

尺度核函数支撑矢量机

张 莉,周伟达,焦李成

(西安电子科技大学雷达信号处理重点实验室,陕西西安 710071)

摘 要: 本文提出了一种可容许的支撑矢量机核—尺度核. 该尺度核函数可以被看作是一个具有平移因子的多维尺度函数,它可作为平方可积空间的子空间上一组完备的基函数. 在此意义上,采用尺度核函数的支撑矢量机,可以认为是在尺度空间中寻找最佳的尺度系数. 因此在理论上尺度核函数支撑矢量机能够以零误差逼近某一空间上的任何目标函数,文中给出的仿真实验进一步验证了它的可行性和有效性.

关键词: 支撑矢量机; 支撑矢量核; 尺度核函数; 高斯核

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2002) 04-0527-03

Scaling Kernel Function Support Vector Machines

ZHANG Li, ZHOU Wei-da, JIAO Li-cheng

(National Key Lab for Radar Signal Processing, Xidian University, Xi'an, Shaanxi 710071, China)

Abstract: A kind of admissible support vector kernel called scaling kernels are presented in this paper. In fact, scaling kernels are the multi-dimensional scaling function with translation vectors and they are a set of complete bases in the sub-space of the square and integrable space. Hence, the goal of scaling kernel function support vector machines is to find the optimal scaling coefficients in a scaling space. In terms of theory, scaling kernel function SVMs can approximate any objective function in some space by any precision. The results obtained by our simulations show the feasibility and validity of scaling kernels.

Key words: support vector machine; support vector kernel; scaling kernel; gaussian kernel

1 引言

无论是对模式识别,还是对回归分析来说,支撑矢量机的非线性处理能力都是通过“核映射”^[1-3]的方法来实现的. “核映射”把输入空间的样本映射到高维的特征空间 H , 在 H 空间中,我们只需进行线性的识别或回归就能达到非线性处理的目的. 但上述的核函数有一限制,须满足 Mercer 条件^[4,5]. 已有的如高斯核函数就是一种普遍使用的支撑矢量核,它在模式识别和回归分析中都表现出了良好的映射性能. 但是我们在实际运用中发现,对于目前已有的支撑矢量核,支撑矢量机是不可能逼近某一 $L_2(R^N)$ (平方可积空间)子空间上的任意函数的,因为已有的核函数通过平移不可能生成该子空间上的一组完备的基. 这种基的不完备性导致了分类支撑矢量机不能逼近该子空间上的任意分类界面,同样回归支撑矢量机也不能逼近任意的目标曲线.

基于以上的考虑,我们需要寻找一种核函数,这种核函数仅通过平移就能生成 L_2 空间上的一组完备的基. 其实这样的函数上早已给出,就是尺度函数. 本文证明了由尺度函数构成的尺度核函数是满足 Mercer 条件的,它可以作为支撑矢量机核函数来使用. 尺度核函数实际上就是具有平移因子的多维尺度函数,因此尺度核函数就是一组完备的基. 尺度核函数支撑矢量机,就相当于在以多维尺度函数及其平移为基张成的尺度空间中,寻找问题的最佳逼近系数,即支撑矢量机中的 Lagrange 乘子. 我们用仿真验证了尺度核函数的逼近性能,并对尺度核函数和高斯核函数的逼近能力进行了比较. 结论表明尺度核的逼近性能相当好. 本文提出的尺度核函数也可

用在分类问题及其他可用支撑矢量算法解决的问题中.

2 支撑矢量机及支撑矢量核

2.1 支撑矢量机

我们以回归支撑矢量机为例来简单地回顾支撑矢量机. 在回归支撑矢量机中,我们利用某种非线性映射,把输入空间的样本 x 映射到某一特征空间 H 中. 在这个特征空间中,我们可以进行线性回归^[1,3]分析:

$$f(x) = w \cdot (x) + b \quad (1)$$

其中 $b \in R$ 表示阈值, $x \in R^N$, $w \in H$, $\cdot: R^N \rightarrow H$. 这里要求输入空间的训练样本 $\{(x_i, y_i) | x_i \in R^N, y_i \in R, i = 1, \dots, l\}$ 是独立同分布的. 因此在高维特征空间中的线性回归对应于输入空间的非线性回归. 式(1)中的 $w \cdot (x)$ 需要在特征空间中计算,同时对于实际一些复杂的应用,我们希望 (x) 能把输入空间映射到一个较高的高维空间,这样将带来计算量的迅速增加. 而支撑矢量机巧妙地通过引入支撑矢量核

$$K(x_i, x_j) = (x_i) \cdot (x_j) \quad (2)$$

解决了这一问题. 从后面可以看到,正是由于支撑矢量核的应用,我们才能够避免对式(1)的直接计算.

回归支撑矢量机是即以最小风险代价找到函数 f . Vapnik 建议采用下式来最小化风险代价来得到函数 f ^[3,5,6]:

$$\min \frac{1}{2} \|w\|^2 + \frac{C}{l} \sum_{i=1}^l |y_i - f(x_i)| \quad (3)$$

其中 $C > 0$ 是惩罚因子. 表达式(3)中的第二项为 ϵ -insensitive 损失函数:

$$|y - f(x)| = \begin{cases} 0, & \text{当 } |y - f(x)| \leq \epsilon \\ |y - f(x)| - \epsilon, & \text{其他情况} \end{cases} \quad (4)$$

收稿日期:2000-09-18;修回日期:2001-08-17

基金项目:国家“863”计划(No. 863-306-06-06-1);教育部博士点基金项目资助

在确定了核函数后,支撑矢量机把上述问题最终归结为如下的一个半正定凸二次规划问题:

$$\begin{aligned} \max \quad & W(\lambda^*) = - \sum_{i=1}^l (\lambda_i^* + \mu_i) + \sum_{i=1}^l (\lambda_i^* - \mu_i) y_i \\ & - \frac{1}{2} \sum_{i,j=1}^l (\lambda_i^* - \mu_i) (\lambda_j^* - \mu_j) K(x_i, x_j) \quad (5) \\ \text{s. t.} \quad & \sum_{i=1}^l (\lambda_i^* - \mu_i) = 0 \\ & \lambda_i^* \in [0, \frac{c}{l}] \end{aligned}$$

其中 λ_i^* 是相应训练样本 x_i 的 Lagrange 乘子. 为了简略, λ_i^* 表示带星号和不带星号的 Lagrange 乘子, ν 为 Vapnik 定义的 ν -insensitive 损失函数阈值. 同时得到:

$$w = \sum_{i=1}^l (\lambda_i^* - \mu_i) \phi(x_i) \quad (6)$$

把式(6)代入到(1)中,有

$$f(x) = \sum_{i=1}^l (\lambda_i^* - \mu_i) K(x, x_i) + b \quad (7)$$

从式(7)中,我们不仅可以看到非线性映射 ϕ 不再出现,取而代之的是核函数 K ,我们即使不知道非线性映射 ϕ 的具体形式,也能进行非线性处理. 同时我们从函数逼近的角度还能发现, $K(x, x_i)$ 可以看成是一组基函数(不一定正交),对目标函数的逼近可以用这一组基函数的线性组合再加上一截距 b 来完成,留下的问题就是寻找一组线性组合的系数 $(\lambda_i^* - \mu_i)$. 这样支撑矢量机就有了一种逼近意义上的解释,在支撑矢量处,须用一个中心在该支撑矢量处的基函数来逼近,而在非支撑矢量处,可以用周围支撑矢量点上的基函数的旁瓣来叠加而成.

2.2 支撑矢量核

支撑矢量核可以是特征空间中的点积形式,如 $K(x, x) = K(\langle x, x \rangle)$,也可以是平移不变形式 $K(x, x) = K(x - x)$ ^[5]. 实际上一个函数只要满足 Mercer 条件,这个函数就是一个可容许的支撑矢量核函数.

Mercer 条件^[4,5]可以写成

$$\iint_{\mathcal{O}L_2} K(x, x) g(x) g(x) dx dx \geq 0 \quad (8)$$

其中 $\forall g(x) \in L_2(R^N), K(x, x) \in L_2(R^N) \otimes L_2(R^N)$.

如果一个函数 K 满足式(11),则它可以被分解为两个函数的点积,如式(2).

对于平移不变函数,由于很难将之分解为两个相同函数的点积形式,所以文献[4]提供了另外的条件:如果一个函数 $K(x)$ 的 Fourier 变换满足下式

$$F[K](\omega) = (2\pi)^{-N/2} \int_{R^N} \exp(-i(\omega \cdot x)) K(x) dx \geq 0 \quad (9)$$

其中 $x \in R^N, i = \sqrt{-1}$,那么该函数就是一种可容许的支撑矢量核,且有如下形式

$$K(x, x) = K(x - x) \quad (10)$$

我们可以利用这两个条件来构造可容许的支撑矢量核.

3 尺度核函数

在子波理论中,尺度空间,尺度函数,子波空间,子波函数

都是非常重要的概念^[7,8]. 当 N 充分大时,近似地有 $L_2(R) = V_N$. 因此可以对 $L_2(R)$ 空间进行分解:

$$\begin{aligned} L_2(R) &= V_N = V_{N-1} \oplus W_{N-1} = V_{N-2} \oplus W_{N-2} \oplus W_{N-1} = \dots \\ &= V_0 \oplus W_0 \oplus W_1 \oplus \dots \oplus W_{N-2} \oplus W_{N-1} \\ &= V \oplus W \oplus \dots \oplus W_{N-1} \oplus W_0 \oplus W_1 \oplus \dots \oplus W_{N-2} \oplus W_{N-1} \end{aligned}$$

这样如果我们将 V_N 空间的尺度函数作为基函数,由尺度函数的性质可知,可以通过平移来生成 V_N 空间中的一组完备的基,则在此空间中的任意函数都可以表示为尺度函数的线性组合.

若记一维尺度函数为 $\phi(x)$,那么可分离的多维尺度函数的特殊情况可以写成^[8]

$$\phi(x) = \prod_{i=1}^N \phi(x^i) \quad (11)$$

其中 $x = (x^1, \dots, x^N) \in R^N$. 每个一维尺度函数都要满足容许条件^[7,8]

$$\int \phi(x) dx = 1 \quad (12)$$

我们可以构造这样的平移不变尺度核函数:

$$K(x, x) = \prod_{i=1}^N \phi(x^i - x^i) \quad (13)$$

下面我们给出一个具体的尺度函数:Shannon 尺度函数^[7]

$$\phi(x) = \frac{\sin(px)}{px} \quad (14)$$

并用它来构造尺度核函数

$$K(x, x) = \prod_{i=1}^N \frac{\sin p(x^i - x^i)}{p(x^i - x^i)} \quad (15)$$

其中 $p > 0$ 是用户自定义的参数. 实际上,尺度核函数式(13)就是一个带有平移项的多维尺度函数. 我们能够证明式(13)满足平移不变核条件(9),因为该每个一维尺度函数的 Fourier 变换为:

$$\phi(\omega) = \begin{cases} 1/p, & -p \leq \omega < p \\ 0, & \text{其他} \end{cases} \quad (16)$$

采用尺度核函数的支撑矢量机的估计函数为:

$$f(x) = \sum_{i=1}^l (\lambda_i^* - \mu_i) \prod_{j=1}^N \frac{\sin p(x^j - x_j^j)}{p(x^j - x_j^j)} + b \quad (17)$$

其中 x_j^i 表示第 i 个训练样本的第 j 个分量. 由于尺度核的使用,使得估计函数(17)式有了另一种意义:以多维尺度函数在支撑矢量上的平移作为基函数,在尺度空间中寻找问题的最佳逼近的尺度系数,即 Lagrange 乘子 (λ_i^*) .

4 仿真实验

为了验证尺度核函数的逼近性能,我们逼近一个一维的分段函数:

$$f(x) = \begin{cases} -2.186x - 12.864, & -10 \leq x < -2 \\ 4.246x, & -2 \leq x < 0 \\ 10e^{-0.05x-0.5} \sin((0.03x+0.7)x), & 0 \leq x \leq 10 \end{cases} \quad (18)$$

和一个二维函数

$$f(x) = (x_1^2 - x_2^2) \sin(0.5x_1) \quad (19)$$

我们采用文[9]中所定义的逼近误差:

$$= \sqrt{\frac{\sum_{i=1}^l (y_i - \hat{y}_i)^2}{\sum_{i=1}^l (y_i - \bar{y}_i)^2}}, \text{ 且 } \bar{y} = \frac{1}{l} \sum_{i=1}^l y_i \quad (20)$$

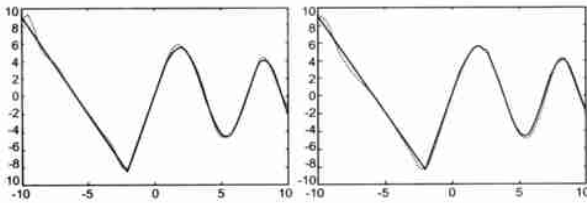


图 1 尺度核对一维函数的逼近曲线 图 2 高斯核对一维函数的逼近曲线

其中 y 表示原函数值, f 表示逼近值.

我们用支撑矢量机对这两个函数进行逼近. 为了便于比较, 我们分别采用了高斯核和尺度核函数作为支撑矢量核. 高斯核是迄今为止找到的最佳支撑矢量核, 因为经高斯核映射

后的特征空间是无穷维的. 高斯核的表达式如下

$$K(x, x) = \exp\left(-\frac{x-x}{2p^2}\right) \quad (21)$$

其中, $p > 0$ 是用户自定义的参数.

对函数式(18)的实验结果见表 1 以及图 1—2. 图 1 是尺度核对一维函数逼近结果, 图 2 是高斯核的逼近结果. 图中实线表示原函数, 虚线表示逼近曲线. 图 3 是二维函数(19)的曲线. 图 4 和图 5 分别是尺度核和高斯核函数对二维函数的逼近曲线. 对二维函数逼近的结果见表 2.

从我们的实验结果来看, 采用尺度核函数得到的逼近误差和采用高斯核得到的逼近误差有相同的数量级, 这说明尺度核和高斯核在逼近能力上是相似的. 也验证了这里所构造的尺度核的确是一种可容许支撑矢量核.



图 3 原函数曲线

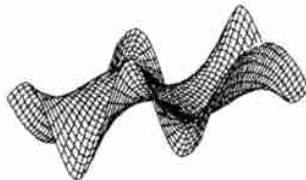


图 4 尺度核对二维函数的逼近曲线

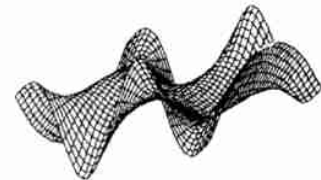


图 5 高斯核对二维函数的逼近曲线

表 1 对一维函数的的逼近结果

核函数	核参数 p	迭代次数	逼近误差
尺度核	0.9158	2000	0.0768
高斯核	1.000	2000	0.0760

表 2 对二维函数的的逼近结果

核函数	核参数 p	迭代次数	逼近误差
尺度核	0.3	2000	0.0171
高斯核	2.9	2000	0.0320

据不同需要, 选用不同的尺度核函数. 我们的仿真结果证实了以上得到的结论.

参考文献:

[1] B E Boser, I M Guyon, V N Vapnik. A training algorithm for optimal margin classifiers [A]. In D. Haussler, editor, Proceedings of the 5th Annual ACM Working on Computational Learning Theory [C]. Pittsburgh, PA: ACM Press, 1992. 144 - 152.

[2] C J C Burges. A tutorial on support vector machines for pattern recognition [J]. Data Mining and Knowledge Discovery, 1998, 2(2) : 1 - 47.

[3] V Vapnik. The Nature of Statistical Learning Theory [M]. New York: Springer Verlag, 1995.

[4] C J C Burges. Geometry and invariance in kernel based methods [A]. in Advance in Kernel Methods-Support Vector Learning [C]. Cambridge, MA: MIT Press, 1999. 89 - 116.

[5] A Smola, B Schölkopf. A tutorial on support vector regression [R]. NeuroCOLT, Rep. 19, 1998.

[6] B Schölkopf, A J Smola, R Williamson, P Bartlett. New support vector algorithms [R]. NeuroCOLT2 Technical Report Series, 1998.

[7] 赵松年, 熊小芸. 子波变换与子波分析 [M]. 北京: 电子工业出版社, 1997.

[8] 张贤达, 保铮. 非平稳信号分析与处理 [M]. 北京: 国防工业出版社, 1998.

[9] Q H Zhang, A Benveniste. Wavelet networks [J]. IEEE Trans Neural Networks, 1992, 3(6) : 889 - 898.

作者简介:

张 莉 女, 1975 年生于贵州省, 西安电子科技大学在读博士生, 主要研究方向有数据挖掘、模式识别和神经网络. E-mail: zhangli@rsp.xidian.edu.cn

周伟达 男, 1974 年生于浙江省, 西安电子科技大学在读博士, 研究领域包括: 机器学习、统计学习理论和智能信号处理.

5 结论

本文给出了以多维尺度函数为核函数的支撑矢量机方法, 由于我们可以构造足够大的尺度空间, 这个空间达到几乎与 $L_2(R)$ 空间重合, 那么在该尺度中所取的一组完备的尺度函数基, 就可以看成是 $L_2(R)$ 空间上的一组完备的基. 这样 $L_2(R)$ 空间的任意函数均可以被这组基线性表出. 对于一般的模式识别问题和回归分析, 所要逼近的函数一般是定义域有限的, 所以他们都可以看成是属于 $L_2(R)$ 空间的. 这样我们可以得到如下的结论: 对于尺度核函数, 不管是分类支撑矢量机, 还是回归支撑矢量机, 在理论上它们都可以以任意精度逼近分界面或目标曲线, 同时又因为支撑矢量机本身的算法保证了支撑矢量机的优化不仅收敛而且能够收敛到最优解上, 从而保证了尺度核函数支撑矢量机的收敛性.

在本文中, 我们以 Shannon 尺度函数为例, 一方面是因为 Shannon 尺度函数具有平移正交性, 同时因为 Shannon 尺度函数具有理想的频域低通性, 通带幅度恒定, 阻带为 0. 经过时移, 这种低通性就变成了带通性, 所以 Shannon 尺度核函数支撑矢量机在信号处理中也将会有很好的运用. 但是由于 Shannon 尺度函数在时间上虽衰减较快, 但不是严格的局域函数, 在远离中心处仍有一定幅度的振荡, 所以若需以很高的精度逼近一个局域函数, 效率不是很高. 在实际运用中, 我们可根