段，机器学习(Machine Learning) ——让机器拥有类似人类的学习能力——是人类的梦想。2004 年 B.Gates 在一次演讲中甚至宣称 “如果在人工智能上有所突破，以至于机器能够学习，那么它将价值 10 个微软。”

机器学习是研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。自从计算机问世以来，赋予其学习能力的梦想就不断推动着机器学习这一研究领域向前发展，并且已经有了十分广泛的应用，例如: 生物特征识别、搜索引擎、医学诊断、检测信用卡欺诈、证券市场分析、DNA 序列测序、语音和手写识别、计算机视觉、战略游戏和机器人运用。如今机器学习已经演变成为人工智能的一个完整分支，而且与其他一些研究领域具有密切联系，例如人们熟知的模式分类(pattern classification) [1] 就可以看作是机器学习的特例。深入开展机器学习的研究，既可以加深理解人类大脑的学习机理，又可以研制出拥有学习能力的智能机器，使人脑的功能得到进一步的延伸和物化，还可以从理论上探索一些人类尚未发现的新学习方法和途径，因此对机器学习进行研究意义十分深远。

尽管机器学习这一研究领域十分活跃而且充满生命力，但迄今为止对“机器学习”一词并没有一个精确和能被公认的定义。学习是人类具有的一种重要智能行为，但究竟什么是学习，长期以来却众说纷纭。社会学家、逻辑学家和心理学家都各有其不同的看法。从事哲学研究的学者注重归纳学习的本质，来自生物学领域的学者注重探寻学习的生物机制; 而统计学家们将学习视为借助统计手段挖掘

样本集内蕴含的规律，人工智能专家则更关心学习策略如何实现等[2]。另一方面，机器学习的模仿对象是人类的学习能力，而学习本身是一种多侧面与综合性的心智活动，它与记忆、思维、知觉和感觉等多种心理行为都有密切的联系，使得人们很难把握其机理与实现。

尽管如此，为了便于进行讨论和研究学科的进展，有必要对机器学习给出定义，即使这种定义是不完全的和不充分的。目前在机器学习领域影响较大的是人工智能大师 Simon 的观点：机器学习是系统的任何增强或改进，这种改进使得系统在重复同样的工作或进行类似的工作时，会比原来做得更好或效率更高。Mitchell[3] 则将机器学习定义为：对于某类任务 T 和性能度量 P，如果一个计算机程序在 T 上以 P 衡量的性能随着经验 E 而自我完善，那么称机器从经验 E 中学习。而 Vapnik[4] 认为机器学习是基于有限观测样本寻求样本与其目标值(targets) 间未知的依赖关系。通俗地说就是从一些观测样本出发得出目前尚不能通过原理分析得到的规律，利用这些规律去分析客观现象，对未来样本或无法观测的样本进行预测。Vapnik 同时认为在实践中有三类最基本的机器学习问题，它们分别是分类、回归以及概率密度估计。本书遵循 Vapnik 的观点，从统计角度看待学习问题，因此下文如无特别说明，所提及的学习均是指基于样本的学习。

机器学习是人工智能研究较为年轻的分支，它的发展过程大体上可分为三个时期。第一个时期是从 20 世纪 50 年代直至 80 年代中叶，研究热点经历了由“无知”学习、概念学习到探索各种学习策略的转变。第二个时期是从 80 年代中叶至 90 年代中叶，机器学习领域内无疑是人工神经网络(ANN) 独领风骚，各种网络学习方法如 Hopfield 网和波尔茨曼机等应运而生，以至于 1989 年 Carbonell[5] 将 ANN 列为当时机器学习的四大发展趋势之首。ANN 虽然有强大的学习能力和自适应能力，但是由于缺乏定量的分析与完备的理论基础支持，在实际应用中往往需要经过费时的摸索才能确定合适的网络模型以及参数设置，其应用效果完全取决于使用者的经验。因此八年后的 1997 年，Dietterich[6] 指出分类器集成、海量样本学习、增强学习和学习复杂随机模型已经成为机器学习新的发展趋势。斗转星

移,如今,机器学习研究已经进入第三个时期,在机器学习领域内也涌现出不少新的研究热点,而对 ANN 的研究却逐渐归于平静。本书无力也无意评介机器学习当前的发展趋势,只是指出在当前机器学习领域的诸多研究热点中,基于再生核的学习尤其令人振奋,它得益于早期研究所形成的牢固基础和如今日益强大的计算能力,部分反映了机器学习当前的发展趋势。

基于再生核的学习(以下简称为基于核的学习,在有些文献中被称为核机器学习或核机器)[7—8]近年在机器学习领域内掀起了阵阵波澜。基于核的学习,首先以支持向量机(support vector machine,SVM)的形式出现,然而,很快就产生了基于核的其他算法,它能够解决分类以外的问题。人们越来越清楚地认识到,这种方法引起了模式分析领域的一场革命。其基本思想是对一些只涉及样本间内积运算的学习方法,通过改变内积定义的方式,用事先定义的核函数取代内积,从而得到与原学习方法对应的非线性版本——基于再生核的学习方法(以下简称为核方法)。用核函数取代内积的这种方式被称为“核技巧(kernel trick)”,通过应用核技巧获得核方法的过程被称为“核化(kernelizing)”。

基于核的分析,对于数学家、科学家和工程师来说,是一个强大的新工具。它提供了非常丰富的方法,可以应用在模式分析、信号处理、句法模式识别和其他模式识别(从样条到神经网络)领域。简而言之,它提供了一个崭新的视角,我们仍然远没有了解它的全部潜力。考究基于核的学习受关注的原因,得益于其通过运用核技巧而带来的一些优势:

第一,提供了学习方法非线性化的新途径,在线性与非线性间架设起了一座桥梁。核方法使得人们能够高效率地分析蕴藏在样本集内部的非线性关系,而这种高效率原先只有线性学习方法才能够达到。一些早期的方法比如主成分分析(PCA),在揭示样本间的线性关系时很有效率,但对处理非线性关系却无能为力。而与之对应的核方法可以将线性学习方法的理论性,与非线性学习方法的灵活性与适用性相结合,从而形成了一类新的强有力的、稳健的学习方法。

第二,简化了欲解决的问题。以分类为例,基于核的学习首先通

过一个非线性映射将样本从输入空间(input space) 映射到某高维特征空间(feature space) 中,然后在特征空间内进行分类。正如Cover[9]所指出的: 一个复杂的模式分类问题被非线性映射到高维空间中以后,该模式比在原始空间中更可能线性可分。

第三,借助核技巧规避了特征空间内的内积运算因映射函数而可能引发的维数灾难。一方面,基于核的学习给定的非线性映射是由再生核诱导的隐映射,仅在推导过程中概念性地使用这种映射,并不需要知道映射的具体形式。另一方面,特征空间中的内积运算实际上是由核函数完成的,映射函数并不真正参与运算,故而算法复杂度没有因映射的存在而增加。

第四,基于核的学习具有“可重用性”。这意味着同一个核方法,若选择的核不同就可以应用到不同的领域。例如通过设计特殊的核,可使核方法在处理非向量型数据(比如串与树等) 方面非常有效,这样就能方便地将之应用到图像与视频处理等领域。

对再生核的研究可以追溯到 1909 年 Mercer[10] 提出的 Mercer 定理。1964 年 Aizerman[11] 给出了势函数方法的收敛性证明,这是核技巧的首次运用。1992 年 Boser 与 Vapnik[12] 借用核技巧构建支持向量机(SVM) 。1998 年 Schölkopf[13] 把核技巧进一步推广至任何包含点积运算的算法中; 随后许多学者基于 Schölkopf 的思路构建了众多核方法。从以上发展历程可以看出,再生核的定义和使用并非源自机器学习本身。但是再生核作为一种学习方法的提出,则与统计学习理论和以此为基础的 SVM 的研究发展密不可分,再生核的许多性质正是在对 SVM 的研究中不断发展并得以推广应用。

对核方法的关注程度从其广阔的应用领域可见一斑。近年来,核方法已在入侵检测[14]、目标跟踪[15]、地球物理反演[16]、与离散余弦变换相结合用于图像压缩[17]、传感器网络中的分散检测(decentralized detection) [18]、说话人自适应[19—20] 以及超光谱(hyperspectral) 图像分析[21—22] 等领域取得了成功地应用。

当前,与核方法相关的文献在各种学术杂志上屡见不鲜。同时核方法已成为机器学习类会议(例如 ICML,NIPS 与 IJCNN 等) 的讨论主题之一。在国外,学者们撰写了众多关于核方法的综

述[7—8,23—25]，建立了相关网站[26—28]，出版了一系列论著[29—35]。在国内，文献[42]是最早介绍 SVM 的中文文献，目前已出版了几本关于 SVM 的论著[39—40]以及关于核方法的译著[36—38]，但还没有一本全面研究核方法的论著。

1.2　核方法的类型概述

核方法可分为有监督型和无监督型两种类型。前者所处理的样本集的类别归属已事先标定，后者主要用来处理未被标定的样本集。本节对当前受关注的核方法加以整理和归并，并注重方法间的比较。

1.2.1　有监督型核方法

1.支持向量机(Support Vector Machine，SVM)

SVM 以统计学习理论(SLT)[41]为基础，通过最小化期望风险(expectation risk)的上界构建，它遵循结构风险最小化(SRM)准则而不是传统的经验风险最小化(ERM)准则。SVM 设计精巧，在解决小样本、非线性、高维学习问题时有独特优势。以往机器学习领域内困扰人们很久的问题，比如模型选择、维数灾难和局部极小点等问题，借助 SVM 都能得到很大程度的解决。而且 SVM 与许多传统学习方法有紧密的联系。因此当前 SVM 与 SLT 一起被很多学者认为是研究机器学习的一个基本框架，SVM 也成为近年来最受关注的核方法。

根据功能差异，SVM 又可细分为支持向量分类机(SVMC)[42]、支持向量回归机(SVMR)[43]和用于逼近的支持向量机(SVMA)[44]等，有时如无必要细分几者则统称为 SVM。因此本书中“SVM”一词有时是上述方法的统称，有时指代单个方法，具体含义由上下文决定。

2.最小最大概率机(Minimax Probability Machine，MPM)

最小最大概率机(MPM)包括 2002 年 Lanckriet[45—46]提出的最小最大概率分类机(MPMC)与稳健最小最大概率机(RMPM)、2004 年 Huang[47]提出的偏置最小最大概率机(BMPM)和 2002 年 Strohmann[48]提出的最小最大概率回归机(MPMR)等。SVM 通过使期望风险的上界最小化而构建，而 MPM 通过使分类错误率的上界最小化而构建。本书第 2 章第 3 节将会指出，期望风险与分类错误率是泛

化错误(generation error)的不同表现形式,而且两者在一定条件下是等价的,从这一点出发可推知 MPM 与 SVM 有密切联系。像 MPM 与 SVM 这样"通过使泛化错误的上界最小化来导出新的学习方法"提供了构建机器学习方法的新途径,上述新途径也促使学者们对泛化错误的界进行深入的研究。

3.**核最近邻**(Kernel Nearest Neighbor, KNN)

最近邻(NN)是一种经典的机器学习方法,主要用于解决分类问题。K 最近邻(k-nearest neighbor,KNN)分类算法,是一个理论上比较成熟的方法,也是最简单的机器学习算法之一。该方法的思路是:如果一个样本在特征空间中的 k 个最相似(即特征空间中最邻近)的样本中的大多数属于某一个类别,则该样本也属于这个类别。Yu[49]用核距离取代欧氏距离,提出了核最近邻(KNN)方法。与 NN 相比,KNN 仅仅改变了距离的计算公式,因此计算复杂度并没有增加;但由于核映射降低了分类难度,因此 KNN 的分类性能比 NN 高。

4.**核 Fisher 判别**(Kernel Fisher Discriminant,KFD)

线性判别分析,简称判别分析,是统计学上的一种分析方法,用于在已知的分类之下遇到有新的样本时,选定一个判别标准,以判定如何将新样本放置于哪一个类别之中。Fisher 判别的核心思想是利用投影降维,判别依据是使类内离差尽可能小,而不是类间投影的离差尽可能大。Fisher 判别通过使广义 Rayleigh 商 $J(w)=(w^TS_Bw)/(w^TS_Ww)$ 最大化求得判别函数的法向量 $w=S_W^{-1}(\hat{\mu}_1-\hat{\mu}_2)$,其中$\hat{\mu}_1$ 和$\hat{\mu}_2$ 分别是两类样本的均值向量,S_B 和S_W 分别是两类样本的类间和类内散度矩阵。KFD 实际上是在特征空间中做 Fisher 线性判别[50],此时由于S_B 和S_W 均是 φ 的函数,必须将广义 Rayleigh 商变形为 $J(\alpha)=\alpha^TA\alpha/\alpha^TB\alpha$,其中矩阵 A 和 B 均与 φ 无关,因此可通过最大化 $J(\alpha)$ 求出 α。尽管此时 w 仍无法求出,但能求出特征空间中的测试样本 $\varphi(x)$ 在 w 上的投影。

5.**核感知器**(Kernel Perceptron,KP)

Rosenblatt 提出的感知器是机器学习史上的里程碑,标志着对学习过程进行数学研究的真正开始,它通过使感知准则极小化而求出判别函数的法向量。Xu[51—52]将感知器扩展到核空间中,提出了核感

知器(KP) 。感知器只能对线性可分的两类样本进行判别,而核感知器则可适用于非线性可分的情形。

1.2.2　无监督型核方法

1.核主成分分析(Kernel Principal Component Analysis, KPCA)

在统计学中,主成分分析是一种简化数据集的方法。它是一个线性变换,在样本均值 $\hat{\mu}=0$ 的前提下通过样本的协方差矩阵 $\sum$ 的特征值分解 $\sum v_i = \lambda_i v_i$ 可求出各个主成分方向v_i。这个变换把数据变换到一个新的坐标系统中,使得任何数据投影的第一大方差在第一个坐标(称为第一主成分) 上,第二大方差在第二个坐标(第二主成分) 上,以此类推。主成分分析经常用减少数据集的维数,同时保持数据集的对方差贡献最大的特征。KPCA 是在核空间中进行主成分分析[13],但在核空间中 $\sum$ 是 φ 的函数,无法通过 $\sum$ 的分解求得 v_i;因此 Schölkopf 令 $v_i = \sum_{j=1}^{N} \alpha_{ij}\varphi(x_j)$,然后通过核矩阵$K_N$ 的特征值分解 $N\lambda_i \alpha_i = K_N \alpha_i$ 求取 α。尽管此时v_i 仍无法求出,但能求出核空间中的测试样本 $\varphi(x)$ 的第 i 个主成分分量 $P_i\varphi(x) = \varphi(x) \cdot v_i = \sum_{j=1}^{N} \alpha_{ij} (K_N)_{ij}$。上述做法可行的前提是核空间中的样本均值也为零,为了去掉此前提还需对核空间中的样本进行归一化处理。

2.核独立成分分析(Kernel Independent Component Analysis,KICA)

在统计学中,独立成分分析或独立分量分析(Independent components analysis,ICA) 是一种利用统计原理进行计算的方法。它是一个线性变换。这个变换把数据或信号分离成统计独立的非高斯的信号源的线性组合。独立成分分析是盲信号分离(blind source separation) 的一种特例。PCA 得到的各主成分只考虑了二阶统计量,对于非高斯分布的随机变量,其高阶统计量还带有许多信息不能忽略,此时可用独立成分分析抽取独立分量。Bach[53] 最早在核空间中构建了与 ICA 对应的 KICA 方法。KICA 分为两类,一类是 KICA-KCCA,即用 KCCA 来计算 KICA 中的对照(contrast) 函数; 另一类是 KICA-KGV,是用核广义方差(kernel generalized variance, KGV) 算法计算对

照函数。与上述 Bach 的方法不同,文献[54]提出的 KICA 方法分为两步:首先将样本映射到特征空间中,并用 KPCA 进行白化(whiten)处理,然后对处理结果进行传统的 ICA 分离。

3.**核聚类**(kernel-based clustering)

聚类是把相似的对象通过静态分类的方法分成不同的组别或者更多的子集(subset),这样让在同一个子集中的成员对象都有相似的一些属性,常见的包括在坐标系中更加短的空间距离等。Girolami[55]最早提出了核聚类方法。经典的聚类方法比如 K 均值聚类要求样本具有近似球状分布,而核聚类在输入空间中样本不呈球状分布时也能对其聚类。假设输入空间中 N 个样本可聚成 K 类,现寻找最好的聚类模式 Z,使得样本到各自类内中心 $\hat{\mu}_k$ 的距离的平方和最小,即 $Z=\arg_z \min T_r(S_W)$,其中 S_W 是样本的类内散度矩阵。核聚类方法同样通过使 $T_r(S_W)$ 最小求 Z,只不过此时 $T_r(S_W)$ 的表达式中含有映射函数 φ,但采用核技巧可以将之化成与 φ 无关的形式。

4.**核自组织映射**(kernel-based self-organized maps, KSOM)

自组织神经网络(SOM)是基于无监督学习方法的神经网络的一种重要类型,具有聚类、自组织、自学习以及可视化的功能。自组织映射网络通过寻找最优参考矢量集合来对输入模式集合进行分类。自组织神经网络是神经网络最富有魅力的研究领域之一,它能够通过其输入样本学会检测其规律性和输入样本相互之间的关系,并且根据这些输入样本的信息自适应调整网络,使网络以后的响应与输入样本相适应。SOM 在对竞争获胜的神经元及其邻域内的神经元的权值进行调整时,以欧氏距离为度量,这将导致当输入样本分布结构呈高度非线性时,其聚类能力下降。Donald[56]最早将 SOM 核化并提出了核自组织映射(KSOM)的概念,从而使聚类可以形成于映射后的高维特征空间中,但其缺点是聚类结果不易在输入空间中获得直观解释,因为在特征空间中的获胜神经元并不一定在原输入空间中存在相应的原像[57]。

1.3　核方法研究进展

1.3.1　支持向量机研究进展

基于统计学习理论,AT&T Bell 实验室的 Vapnik 提出的支持向量机(Support Vector Machine,SVM) 已成为机器学习领域新的研究热点[58]。支持向量机方法是建立在统计学习理论的 VC 维理论和结构风险最小原理基础之上的,根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折中,以期获得最好的推广能力。统计学习理论是建立在一套较坚实的理论基础之上的,为解决有限样本学习问题提供了一个统一的框架。支持向量机方法的几个主要优点有:

第一,它是专门针对有限样本情况的,其目标是得到现有信息下的最优解而不仅仅是样本数趋于无穷大时的最优值。

第二,算法最终将转化成为一个二次型寻优问题,从理论上说,得到的将是全局最优点,解决了在神经网络方法中无法避免的局部极值问题。

第三,算法将实际问题通过非线性变换转换到高维的特征空间(feature space) ,在高维空间中构造线性判别函数来实现原空间中的非线性判别函数,特殊性质能保证机器有较好的推广能力,同时它巧妙地解决了维数问题,其算法复杂度与样本维数无关。

支持向量机的本质是求解一个二次规划(quadratic programming, QP) 问题。在二次型寻优过程中要进行大量的矩阵运算,多数情况下,寻优算法是占用算法时间的主要部分。通常,训练算法改进的思路是把要求解的问题分成许多子问题,然后通过反复求解子问题来求得最终的解。1995 年,Cortes 和 Vapnik 提出 Chunking 算法[59],其出发点是删除矩阵中对应拉格朗日乘数为零的行和列将不会影响最终的结果。因此,可将一个大型 QP 问题分解为一系列较小规模的 QP 问题。但是该算法的一个前提是: 支持向量的数目比较少,如果支持向量的数目本身就比较多,那么随着训练迭代次数的增加,工作样本数也越来越大,就会导致算法无法实施。针对大训练样本问题,Platt 提出了序贯最小优化(sequential minimal optimization, SMO) 算