

支持向量机理论及其应用

卢虎¹, 李彦², 肖颖²

(1. 西北工业大学, 陕西 西安 710072; 2. 空军工程大学 电讯工程学院, 陕西 西安 710077)

摘要: 作为当前国际机器学习前沿热点的支持向量机是一种新型的机器学习算法, 具有卓越的学习效果。文中分析了该方法的核心思想及常用训练算法, 并给出其具体应用。

关键词: 支持向量机; 统计学习理论; 机器学习

中图分类号: TP183 **文献标识码:** A **文章编号:** 1009-3516(2003)04-0089-03

支持向量机(Support Vector Machines, 简称 SVM)是一种基于统计学习理论的新型机器学习算法, 由于其具有出色的学习性能, 已成功应用于人脸识别、手写体数字识别、图像检索等许多领域。本文在介绍最优超平面概念的基础上, 分析 SVM 的特点, 给出常用的 SVM 训练算法及其具体应用。

1 最优超平面

设给定训练数据为 $\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$, 其中 $x \in \mathbf{R}^n$, $y \in \{-1, +1\}$, 又若 n 维空间中的线性判别函数一般形式为: $g(x) = (w \cdot x) + b$, 且集合中的所有数据都可以被分类面 $(w \cdot x) + b = 0$ 所正确划分, 则该分类面就是最优超平面。而距离该最优超平面最近的异类向量就是所谓的支持向量(Support Vector), 支持向量与超平面之间的距离最大(即边缘最大化), 一组支持向量可以唯一的确定一个超平面。如图 1 所示。

图中, 实心点和空心点代表两类样本, H 为分类线, H_1, H_2 分别为过各类中离分类线最近的样本且平行于分类线的直线, 它们之间的距离叫做分类间隔(Margin)。由于支持向量与超平面之间的距离为 $1/\|w\|$, 则支持向量间距为 $2/\|w\|$, 寻找超平面的问题, 可转化为求解以下二次规划问题:

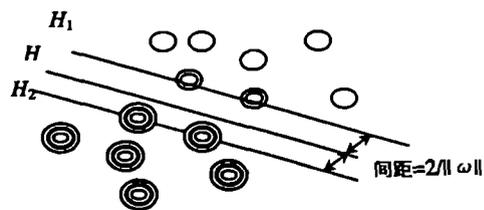


图1 SVM 算法原理示意图

$$\Phi(w) = \frac{1}{2}(w \cdot w) \quad (1)$$

约束条件为不等式: $y_i[(w \cdot x_i) + b] \geq 1 \quad i = 1, 2, \dots, l$ (2)

对一个规范超平面子集, 其 VC 维 h 满足不等式:

$$h \leq \min([R^2 A^2 + b], n) + 1 \quad (3)$$

式中 n 为向量空间的维数, R 为覆盖所有向量的超半球半径, $\|w\| \leq A$ 。

由式(3)可以得到, 通过最小化 $\|w\|$ 可使 VC 置信度最小, 若固定经验风险, 则最小化期望风险的问题就是最小化 $\|w\|$ 的问题, 此即 SVM 算法的理论出发点。

2 构造最优超平面

可知, 在线性条件下式(1)的最优解为下面 Lagrange 函数的鞍点:

收稿日期: 2002-12-31

作者简介: 卢虎(1975-), 男, 陕西西安人, 硕士生, 主要从事信号与信息处理研究。

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i \{y_i(w \cdot x + b) - 1\} \tag{4}$$

式中, α 为非负 Lagrange 乘子。在鞍点处, 由于 w 和 b 的梯度均为零, 则可知最优超平面系数 α_i 满足: $\sum_{i=1}^l \alpha_i y_i = 0$, 此时最优超平面是训练集中的向量的线性组合:

$$w = \sum_{i=1}^l \alpha_i y_i x_i \tag{5}$$

因为只有支持向量可以在 w 的展开式中具有非零系数 α_i , 这时的支持向量就是使式(2)成立的向量, 即只有支持向量影响最终的分类结果, 数学表述如下:

$$w = \sum_{\text{支持向量}} \alpha_i y_i x_i \tag{6}$$

根据 Kuhn - Tucker 条件可知, 最优解满足:

$$\alpha_i \{[(w \cdot x) + b] y_i - 1\} = 0 \tag{7}$$

将式(6)、(7)代入式(5)中, 构造最优超平面的问题就转化为一个较为简单的二次规划问题, 可知分类平面为

$$f(x) = \text{sgn}(\sum \alpha_i y_i x_i \cdot x + b) \tag{8}$$

对于线性不可分的情况, SVM 引入松弛变量和惩罚因子, 使目标函数变为

$$\Phi(w, \zeta) = \frac{1}{2} (w \cdot w) + C (\sum_i \zeta_i) \tag{9}$$

此外, SVM 通过非线性变换将输入空间变换到高维空间, 然后在新空间中求解最优分类面, 线性可分情况下的点积运算变为 $k(x, y) = (\Phi(x) \cdot \Phi(y))$ 。由此得到的分类函数为

$$f(x) = \text{sgn}(\sum \alpha_i y_i \cdot k(x_i, x) + b) \tag{10}$$

3 支持向量机

支持向量机实现的是如下的思想: 通过事先选择好的非线性映射将输入向量 x 映射到一个高维特征空间 Z , 在这个空间构造最优分类超平面。SVM 的分类函数在形式上类似神经网络, 输出是中间节点的线性组合, 每个中间节点对应一个支持向量, 如图 2 所示。

由以上对最优超平面的讨论可知, 向量之间只进行点积运算。因此, 如果采用核函数(Kernal Function), 就可以避免在高维特征空间进行复杂的运算。该过程可以表述如下: 首先将输入向量 x 通过映射: $\Psi: \mathbf{R}^n \rightarrow H$ 映射到高维 Hilbert 空间 H 中。该核函数 K 满足: $k(x_i, y_j) = (\Phi(x) \cdot \Phi(y))$, 显然不同的核函数将形成不同的算法(即不同的支持向量机), 目前, 主要有: ①多项式核函数: $k(x, x_j) = [(x \cdot x_j) + 1]^d$, 得到 d 阶多项式分类器; ②径向基函数(RBF): $k(x, y) = \exp[-\frac{\|x-y\|^2}{\sigma^2}]$, 所得分类器与传统 RBF 的区别是: 这里每个基函数中心对应一个支持向量, 它们及输出权值都是由算法自动确定的。另外, 还有其它一些核函数。

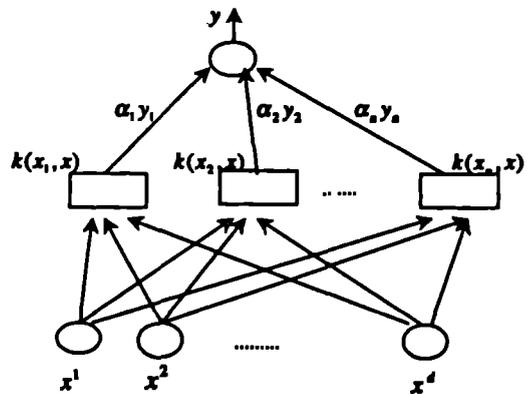


图2 支持向量机示意图

4 应用研究

SVM 是一种未成熟的新技术, 仍然存在很多局限。但是由于 SVM 找到的是全局最优解。因此在很多问题上都有着其它统计学习技术难以比拟的优越性, 并在一些领域已经获得成功。

模式识别中,最突出的应用研究是字符识别。表1是贝尔实验室对美国邮政手写字体识别的实验结果。

表1 各种分类器结果比较

分类器	错误率
人工表现	2.5%
决策树 C4.5	16.2%
5层神经网络	5.1%
SVMRBF 分类器	4.1%
SVM 多项式分类器	4.0%

表1一方面说明SVM方法较传统方法有明显优势,同时也看到不同SVM方法得到相近的结果,说明SVM对不同方法具有一定的不敏感性,(目前该点只能从试验说明,尚未找到数学证明)而不象神经网络那样对模型的依赖型很强。

在函数拟合方面,SVM主要应用于函数逼近、时间序列预测和数据压缩等。

图3是原始数据的曲线,图4是采用RBF函数为核函数进行拟合的结果,可以看到曲线拟合效果良好。(图4中的小圆代表对原始数据进行插值的点。)

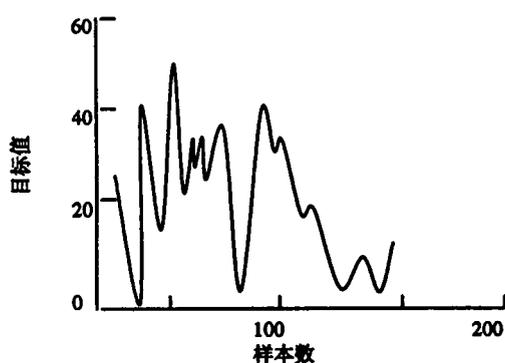


图3 原始数据

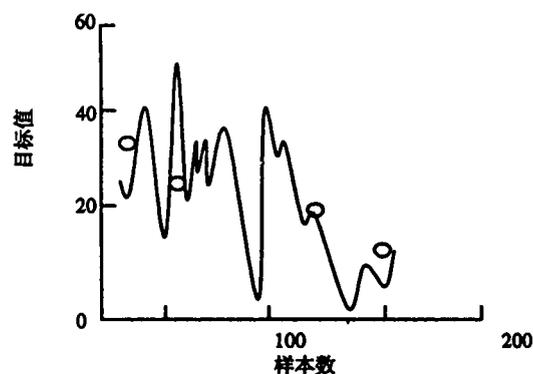


图4 SVM拟合结果

参考资料:

- [1] Vapnik V. The Nature of Statistical Learning Theory[M]. New York: Springer Verlag, 1995.
- [2] Cortes C, Vapnik V. Support Vector Networks[J]. Machine Learning, 1995, 20: 273 - 297.
- [3] 马云潜, 张学工. 支持向量机函数拟合在分形插值中的应用[J]. 清华大学学报, 2000, 40(3): 76 - 78.
- [4] 张 磊, 林福宗. 基于支持向量机的相关反馈图像检索算法[J]. 清华大学学报, 2002, 42(1): 80 - 83.

(编辑: 门向生)

Support Vector Machines and its Application

LU Hu¹, LI Yan², XIAO Ying²

(1. Dept. Electronic Engineering, Northwestern Polytechnic University, Xi'an, Shaanxi 710072, China; 2. The Telecommunication Engineering Institute, Air Force Engineering University, Xi'an, Shaanxi 710077, China)

Abstract: Support Vector Machines are a kind of novel machine learning methods, which has become a hotpot of machine learning because of their excellent learning effect. In this paper SVM's core thinking and popular training algorithms are presented and the concrete applications are given.

Key words: support vector machines; statistical learning; machine learning