

# 处理非线性分类和回归问题的一种新方法(I) ——支持向量机方法简介\*

陈永义 俞小鼎 高学浩

(中国气象局培训中心, 北京 100081)

冯汉中\*\*

(成都气象中心, 成都 610071)

## 摘 要

简要介绍了近年来倍受瞩目的一种处理高度非线性分类、回归等问题的计算机学习的新方法——支持向量机(SVM)方法;分析了这一方法的特点及其在数值预报产品释用及气象研究业务中的应用前景。SVM是一种有坚实理论基础的新颖的小样本学习方法。它基本上不涉及概率测度及大数定律等,因此不同于现有的统计方法。从本质上看,它避开了从归纳到演绎的传统过程,实现了高效的从训练样本到预报样本的“转导推理”(transductive inference),大大简化了通常的分类和回归等问题。SVM的最终决策函数只由少数的支持向量所确定,计算的复杂性取决于支持向量的数目,而不是样本空间的维数,这在某种意义上避免了“维数灾”。

关键词: 机器学习 支持向量机(SVM) 模式识别 回归估计 天气预报

## 引 言

用人类的思维揭示学习的秘密是一个历史性的难题。如果说在几十年前它还只是少数哲学大师和学术先觉的思考课题,那么由于信息技术的迅猛发展和计算机的广泛应用,学习问题,特别是机器学习问题已成为广大研究、技术人员必须面对的实际问题。

学习的本质是归纳,对此经典逻辑很少办法。模式识别、回归分析、密度估计是学习的三个基本内容。学者们应用概率统计和函数逼近等方法取得了很多研究成果,特别是在处理线性问题上,有些甚至是完满的。但对于本质上非线性的问题还缺少较好的结果。

当我们面对数据而又缺乏理论模型时,统计分析方法是最先采用的方法。然而传统的统计方法只有在样本数量趋于无穷大时才能有理论上的保证。而在实际应用中样本数目通常都是有限的,甚至是小样本,对此基于大数定律的传统统计方法难以取得理想的效果。

Vapnik 等人提出的统计学习理论<sup>[1,2]</sup>是一种专门的小样本理论,它避免了人工神经网络等方法的网络结构难于确定、过学习和欠学习以及局部极小等问题,被认为是目前针对小样本的分类、回归等问题的最佳理论。这一方法数学推导严密,理论基础坚实。基于

\* 本研究得到国家自然科学基金资助(60072006)。

\*\* 2002 年中国气象局培训中心访问学者。

2003-01-25 收到,2003-09-17 收到修改稿。

这一理论近年提出的支持向量机(Support Vector Machines 简称 SVM)方法,为解决非线性问题提供了一个新思路。

早在 20 世纪 60 年代, Vapnik 就已完成了统计学习的基本理论结果,如经验风险最小化原则下统计学习一致性的条件(收敛性、收敛的可控性、收敛与概率测度定义的无关性,号称机器学习理论的“三个里程碑”)、关于统计学习方法推广能力的界的理论以及在此基础上建立的小样本归纳推理原则等。直到 20 世纪 90 年代中期,实现统计学习理论和原则的实用化算法——SVM 方法才被逐渐完整提出<sup>[3,4]</sup>,并在模式识别等人工智能领域成功应用,受到广泛关注。

SVM 方法的基本思想是:基于 1909 年 Mercer 核展开定理<sup>[5]</sup>,可以通过非线性映射  $\varphi$ ,把样本空间映射到一个高维乃至无穷维的特征空间(Hilbert 空间),使在特征空间中可以应用线性学习机的方法解决样本空间中的高度非线性分类和回归等问题。

降维(即把样本空间向低维空间做投影)是人们处理复杂问题常用的简化方法之一,这样做可以降低计算的复杂性。而升维,即向高维空间做映射,一般只会增加计算的复杂性,甚至会引起“维数灾”,因而人们很少问津。但是作为分类、回归等问题来说,很可能在低维样本空间无法线性处理的样本集,在高维特征空间却可以通过一个线性超平面实现线性划分(或回归),而与特征空间的线性划分(或回归)相对应的却是样本空间的非线性分类(或回归)。自然发生的两个问题是如何求得非线性映射  $\varphi$  和解决算法的复杂性。

SVM 方法巧妙地解决了这两个难题:由于应用了核函数的展开定理,所以根本不需要知道非线性映射的显式表达式;由于是在高维特征空间中应用线性学习机的方法,所以与线性模型相比几乎不增加计算的复杂性,这在某种程度上避免了“维数灾”。

## 1 分类和回归问题的提法

给定训练样本集

$$(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l) \quad (1)$$

其中  $x_i \in \mathbf{R}^N$ , 为  $N$  维向量,  $y_i \in \{-1, 1\}$  或  $y_i \in \{1, 2, \dots, k\}$  或  $y_i \in \mathbf{R}$ ; 通过训练学习寻求模式  $M(x)$ , 使其不但对于训练样本集满足  $y_i = M(x_i)$ , 而且对于预报数据集:

$$x_{l+1}, x_{l+2}, \dots, x_m \quad (2)$$

同样能得到满意的对应输出值  $y_i$ 。

为与国际文献的术语一致,以下称建立的分类模式  $M(x)$  为学习机或分类机。

当  $y_i \in \{-1, 1\}$  时为最简单的二类划分, 当  $y_i \in \{1, 2, \dots, k\}$  时为  $k$  类划分, 当  $y_i \in \mathbf{R}$  时为函数估计, 即回归分析。

限于篇幅本文主要介绍 SVM 分类方法, 对于 SVM 回归方法只在第 5 节作一简介。

多类划分可看成是做多个二类划分(判断是否属于某一类)。当分类机  $M(x)$  为线性函数(直线或线性超平面)时对应线性划分, 否则为非线性分类。

线性划分的理想情况是训练样本集可以完全线性分离。当训练样本集不能线性分离(训练样本有重叠现象)时, 可以通过引入松弛变量而转化为可线性分离的情况。

本文首先讨论线性可分离的二类划分的SVM分类机,进而解决线性不可分离的二类划分和多类线性分类问题,最后通过引入非线性映射和核函数,将非线性分类问题转化为线性SVM分类问题加以解决。图1给出SVM分类问题及解决思路的框图。

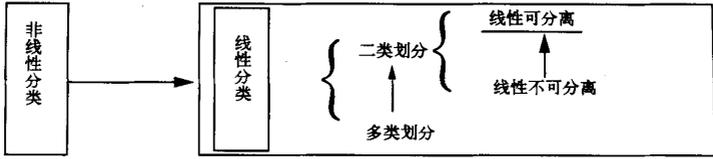


图1 SVM分类问题及解决思路

## 2 最优划分线性超平面和支持向量

对于式(1)给出的训练样本集的线性二类划分问题,就是寻求函数:

$$y = f(x) = \text{Sgn}((w \cdot x) + b) \tag{3}$$

使对于  $i = 1, 2, \dots, l$  满足条件:

$$y_i = f(x_i) = \text{Sgn}((w \cdot x_i) + b) \tag{4}$$

其中  $w, x, x_i \in \mathbf{R}^N, b \in \mathbf{R}, w, b$  为待确定的参数,  $\text{Sgn}$  为符号函数。显然  $(w \cdot x) + b = 0$  为划分超平面,  $w$  为其法方向向量。条件(4)又可写成等价形式:

$$y_i((w \cdot x_i) + b) > 0 \tag{5}$$

对于线性可分离的问题,满足上述条件形如式(3)的线性决策函数是不唯一的。图2给出二维情况下满足条件的划分直线的分布区域图。落在深色区域内的任一直线都可作为决策函数。于是人们要问:众多决策函数中有否最优?若有,是否唯一?如何寻求?

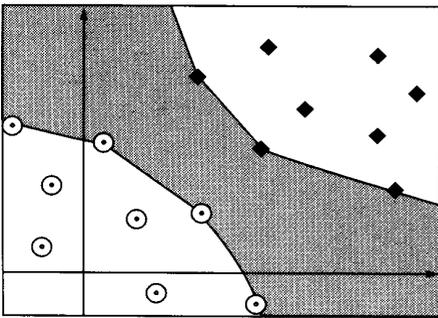


图2 划分直线的分布区域图

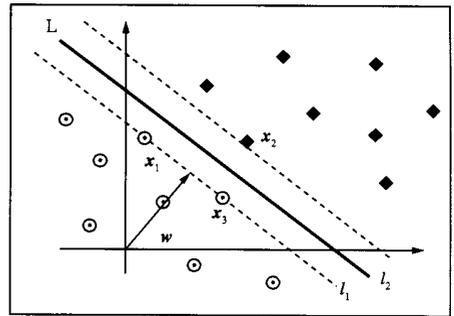


图3 最优划分超平面示意图

判断优劣要有原则,通常采用误差最小化原则,即寻求决策函数使对训练样本集的分类误差“总和”(有多种汇总方法)最小。按此原则,落在虚线区域内的任一直线都是最优,因为都使总分类误差为零。

Vapnik 提出一个最大边际化(maximal margin)原则<sup>[2]</sup>,所谓边际又称间隔,是指训练

样本集到划分超平面的距离,它是所有训练样本点到划分超平面的(垂直)距离中的最小者:

$$\text{Min}(\|x - x_i\| : x \in \mathbf{R}^N, (w \cdot x) + b = 0, i = 1, \dots, l)$$

所谓最大边际化原则是指寻求使间隔达到最大的划分为最优,即是对  $w, b$  寻优,求得最大间隔:

$$\text{Max}_{w, b}(\text{Min} \|x - x_i\| : x \in \mathbf{R}^N, (w \cdot x) + b = 0, i = 1, \dots, l)$$

对应最大间隔的划分超平面称为最优划分超平面,简称为最优超平面,如图 3 中的  $L$ 。图 3 中两条平行虚线  $l_1, l_2$ (称为边界)距离之半就是最大间隔。可以证明最大间隔是唯一的,但达到最大间隔的最优超平面可能不唯一。

最大间隔和最优超平面只由落在边界上的样本点完全确定,我们称这样的样本点为支持向量,如图 3 中的  $x_1, x_2, x_3$  样本点。

只由少数训练样本点(支持向量)就把最大间隔和最优超平面完全确定,其余非支持向量的样本点均不起作用,这具有重要的意义。它说明间隔最大化原则下的最优划分不是依赖于所有点,而只是由支持向量决定。求最优超平面和最大间隔等同于确定各个样本点是否支持向量。这预示着该方法具有好的鲁棒性(robustness)和好的算法复杂性。

由支持向量确定的线性分类机称为线性支持向量机。

### 3 线性支持向量机

#### 3.1 线性可分离的情况

对于式(1)给定的训练样本集,如果样本是线性可分离的,如何建立线性支持向量机?设图 3 中的划分超平面  $L$  的方程为:

$$(w \cdot x) + b = 0$$

两条边界  $l_1, l_2$  的方程(经过恒等变形后)为:

$$(w \cdot x) + b = \pm 1$$

设  $x_1$  在  $l_1$  上,  $x_2$  在  $l_2$  上,即:

$$(w \cdot x_1) + b = -1, \quad (w \cdot x_2) + b = +1$$

两式相减有:

$$(w \cdot (x_2 - x_1)) = 2$$

进而有:

$$\left(\frac{w}{\|w\|} \cdot (x_2 - x_1)\right) = \frac{2}{\|w\|} \quad (6)$$

式(6)左边恰好就是连接  $x_1, x_2$  的向量在划分超平面法方向上的投影,它是最大间隔的二倍。求最大间隔等价于求  $\|w\|$  或  $\|w\|^2$  或  $\frac{1}{2}\|w\|^2$  的最小值。

考虑到要使所有训练样本点分类正确,应成立:

$$\begin{aligned} (w \cdot x_i) + b &\geq 1, \text{ 若 } y_i = 1 \\ (w \cdot x_i) + b &\leq -1, \text{ 若 } y_i = -1 \end{aligned}$$

两式可以合并为:

$$y_i((w \cdot x_i) + b) \geq 1$$

这样,建立线性支持向量机的问题转化为求解如下一个二次凸规划问题:

$$\begin{cases} \min \frac{1}{2} \|w\|^2 \\ \text{约束条件: } y_i((w \cdot x_i) + b) \geq 1 \end{cases} \quad (7)$$

由于目标函数和约束条件都是凸的,根据最优化理论,这一问题存在唯一全局最小解。应用 Lagrange 乘子法并考虑满足 KKT 条件(Karush-Kuhn-Tucker):

$$a_i(y_i((x \cdot x_i) + b) - 1) = 0 \quad (8)$$

可求得最优超平面决策函数为:

$$M(x) = \text{Sgn}((w^* \cdot x) + b^*) = \text{Sgn}(\sum_{s.v.} a_i^* y_i(x \cdot x_i) + b^*) \quad (9)$$

其中  $a_i^*$ ,  $b^*$  为确定最优划分超平面的参数,  $(x \cdot x_i)$  为两个向量的点积。

由式(8)知:非支持向量对应的  $a_i$  都为零,求和只对少数支持向量进行。

### 3.2 线性不可分的情况

对于线性不可分的情况,通过引入松弛变量  $\xi \geq 0$ ,修改目标函数和约束条件,应用完全类似的方法可以求解。与式(7)类似的新的凸规划问题为:

$$\begin{cases} \min \frac{1}{2} \|w\|^2 + C \sum_i \xi \\ \text{约束条件: } y_i((w \cdot x_i) + b) \geq 1 - \xi \end{cases} \quad (10)$$

若  $\xi$  都为零,式(10)就变成了线性可分问题式(7)。式(10)中大于零的  $\xi$  对应错分的样本,参数  $C$  为惩罚系数。

### 3.3 线性多类分类问题

$k$  类分类问题可以转化为  $k$  个二类划分问题。这时对应的每一个二类划分的决策函数为:

$$f_i(x) = \text{Sgn}((w_i \cdot x) + b_i) \quad i = 1, 2, \dots, k$$

其中  $f_i(x)$  的定义为:

$$f_i(x) = \begin{cases} 1 & x \text{ 属于第 } i \text{ 类} \\ -1 & x \text{ 属于其他类} \end{cases}$$

$k$  类分类问题的总决策函数为:

$$M(x) = \arg(\max_i ((w_i \cdot x) + b_i)) \quad (11)$$

上式的  $\arg$  为选取指标函数,其含义为:选取使样本点  $x$  对于  $k$  个决策函数值最大的  $f_i(x)$  的指标  $i$  做为  $M(x)$  的值。 $i$  所对应的类,作为样本点应该归属的类。

## 4 非线性支持向量机

SVM方法真正有价值的应用是用来解决非线性问题,方法是通过一个非线性映射  $\varphi$ ,把样本空间映射到一个高维乃至无穷维的特征空间,使在特征空间中可以应用线性支

持向量机的方法解决样本空间中的高度非线性分类和回归等问题。图 4 对此给出二维样本数据的直观示意图。

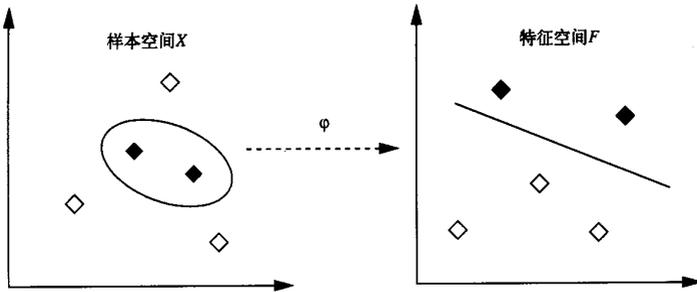


图 4 样本空间到特征空间的非线性映射示意

### 4.1 Mercer 核和 Mercer 定理

一个二元函数  $K(x, y)$  通常称为是一个核函数 (简称核)。给定核  $K(x, y)$ , 若有实数  $\lambda$  和非零函数  $\phi(x)$  使成立

$$\int_a^b K(x, y) \phi(x) dx = \lambda \phi(y)$$

则称  $\lambda$  为核的一个特征值, 称  $\phi(x)$  为核的关于特征值  $\lambda$  的一个特征函数。

对称正定的连续核称为 Mercer 核。关于 Mercer 核有如下定理:

**Mercer 定理<sup>[5]</sup>:** Mercer 核  $K(x, y)$  可以展开成一致收敛的函数项级数:

$$K(x, y) = \sum_i \lambda_i \phi_i(x) \phi_i(y) \tag{12}$$

其中  $\lambda_i, \phi_i(x)$  分别为核  $K(x, y)$  的特征值和特征向量, 它们的个数可能有限或无穷。

Mercer 核很多, 如径向基函数核、双曲正切函数核等。由已知的 Mercer 核经过某些运算可以生成新的 Mercer 核。特别是, 由点积定义的核必是 Mercer 核:

$$K(x, y) = (\varphi(x) \cdot \varphi(y)) \tag{13}$$

### 4.2 特征映射和特征空间

称由 Mercer 核的特征函数张成的函数集为特征空间, 记为  $F$ 。原样本空间记为  $X$ 。

如果我们作如下样本空间  $X$  到特征空间  $F$  的非线性映射  $\phi$ :

$$\varphi(x) = (\sqrt{\lambda_1} \phi_1(x), \sqrt{\lambda_2} \phi_2(x), \dots, \sqrt{\lambda_k} \phi_k(x), \dots)$$

则显然有:

$$K(x, y) = \sum_i \lambda_i \phi_i(x) \phi_i(y) = (\varphi(x) \cdot \varphi(y)) \tag{14}$$

从此可以看出: 当我们把样本空间通过非线性映射映入特征空间时, 如果只用到映象的点积, 则可以用相对应的核函数来代替, 而不需要知道映射的显式表达式。这是从线性支持向量机到非线性支持向量机的关键一步。

### 4.3 非线性支持向量机

在特征空间  $F$  中应用线性支持向量机的方法, 分类决策函数式(9)变为:

$$M(x) = \text{Sgn}((w^* \cdot \varphi(x)) + b^*) = \text{Sgn}(\sum_{s.v.} a_i^* y_i (\varphi(x) \cdot \varphi(x_i)) + b^*) \tag{15}$$

与式(9)相比,这里只是用  $\varphi(x)$  和  $\varphi(x_i)$  代替了  $x$  和  $x_i$ ,因此计算过程相同。

考虑到 Mercer 定理和式(14),式(15)可以化简为:

$$M(x) = \text{Sgn} \left( \sum_{S.V.} a_i^* y_i K(x, x_i) + b^* \right) \quad (16)$$

这就是非线性支持向量学习机的最终分类决策函数。虽然用到了特征空间及非线性映射,但实际计算中并不需要知道他们的显式表达。只要求出支持向量及其支持的“强度”和阈值,通过核函数的计算,即可得到原来样本空间的非线性划分输出值。

这样我们就通过核函数和线性 SVM 方法解决了非线性 SVM 问题。而线性 SVM 的算法归结为一个凸约束条件下的二次凸规划问题,对此已有很多成熟的算法和应用软件可资使用<sup>[6]</sup>。

## 5 SVM 回归方法简介

回归分析又称函数估计,它要解决的问题是:根据给定的样本数据集  $\{(x_i, y_i) | i=1, \dots, k\}$ ,其中  $x_i$  为预报因子值,  $y_i$  为预报对象值,寻求一个反映样本数据的最优函数关系  $y = f(x)$ 。

这里的最优是指按某一规定的误差函数计算,所得函数关系对样本数据集拟合的“最好”(累计误差最小)。图 5 中的 a, b, c 为多元统计分析中常用的误差函数, d 为 SVM 回归中常用的  $\varepsilon$ -不灵敏误差函数 ( $\varepsilon \geq 0$ )。当  $\varepsilon = 0$  时, d 等同于 b。

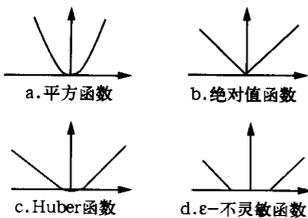


图 5 误差函数

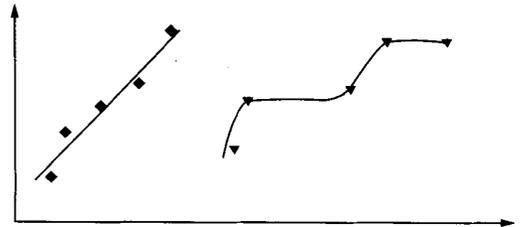


图 6 线性与非线性回归图示

如果所得函数关系  $y = f(x)$  是线性函数,则称线性回归,否则为非线性回归。图 6 给出二维数据的线性与非线性回归的图示。

与 SVM 分类问题不同的是:SVM 回归的样本点只有一类,所寻求的最优超平面不是使两类样本点分得“最开”,而是使所有样本点离超平面的“总偏差”最小。这时样本点都在两条边界线之间,求最优回归超平面同样等价于求最大间隔,推导过程与 SVM 分类情况相同,这里略去。

如果采用图 5d 的  $\varepsilon$ -不灵敏函数作为误差函数,当所有样本点到所求超平面的距离都可以不大于  $\varepsilon$  时(这相当于 SVM 分类时的线性可分的情况),寻求最优回归超平面的问题转化为求解如下一个二次凸规划问题:

$$\min \frac{1}{2} \|w\|^2$$

$$\text{约束条件: } \begin{cases} y_i - (w \cdot x_i) - b \leq \varepsilon \\ (w \cdot x_i) + b - y_i \leq \varepsilon \end{cases} \quad (17)$$

当个别样本点到所求超平面的距离大于  $\varepsilon$  时(这相当于 SVM 分类时的线性不可分的情况),  $\varepsilon$  不灵敏函数使超出的偏差相当于 SVM 分类中引入的松弛变量  $\xi$ , 引入容错惩罚系数  $C$ , 寻求最优回归超平面的二次凸规划问题变成:

$$\min \frac{1}{2} \|w\|^2 + C \sum_i (\xi_i + \xi_i^*),$$

$$\text{约束条件: } \begin{cases} y_i - (w \cdot x_i) - b \leq \varepsilon + \xi_i \\ (w \cdot x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (18)$$

对于最优化问题(17) (18), 类似于 SVM 分类方法, 可求得最优超平面线性回归函数为:

$$f(x) = (w \cdot x) + b = \sum_{s.v.} (a_i - a_i^*) (x \cdot x_i) + b \quad (19)$$

$a_i$ 、 $a_i^*$  和  $b$  通过约束条件求得, 为确定最优超平面的参数。上述求解过程有多种高效算法和成熟的计算机程序可资利用。最后的结果表明: 最优回归超平面只由作为支持向量的样本点完全确定。

式(19)中出现的点积提示我们可以同样引入核函数从而实现非线性回归。将样本空间中的点  $x$  和  $x_i$  用映射的像  $\phi(x)$  和  $\phi(x_i)$  代替, 再应用  $K(x, x_i) = (\phi(x) \cdot \phi(x_i))$  我们得到:

$$f(x) = (w \cdot \phi(x)) + b = \sum_{i=1}^L (a_i - a_i^*) K(x, x_i) + b \quad (20)$$

这就是 SVM 方法最终确定的非线性回归函数。

## 6 SVM 方法的特点及应用展望

SVM 是一种有坚实理论基础的新颖的小样本学习方法。它基本上不涉及概率测度及大数定律等, 因此不同于现有的统计方法。从本质上看, 它避开了从归纳到演绎的传统过程, 实现了高效的从训练样本到预报样本的“转导推理”(transductive inference), 大大简化了通常的分类和回归等问题。

从式(9)和式(16)可以看出, SVM 的最终决策函数只由少数的支持向量所确定, 计算的复杂性取决于支持向量的数目, 而不是样本空间的维数, 这在某种意义上避免了“维数灾”。如果说神经网络方法是对样本的所有因子加权的话, SVM 方法是对只占样本集少数的支持向量样本“加权”。当预报因子与预报对象间蕴涵的复杂非线性关系尚不清楚时, 基于关键样本的方法可能优于基于因子的“加权”。

少数支持向量决定了最终结果, 这不但可以帮助我们抓住关键样本“剔除”大量冗余

样本,而且注定了该方法不但算法简单,而且具有较好的“鲁棒”性。这种“鲁棒”性主要体现在:

- ① 增、删非支持向量样本对模型没有影响;
- ② 支持向量样本集具有一定的鲁棒性;
- ③ 有些成功的应用中,SVM方法对核的选取不敏感。

由于有较为严格的统计学习理论做保证,应用SVM方法建立的模型具有较好的推广能力。SVM方法可以给出所建模型的推广能力的严格的界,这是目前其它任何学习方法所不具备的。建立任何一个数据模型,人为的干预越少越客观。与其他方法相比,建立SVM模型所需要的先验干预较少。但核函数的选定及有关参数的优化仍是目前尚未解决的问题。

近年来SVM方法已经在图像识别、信号处理和基因图谱识别等方面得到了成功的应用,显示了它的优势<sup>[7]</sup>。SVM通过核函数实现到高维空间的非线性映射,所以适合于解决本质上非线性的分类、回归和密度函数估计等问题。支持向量方法也为样本分析、因子筛选、信息压缩、知识挖掘和数据修复等提供了新工具。气象预报预测领域的很多问题都带有显著的非线性特性,这一方法有望在释用数值预报产品和多项气象预报预测研究、业务中得到推广应用。虽然支持向量机方法与基于概率测度和大数定律的传统统计方法风格迥异,但它与通常的统计分析预测方法有天然的联系(有大致相同的问题表述和数据预处理等)。我们只作了SVM在分类预报和回归预报的应用个例(见本文的下篇),收到了较好的效果。对于应用了统计分析预测方法的其他研究和业务课题(诸如判别分析、主分量分析、时间序列分析等),不妨用SVM方法一试并加以比较,也许会有新的结果。

## 参 考 文 献

- 1 Vapnik V N. *Statistical Learning Theory*. John Wiley & Sons, Inc., New York, 1998.
- 2 Vapnik V N. *The Nature of Statistical Learning Theory*. Springer Verlag, New York, 2000.(有中译本:张学工译.统计学习理论的本质.北京:清华大学出版社,2000.)
- 3 Cristianini N and Shawe-Taylor J. *An Introduction of Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.
- 4 Burges C J. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 1998, 2: 127~167.
- 5 Courant R and Hilbert D, *Method of Mathematical Physics, Volume I*. Springer Verlag, 1953.
- 6 <http://www.kernel-machines.org/>
- 7 Scholkopf B, Burges Ch J C and Smola A J, edited. *Advances in Kernel Methods-Support Vector Learning*. MIT Press, Cambridge, 1999.

# A NEW METHOD FOR NON-LINEAR CLASSIFY AND NON-LINEAR REGRESSION I : INTRODUCTION TO SUPPORT VECTOR MACHINE

Chen Yongyi Yu Xiaoding Gao Xuehao

(*Training Center of China Meteorological Administration, Beijing 100081*)

Feng Hanzhong

(*Chengdu Meteorological Center, Chengdu 610071*)

## Abstract

A brief introduction to an increasingly popular machine learning technique, SVM (support vector machine) is presented for solving nonlinear classification and regression problems. Properties of SVM are discussed together with potentials of applying SVM to numerical weather forecast. SVM is a novel learning method that has solid theoretical basis and requires only small amount of sample. It does not rely on probability measures and Law of Large Numbers, hence is different from many other statistical methods. In essence, SVM smartly evades the traditional inference process from induction to deduction. Instead, it employs transductive inference from training sample to predicting sample, which greatly simplifies classification and regression problems. The decision function of SVM is only determined by a few support vectors. The complexity of computation relates to the number of support vectors rather than the dimension of the sample space. Thus, to some degree, SVM avoids the "curse of dimensionality".

**Key words:** Machine learning Support vector machines (SVM) Pattern recognition Regression estimation Weather forecast