第28卷 第3期
2002年5月
自 动 化 学 报
ACTA AUTOMATICA SINICA
Vol. 28, No. 3
May, 2002

短文

基于变异函数的径向基核函数参数估计[1)]

阎 辉 张学工 马云潜 李衍达
(清华大学自动化系，智能技术与系统国家重点实验室 北京 100084)
(E-mail: yanhui@simba.au.tsinghua.edu.cn)

摘 要 研究了支持向量机(support vector machine, SVM)方法在一定假设条件下，核函数取为样本协方差函数时解的具体形式，得出了在该假设情况下SVM方法等价于克立格方法的结论，提出了用协方差函数作为SVM核函数的思想. 考虑到在某些情况下协方差函数可能不存在，因此考虑用变异函数来代替协方差函数估计径向基核函数的宽度参数. 这样不仅解决了SVM中径向基核函数宽度参数的确定问题，而且把这种情况下的SVM拟合与概率统计学中的克立格方法联系了起来，赋予了SVM方法新的统计上的意义.

关键词 支持向量机，克立格方法，协方差函数，径向基核函数
中图分类号 TP18

THE PARAMETER ESTIMATION OF RBF KERNEL FUNCTION BASED ON VARIOGRAM

YAN Hui ZHANG Xue-Gong MA Yun-Qian LI Yan-Da
(Department of Automation, Tsinghua University,
State Key Laboratory of Intelligent Technology and Systems, Beijing 100084)
(E-mail: yanhui@simba.au.tsinghua.edu.cn)

Abstract The paper studies the form of the result of SVM with covariance function as the kernel function under some conditions, and draws the conclusion that Kriging is equivalent to SVM under those conditions. Based on the conclusion, we put forward the idea of using the covariance function as a substitute for the RBF kernel function of SVM. Considering that the covariance function could not exist in some conditions, we put forward the idea of using the variogram function as a substitute for the covariance function and prove their equivalence. It not only solves the problem of parameter estimation of SVM kernel function, but also connects the SVM with the Kriging, which gives new statistical meaning to SVM.

1) 国家自然科学基金"基于统计学习理论模式识别方法研究"(69885004)资助
收稿日期 2000-06-06 收修改稿日期 2001-01-10

Key words Support vector machine, Kriging, covariance function, RBF kernel function

1 引言

本文的研究目的是希望通过对 SVM 与克立格方法的等价性研究,得到 SVM 等价于克立格方法的充分条件,在此基础上借助克立格方法来确定 SVM 的核函数. 本文介绍了 SVM 用于模式识别和函数逼近的基本原理,以及克立格方法原理,在此基础上推导了在一定的假设下 SVM 方法得到的结论等价于克立格方法得到的结论. 根据该结论提出了利用协方差函数确定 SVM 核函数参数的思想,并且在该思想的指导下尝试用协方差函数来确定径向基核函数的宽度参数.

2 SVM 与普通克立格方法的关系

SVM 方法是从线性可分情况下的最优分类面提出的,其基本思想概括起来就是通过非线性变换将输入空间变换到一个高维空间. 在这个新空间中求取最优线性分类面,所求得的分类函数形式上类似于一个神经网络,其输出是若干中间层节点的线性组合,而每一个中间层节点对应于输入样本与一个支持向量的内积[1~4].

在 SVM 基础上,近来提出了基于 SVM 的函数回归方法[1],将 SVM 的思想推广到函数回归和函数估计领域. 当存在函数 $f(\boldsymbol{x})=(\boldsymbol{w}\cdot\boldsymbol{x})+b$ 以精度 ε 逼近所有的观测样本对 $(\boldsymbol{x}_i,y_i)(i=1,\cdots,n,\ \boldsymbol{x}_i\in R^d,\ y_i\in R)$时,支持向量机函数拟合的问题可表示为

$$\min\quad F(\boldsymbol{w},\xi,\xi^*)=\frac{1}{2}\|\boldsymbol{w}\|^2+C\Big(\sum_{i=1}^{n}\xi_i^2+\sum_{i=1}^{n}\xi_i^{*2}\Big) \tag{1a}$$

$$\text{s.t.}\ \begin{cases}(\boldsymbol{w}\cdot\boldsymbol{x}_i)+b-y_i\leqslant\varepsilon+\xi_i\\ y_i-(\boldsymbol{w}\cdot\boldsymbol{x}_i)-b\leqslant\varepsilon+\xi_i^*\\ \xi_i\geqslant 0\\ \xi_i^*\geqslant 0\end{cases} \tag{1b}$$

当用核函数 $K(\boldsymbol{x}_i,\boldsymbol{x}_j)$代替输入空间的内积$(\boldsymbol{x}_i\cdot\boldsymbol{x}_j)$时,其对应的对偶问题为

$$\max W(\alpha,\alpha^*)=-\varepsilon\sum_{i=1}^{n}(\alpha_i^*+\alpha_i)+\sum_{i=1}^{n}y_i(\alpha_i^*-\alpha_i)-$$
$$\frac{1}{2}\Big(\sum_{i,j=1}^{n}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)K(\boldsymbol{x}_i\cdot\boldsymbol{x}_j)+\frac{1}{C}\sum_{i=1}^{n}(\alpha_i^*)^2+\frac{1}{C}\sum_{i=1}^{n}(\alpha_i)^2\Big) \tag{2a}$$

$$\text{s.t.}\ \sum_{i=1}^{n}\alpha_i^*=\sum_{i=1}^{n}\alpha_i,\quad 0\leqslant\alpha_i^*\leqslant C,\ 0\leqslant\alpha_i\leqslant C,\ i=1,\cdots n \tag{2b}$$

所求的权向量 $\boldsymbol{w}$ 为

$$\boldsymbol{w}=\sum_{i=1}^{n}(\alpha_i^*-\alpha_i)\boldsymbol{x}_i \tag{2c}$$

普通克立格方法解决的问题可以表示为已知训练样本对$(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_n,y_n)$,目

的是根据这些训练样本对，对局部空间区域内的任何一个点 $\boldsymbol{x}$ 给出估计值 y，并且估计值要写成已知样本点加权和的形式 $f(\boldsymbol{x})=\sum_{i=1}^{n}\lambda_i y_i$. 对于给出的估计而言，该估计量应该是线性无偏，并且估计方差最小. 可以证明通过选择 λ_i 使其满足 $\sum_{i=1}^{n}\lambda_i=1$，则得到的估计量必是线性无偏的. 而欲使估计方差为最小，需要使误差的方差最小化，即满足条件 $\min E[y-f(\boldsymbol{x})]^2$. 这是一个等式约束的最小化问题，可以借助拉格朗日乘子法求出其中的 λ[5]

$$\lambda=\Sigma^{-1}(C+\mu I)=\Sigma^{-1}C+\mu\Sigma^{-1}I \tag{3}$$

将求得的 λ_i 代入估计式得到

$$\boldsymbol{y}=f(\boldsymbol{x})=\lambda^{\mathrm{T}}Y=(\Sigma^{-1}C+\mu\Sigma^{-1}I)^{\mathrm{T}}Y=C'(\Sigma^{-1})^{\mathrm{T}}Y+(\mu\Sigma^{-1}I)^{\mathrm{T}}Y \tag{4}$$

其中 Σ 为随机向量 $\boldsymbol{y}$ 的协方差阵，在二阶平稳假设下可以写成变量 $\boldsymbol{x}$ 的函数，即 $\mathrm{cov}(f(\boldsymbol{x}_i),f(\boldsymbol{x}_j))=C(\boldsymbol{x}_i,\boldsymbol{x}_j)$，$C=[C(\boldsymbol{x},\boldsymbol{x}_1),C(\boldsymbol{x},\boldsymbol{x}_2),\cdots,C(\boldsymbol{x},\boldsymbol{x}_n)]^{\mathrm{T}}$. 如果假设 $Ef(\boldsymbol{x})=0$ 成立，可知约束 $\sum_{i=1}^{n}\lambda_i=1$ 将不起作用，此时(4)式变为

$$\boldsymbol{y}=f(\boldsymbol{x})=C^{\mathrm{T}}(\Sigma^{-1})^{\mathrm{T}}Y \tag{5}$$

可以证明，在 $Ef(\boldsymbol{x})=0$ 假设下，如果令 $K(\boldsymbol{x},\boldsymbol{y})=C(\boldsymbol{x},\boldsymbol{y})$，$\varepsilon=0$，$C->\infty$，则由SVM方法得到的函数估计结果与由普通克立格方法得到的结果是等价的，即 $\boldsymbol{y}=f(\boldsymbol{x})=C^{\mathrm{T}}(\Sigma^{-1})^{\mathrm{T}}Y$. 如果令 $\varepsilon=0$，则 $-\varepsilon\sum_{i=1}^{n}(\alpha_i^*+\alpha_i)=0$，如 C 取一个比较大的值，则 $\left(\frac{1}{C}\sum_{i=1}^{n}(\alpha_i^*)^2+\frac{1}{C}\sum_{i=1}^{n}(\alpha_i)^2\right)\to 0$. 等于约束条件(2b)可以忽略，此时对偶问题(2)成为

$$\max\quad W(\alpha,\alpha^*)=\sum_{i=1}^{n}y_i(\alpha_i^*-\alpha_i)-\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)K(\boldsymbol{x}_i,\boldsymbol{x}_j) \tag{6a}$$

$$\text{s.t.}\quad \sum_{i=1}^{n}(\alpha_i-\alpha_i^*)=0 \tag{6b}$$

$\varepsilon=0$ 的结果使得所有偏离监督信号的输出都受到了惩罚($\varepsilon\neq 0$ 时所有偏离监督信号的输出如果小于 ε 将不受惩罚)，而 C 取一个比较大的值，使得对拟合误差的惩罚在总惩罚项(1a)占较大的比例. 采用这两项简化后得到的二次规划问题(6)仍然对应着问题(2)，得到的解仍然具有结构风险最小化的特点. 采用拉格朗日方法，约束优化问题(6)转化为最小化

$$\min L(\alpha_i,\alpha_i^*,\lambda)=\sum_{i=1}^{n}y_i(\alpha_i^*-\alpha_i)-\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)K(\boldsymbol{x}_i,\boldsymbol{x}_j)-\lambda\sum_{i=1}^{n}(\alpha_i-\alpha_i^*) \tag{7}$$

令 $\beta_i=\alpha_i^*-\alpha_i(i=1,\cdots,n)$，将其代入(7)式得

$$\min L(\beta,\lambda)=\sum_{i=1}^{n}y_i\beta_j-\frac{1}{2}\sum_{i,j=1}^{n}\beta_i\beta_jK(\boldsymbol{x}_i\cdot\boldsymbol{x}_j)-\lambda\sum_{i=1}^{n}\beta_i \tag{8}$$

(8)式相对于 β_i，λ 求导，并令导数为0，则得

$$Y-H\beta-\lambda I=0 \tag{9a}$$

$$\sum_{i=1}^{n}\beta_i=0\Leftrightarrow\beta' I=0 \tag{9b}$$

其中 $Y=\begin{bmatrix}y_1\\y_2\\\cdots\\y_n\end{bmatrix}$，$H=\begin{bmatrix}K(\boldsymbol{x}_1,\boldsymbol{x}_1) & K(\boldsymbol{x}_1,\boldsymbol{x}_2) & \cdots & K(\boldsymbol{x}_1,\boldsymbol{x}_n)\\K(\boldsymbol{x}_2,\boldsymbol{x}_1) & K(\boldsymbol{x}_2,\boldsymbol{x}_2) & \cdots & K(\boldsymbol{x}_2,\boldsymbol{x}_n)\\\cdots & \cdots & \cdots & \cdots\\K(\boldsymbol{x}_n,\boldsymbol{x}_1) & K(\boldsymbol{x}_n,\boldsymbol{x}_2) & \cdots & K(\boldsymbol{x}_n,\boldsymbol{x}_n)\end{bmatrix}$，$\beta=\begin{bmatrix}\beta_1\\\beta_2\\\cdots\\\beta_n\end{bmatrix}$.

如果 H 为正定阵,则 H 的逆存在并唯一.(9a)式两边同时乘以 H^{-1}并化简后得

$$\beta = H^{-1}Y - \lambda H^{-1}I = H^{-1}(Y - \lambda I) \tag{10}$$

将(10)式代入 $\boldsymbol{w}$ 的表达式(2c)得

$$\boldsymbol{w} = \sum_{i=1}^{n}\beta_i x_i = X\beta = X(H^{-1}(Y - \lambda I)) \tag{11}$$

故得到的拟合函数表达式为

$$\begin{aligned} f(\boldsymbol{x}) &= K(\boldsymbol{x},\boldsymbol{w}) + b = K(\boldsymbol{x}, XH^{-1}(Y - \lambda I)) + b = \\ &[K(\boldsymbol{x}_1,\boldsymbol{x}), K(\boldsymbol{x}_2,\boldsymbol{x}), \cdots, K(\boldsymbol{x}_n, x)](H^{-1}(Y - \lambda I)) + b \end{aligned} \tag{12}$$

如果 $Ef(\boldsymbol{x})=0$,则在二维情况下可以认为样本是均值为 0,以原点为中心的分布(可以通过对数据进行预处理得到).估计问题变成在高维空间估计一条过原点的超平面.在这个假设下 $b=0$, $f(\boldsymbol{x})=K(\boldsymbol{w},\boldsymbol{x})$.如果 $b=0$,这时约束条件(2b)将消失.经过以上简化后,则(12)式成为

$$f(\boldsymbol{x}) = [K(\boldsymbol{x}_1,\boldsymbol{x}),\ K(\boldsymbol{x}_2,\boldsymbol{x}), \cdots, K(\boldsymbol{x}_n,\boldsymbol{x})]H^{-1}Y \tag{13}$$

若令 $K(\boldsymbol{x},\boldsymbol{x}_i)=C(\boldsymbol{x},\boldsymbol{x}_i)$则有 $H=\Sigma$.将其代入(13)式得

$$f(\boldsymbol{x}) = [C(\boldsymbol{x}_1,\boldsymbol{x}), C(\boldsymbol{x}_2,\boldsymbol{x}), \cdots, C(\boldsymbol{x}_n,\boldsymbol{x})](\Sigma^{-1}Y) = C^{\mathrm{T}}(\Sigma^{-1})^{\mathrm{T}}Y \tag{14}$$

这样,采用 SVM 方法得到的函数回归形式(14)与克立格方法得到的函数回归形式便相互等价了.以上在(14)式中使用了$(\Sigma^{-1})^{\mathrm{T}}=(\Sigma^{\mathrm{T}})^{-1}$的假设,可以证明对于对称正定阵,该假设成立.如果假设条件 $Ef(\boldsymbol{x})=0$ 不满足,两种方法得出的结论最多只相差一个常数.这就说明,如果在 SVM 中用协方差函数 $C(\boldsymbol{x},\boldsymbol{y})$代替核函数 $K(\boldsymbol{x},\boldsymbol{y})$,这样得到的结果将具有克立格方法解的性质.因此,我们可以考虑在 SVM 中用协方差函数代替核函数 $K(\boldsymbol{x},\boldsymbol{y})$,这样得到的解将具有均方最小的意义.

3　用原始样本集合求取径向基核函数的参数

在用支持向量机方法进行函数拟合时,径向基核函数经常被采用,其中的宽度参数 σ 对拟合的结果影响较大.由于目前对 σ 的取值一直是根据经验与试算,因此我们希望能够借助上述的等价性研究来解决该问题.我们尝试提出利用原始样本集合求取径向基核函数的宽度参数,用该径向基函数作为 SVM 的核函数.如果样本均值为零,则样本的方差为 1,样本的协方差函数 $C(h)<=1$.又由于函数 $\exp(h,\sigma)<=1$,故可以令 $C(h)=\exp(h,\sigma)$,从而确定宽度参数 δ.可以在这里假设变量满足二阶平稳假设,根据样本均值是否为 0,分两种情况对径向基函数进行参数估计:

1) 假设样本的均值为 0,则协方差函数满足 $0\leqslant C(h)\leqslant 1$,$C(0)=1$,

$$C(h) = \exp\left(-\frac{h^2}{2\sigma_1^2}\right) \tag{15}$$

2) 假设样本的均值不为 0,即 $0\leqslant C(h)\leqslant\infty$, $C(0)=\alpha$,

$$C(h) = \alpha\exp\left(-\frac{h^2}{2\sigma_1^2}\right) \tag{16}$$

所以,如果事先不对样本进行零均值化处理,则用第二种情况直接估计即可.如果事先

已知样本为零均值,则用第一种情况直接估计. 由于在许多实际工作中,协方差函数并不存在,例如一些自然现象和随机函数,它们具有无限离散性,即无协方差和先验方差. 此时采用(15)和(16)式进行估计将没有意义,而采用变异函数[1)]进行估计. 在一维条件下的变异函数的定义为

$$r(\boldsymbol{x},h)=\frac{1}{2}\operatorname{var}[z(\boldsymbol{x}+h)-z(\boldsymbol{x})]=$$
$$\frac{1}{2}E[z(\boldsymbol{x}+h)-z(\boldsymbol{x})]^2-\frac{1}{2}\{E[z(\boldsymbol{x}+h)]-E[z(\boldsymbol{x})]\}^2 \tag{17}$$

当变异函数 $r(\boldsymbol{x},h)$ 与位置 $\boldsymbol{x}$ 无关,而只依赖于分隔两个样品点之间的距离 h 时,则有

$$r(\boldsymbol{x},h)=\frac{1}{2}E[z(\boldsymbol{x}+h)-z(\boldsymbol{x})]^2 \tag{18}$$

在二阶平稳假设下,可以推出

$$r(h)=C(0)-C(h) \tag{19}$$

如果分别将(15),(16)式代入(19)式,则得

$$r(h)=C(0)-C(h)=1-\exp\left(-\frac{h^2}{2\sigma_1^2}\right) \tag{20a}$$

$$r(h)=C(0)-C(h)=\alpha-\alpha\exp\left(-\frac{h^2}{2\sigma_1^2}\right)=\alpha\left[1-\exp\left(-\frac{h^2}{2\sigma_1^2}\right)\right] \tag{20b}$$

因此,采用 $r(h)$ 估计 σ 与用 $C(h)$ 估计 σ 是等价的,而 $r(h)$ 正是地质统计学中定义的变异函数. 由于变异函数可以在更宽松的条件下得到,并且在统计学中关于 $r(h)$ 的估计存在经验公式可以直接套用[5],因此选择变异函数进行核函数的估计较协方差函数更为方便. 如果知道核函数的类型,可以选择适当的变异函数模型进行核函数参数的估计. 这里我们选择了高斯模型作为变异函数的理论模型,在得到了模型参数的基础上进一步估计核函数参数. 高斯模型[5]的定义为

$$r(h)=\begin{cases}0, & h=0\\ C_0+C\left(1-e^{-\frac{h^2}{a^2}}\right), & h>0\end{cases} \tag{21}$$

若变量均值为 0,则该模型与(20)式完全吻合. 这样,在得到了变异函数模型的基础上,通过估计变异函数,就可以求得径向基函数的宽度参数 σ.

4 结论

本文探讨了 SVM 在条件($Ef(\boldsymbol{x})=0,\ \varepsilon=0,\ C\to\infty$)下,用协方差函数代替核函数时解的具体形式,同时也推导了克立格方法在 $Ef(\boldsymbol{x})=0$ 条件下解的具体形式. 通过对比发现,二者存在着等价性. 在此基础上,分析了一般情况下在 SVM 中用协方差函数代替核函数进行 SVM 拟合的可行性. 考虑到在一些情况下协方差函数可能不存在,提出了利用变异函数进行核函数参数估计的方法,即首先利用变异函数估计径向基核函数宽度参数,然后利用该核函数进行 SVM 拟合. 最后证明了该方法等价于用协方差函数进行参数估计的方法.

1) 经典统计学的缺陷之一就是没有考虑各样本的空间位置,因此即使两批样本的均值和方差都相同,也可能其空间分布很不相同. 为了弥补这一缺陷,在地质统计学中引入了变异函数这一概念,它能够反映区域化变量的空间变化特征,特别是透过随机性反映区域化变量的随机性. 变异函数被称作变差函数和结构函数.

参 考 文 献

1 Vapnik V N. The Nature of Statistical Learning Theory. NY: Springer-Verlag, 1995

2 Cortes C, Vapnik V N. Support-vector networks. *Machine Learning*, 1995, **20**:273～297

3 Burges C J C. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 1998, **2**(2):110～115

4 侯景儒,尹镇南等. 实用地质统计学. 北京: 地质出版社, 1998

5 张学工. 关于统计学习理论与支持向量机. 自动化学报, 2000, **26**(1):32～42

阎　辉　现在清华大学自动化系攻读博士学位. 主要研究方向为模式识别、信号处理理论及其应用.

张学工　博士,现为清华大学自动化系信息技术研究所所长. 主要研究方向为模式识别理论及其应用.

(上接第 377 页)

中国自动化学会 2002 年一般专题学术活动计划

项目名称	主要内容	时间	地点	联系人
系统仿真技术及其应用学术交流会	21 世纪系统仿真技术展望、国内外仿真系统和软件发展方向;连续过程建模与仿真;建模和仿真方法、DEDS 仿真技术、定性仿真技术;航空航天、武器装备、仿真技术及各类仿真系统;应用仿真结构的未来等	3 季度	安徽	合肥市中国科学技术大学自动化系　王雷、戴耀华 电话:(0551)3601514 3606104 邮编:230001 E-mail: chenzh@ustc.edu.cn
全国第十届空间及运动体控制技术学术会议	控制理论、方法及其在空间及运动体控制中的应用;飞行器制导、导航与控制技术;空间实验室动力学和控制技术;航天器交会对接技术;空间机器人等	8 月	湖南	北京控制工程研究所(北京 2729 信箱)科技委　李静 电话:68378650 邮编:100080
全国办公自动化学术会议	电子政府发展趋势、展望以及对信息化发展的促进作用;法律与政策环境研究、技术标准与管理规范;发展模式及应用实例分析等	7 月	沈阳	沈阳浑南高新技术产业开发区——东大软件园东软集团科研部 李春山、高昕 电话:(024)23783377 邮编:110179 E-mail: lics@neusoft.com; Gaoxin@neusoft.com
第十七届青年学术年会	年会主题:未来自动化领域的机遇与挑战	7 月	北戴河	河北省秦皇岛市燕山大学电气工程学院　关新平、罗小元 电话:(0335)8057041　8057034 邮编:066004 E-mail: xpguan@ysu.edu.cn
全国自动化教育学术年会	自动化教育的发展、科研、人才培养以及教育经验等方面的研讨	4 季度	待定	北京清华大学自动化系　肖德云 电话:62785845 邮编:100084
中国机器人竞赛暨研讨会	FIRA 机器人比赛,机器人体系结构、传感技术、视觉、规划与决策技术、协作、竞争与学习等	6 月	上海	北京清华大学计算机系　钱宗华 电话:62788939 邮编:100084
重庆市自动化与仪器仪表学会学术交流会	产品展示、专题报告、新产品介绍	3 季度	重庆	重庆市 1506 信箱　孙怀义 电话:(023)68268572 邮编:400708