用于回归估计的支持向量机
The Support Vector Machines for Regression

李志明,孔令富
Li Zhiming,Kong Lingfu

(燕山大学信息科学与工程学院,河北秦皇岛 066004)
(Coll. of Info. Sci. and Engi. ,Yanshan Univ. ,Qinhuangdao,Hebei,066004,China)

摘要:介绍机器学习的表示方式,分析和比较机器学习中经验风险最小化原则和结构风险最小化原则,引出用于回归估计的支持向量机,并用数学方式阐述其基本思想,讨论支持向量机技术发展中存在的主要问题.

关键词:支持向量机 回归估计 经验风险最小化 结构风险最小化
中图法分类号:TP181 **文献标识码**:A **文章编号**:1002-7378(2005)04-0215-04

Abstract:The expression of machine learning is introduced. The empirical risk minimization and the structural risk minimization in machine learning are analyzed. A support vector machine for regression is presented. The basic idea an dthe main issues in the development of the support vector machine are discussed.

Key words: support vector machine, regression, empirical risk minimization, structural risk minimization

对采集数据的学习归纳出某种系统规律,并利用这些规律对未来数据或无法观测到的数据进行预测是机器学习系统研究的重点.在这类研究方法中,神经网络方法应用最广泛,但神经网络基于经验风险最小化(Empirical Risk Minimization,简称ERM)原则泛化能力较差,存在过学习和局部最优解等到目前还无法克服的问题[1,2].支持向量机(Support Vector Machine,简称SVM)是由Vapnik等[1]提出的一种基于小样本统计理论的学习机,具有完备的理论基础和严格的理论体系.支持向量机是结构化风险最小化原理的近似实现,能够提高学习机的泛化能力.此外,支持向量机算法最终转化为二次规划的凸优化问题,存在全局唯一最优解.基于上述优点,SVM一经提出就得到了广泛的重视.随着Vapnik[3]对不敏感损失函数的引入,SVM已推广到非线性系统的回归估计,并展现了良好的学习和泛化性能.本文描述用于回归估计的支持向量机的基本方法,并讨论目前困扰支持向量机发展的一些问题.

收稿日期:2005-07-05
作者简介:李志明(1978-),男,河北滦县人,硕士研究生,主要从事计算机测控技术和软测量技术研究。

1 机器学习的表示

机器学习的最终目的就是根据给定的训练样本求取系统输入与输出之间的某种依赖关系,并对未知输出作出尽可能准确的预测.机器学习问题可以形式化地表示为:已知变量与输入变量之间存在某种未知依赖关系,即存在一个未知的联合概率$F(x,y)$,机器学习根据n个独立同分布观测样本

$$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n), \tag{1}$$

在一组函数$\{f(x,w)\}$中寻求一个最优的函数$\{f(x,w_0)\}$,使预测期望风险

$$R(w)=\int L(y,f(x,w))\mathrm{d}F(x,y) \tag{2}$$

最小,其中,$\{f(x,w)\}$称作预测函数集,$w\in\Omega$为函数的广义参数,则$\{f(x,w)\}$可以表示任何函数集,通常也称作学习函数、学习模型或学习机器;$L(y,f(x,w))$是由于用$f(x,w)$对y进行预测而造成的损失,称之为损失函数.不同类型的学习问题有不同形式的损失函数.在回归估计问题中,损失函数可以定义为

$$L(y,f(x,w))=(y-f(x,w))^2. \tag{3}$$

损失函数通常采用ε-不灵敏区函数,其定义为:

$$L(x,y)=$$

$$\begin{cases} 0, & |\hat{f}(x)-y|<\varepsilon, \\ |\hat{f}(x)-y|-\varepsilon, & |\hat{f}(x)-y|\geqslant\varepsilon,(\varepsilon>0), \end{cases} \tag{4}$$

其中，$\hat{f}(x)$ 为通过对样本集的学习而构造的回归估计函数；y 是 x 对应的实际目标值；ε 为与函数估计精度直接相关的设计参数，ε- 不敏感损失函数通常形象地被喻为 ε- 管道. 学习的目的是构造 $\hat{f}(x)$，使之与目标值之间的距离小于 ε，同时函数的复杂性最小，这样对于未知样本 x，可最优地估计出对应的目标值.

2 经验风险最小化与结构风险最小化

(2) 式定义的期望风险最小化必须依赖关于联合概率 $F(x,y)$ 的信息. 但是，在实际的机器学习问题中，我们能利用的只有样本(1) 的信息，因此期望风险并无法直接计算和最小化. 根据概率论中大数定理的思想，人们自然想到用算术平均代替(2) 式中的数学期望，于是定义

$$R_{\text{emp}}(w)=\frac{1}{n}\sum_{i=1}^{n}L(y_i,f(x_i,w)) \tag{5}$$

来逼近(2) 式定义的期望风险. 由于 $R_{\text{emp}}(w)$ 是用已知的训练样本(即经验数据) 定义的，因此称作经验风险[4]. 以参数 w 求经验风险 $R_{\text{emp}}(w)$ 的最小值代替求期望风险 $R(w)$ 的最小值，就是所谓的 ERM 原则. 经仔细思考可以发现，由于有限样本的限制使得经验风险最小化原则与经验风险之间没有必然的联系，即使能够保证样本无穷大的条件得到满足，也无法认定在此前提下得到的经验风险最小化方法在样本数有限时仍能具有好的预测效果.

统计学习理论中关于经验风险和实际风险之间关系的重要结论，称作推广性的界. 该结论说明实际风险 $R(w)$ 与经验风险 $R_{\text{emp}}(\omega)$ 之间的关系：

$$R(w)\leqslant R_{\text{emp}}(\omega)+\Phi(n/h), \tag{6}$$

其中，$\Phi(\cdot)$ 为单调递减函数；h 是函数集的 VC 维；n 是样本数. 式(6) 表明在有限训练样本下，学习机器的 VC 维越高，复杂性越高，则置信范围越大，导致真实风险与经验风险之间可能的差别越大. 要取得良好的学习效果，机器学习过程不但要使经验风险最小，还要使 VC 维尽可能小从而达到缩小置信范围的目的，最终取得较小的实际风险和较好的推广性. 这也正是大多数情况下选用较复杂的学习机器即使能够获得较好的记忆功能却得不到令人满意的推广性能的原因.

ERM 准则只强调经验风险最小(训练误差)，没有最小化置信范围值，因此基于 ERM 准则的学习方法的学习能力强，但泛化能力较差，导致出现过学习现象，例如神经网络. 最大化泛化能力不仅需要最小化经验风险，而且应最小化置信范围值. 基于此思想，统计学习理论提出一种新的策略，即把函数集构造为一个函数子集序列，使各个子集按照 VC 维的大小排列，在每个子集中寻找最小经验风险，在子集间折衷考虑经验风险和置信范围，取得实际经验风险最小. 这种思想称作结构风险最小化或有序风险最小化(Structural Risk Minimization，SRM) 准则[5]，如图 1 所示.

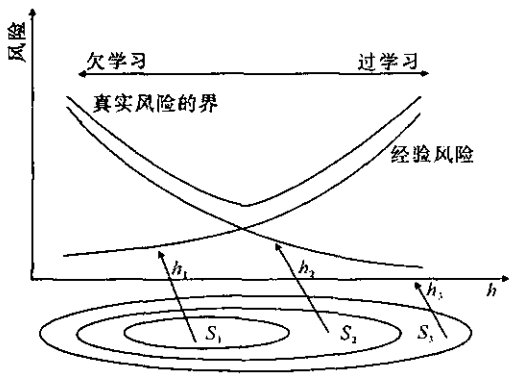

图 1 结构风险最小化

SVM 是结构风险最小化思想的具体实现，它不像神经网络等传统方法那样以训练误差最小化作为优化目标，而是以训练误差作为优化问题的约束条件，以置信范围值最小化作为优化目标.

3 用于回归的支持向量机

回归估计问题中，假设存在一未知函数 $y=f(x)$，$x\in R^d$，$y\in R$，要求函数 $\hat{f}:R^d\to R$，使得函数 f 和 $\hat{f}$ 之间的距离最小，即损失函数 $R(f,\hat{f})=\int L(f,\hat{f})\mathrm{d}x$ 最小，由于函数 f 未知，因而只能根据已知的有限样本来求取 $\hat{f}$.

3.1 线性回归

样本数据为线性时，假定 $\hat{f}(x)$ 为如下形式：

$$\hat{f}=\langle w,x\rangle+b, \tag{7}$$

其中，$\langle w,x\rangle$ 表示 $w\in R^d$ 与 $x\in R^d$ 的内积，$b\in R$. 根据结构风险最小化准则，$\hat{f}$ 应使得：

$$J=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{r}L(\hat{f}(x_i),y_i) \tag{8}$$

最小，其中，C 是平衡因子，$\|\cdot\|$ 表示向量模.

由 ε- 不敏感损失函数可知，用于回归估计的支持向量机可以表示为

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{r}(\xi_i + \xi_i^*), \tag{9}$$

$$\text{s.t.}\begin{cases} y_i - \langle w, x_i\rangle - b \leqslant \varepsilon + \xi_i, \\ \langle w, x_i\rangle + b - y_i \leqslant \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geqslant 0, \end{cases}$$

其中，ξ_i 为目标值之上超出 ε 部分所设的松驰变量，ξ_i^* 为目标值之下超出 ε 部分所设的松驰变量.

在样本数目不多的情况下，使用拉格朗日定理[3] 解凸最优化问题可以使用对偶表示替代原描述. 由于直接处理不等式约束较为困难，而对偶问题是通过引入称为对偶变量的拉格朗日乘子求解，所以对偶问题通常比原问题更易处理. 要解上述问题，首先把拉格朗日函数对于各个原变量的导数置零，然后将得到的关系式带入原拉格朗日函数，将原问题转化为对偶问题并去除原变量的相关性，具体方法如下.

建立拉格朗日方程：

$$L(w,\xi_i,\xi_i^*) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{r}(\xi_i + \xi_i^*) - \sum_{i=1}^{r}\alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i\rangle + b) - \sum_{i=1}^{r}\alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, x_i\rangle - b) - \sum_{i=1}^{r}(\eta_i\xi_i + \eta_i^*\xi_i^*), \tag{10}$$

其中，参数 w, b, ξ_i, ξ_i^* 的偏导都应等于零，即

$$\text{s.t.}\begin{cases} \dfrac{\partial L}{\partial w} = w - \sum_{i=1}^{r}(\alpha_i - \alpha_i^*)x_i = 0, \\ \dfrac{\partial L}{\partial b} = \sum_{i=1}^{r}(\alpha_i - \alpha_i^*) = 0, \\ \dfrac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0, \\ \dfrac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0, \end{cases} \tag{11}$$

根据最优化的充要条件(KKT 条件[3]) 知，在最优点，拉格朗日乘子与约束的乘积为 0，即

$$\alpha_i(\varepsilon + \xi_i - y_i + w \cdot x_i + b) = 0, \tag{12}$$

$$\alpha_i^*(\varepsilon + \xi_i^* - y_i + w \cdot x_i + b) = 0, \tag{13}$$

$$\eta_i\xi_i = 0 \rightarrow (C - \alpha_i)\xi_i = 0, \tag{14}$$

$$\eta_i^*\xi_i^* = 0 \rightarrow (C - \alpha_i^*)\xi_i^* = 0, \tag{15}$$

结合(12) ～ (15) 式，将(11) 式代入(10) 式，得到对偶优化问题

$$\max -\frac{1}{2}\sum_{i,j=1}^{r}(\alpha_i - \alpha_i^*)(\alpha_j - a_j^*)\langle x_i, x_j\rangle - \varepsilon\sum_{i=1}^{r}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{r}y_i(\alpha_i - \alpha_i^*), \tag{16}$$

$$\text{s.t.}\begin{cases} \sum_{i=1}^{r}(\alpha_i - \alpha_i^*) = 0, \\ \alpha_i, \alpha_i^* \in [0, C]. \end{cases}$$

由此，回归估计问题就归结为二次规划问题(16) 式. 求解该二次规划问题，可得

$$w = \sum_{i=1}^{r}(\alpha_i - \alpha_i^*)x_i, \tag{17}$$

根据(12) 与(13) 式通过反证法可以得出

$$\alpha_i \times \alpha_i^* = 0. \tag{18}$$

该式说明如果 α_i 不为 0，则 α_i^* 必为 0，反之亦然. 因此最优化计算得到的 α_i, α_i^* 的取值必然为以下形式之一：

$$\begin{cases} \alpha_i = 0, \alpha_i^* = 0, & (19a) \\ \alpha_i = 0, 0 < \alpha_i^* < C, & (19b) \\ 0 < \alpha_i < C, \alpha_i^* = 0, & (19c) \\ \alpha_i = 0, \alpha_i^* = C, & (19d) \\ \alpha_i = C, \alpha_i^* = 0. & (19e) \end{cases}$$

由(17) 式可知，非支持向量((19a) 所对应的 x_i) 对 w 没有贡献，只有支持向量((19b) ～ (19e) 所对应的 x_i) 对 w 有贡献，对应的学习方法称为支持向量机. (19b) 和(19c) 对应的 x_i 称为标准支持向量，它落在 ε- 管道上的数据点上. (19d) ～ (19e) 对应的 x_i 称为边界支持向量，是超出 ε- 管道的数据点. 因此，ε 越大，支持向量数越少，但函数估计精度越低[6].

对于标准支持向量，如果 $\alpha_i = 0, 0 < \alpha_i^* < C$，则由(13)、(15)、(17) 式可得

$$b = y_i - \sum_{x_j \in SV}(\alpha_j - \alpha_j^*)x_j \cdot x_i + \varepsilon. \tag{20}$$

如果 $0 < \alpha_i < C, \alpha_i^* = 0$，则由(12)、(14)、(17) 式可得

$$b = y_i - \sum_{x_j \in SV}(\alpha_j - \alpha_j^*)x_j \cdot x_i - \varepsilon, \tag{21}$$

为了计算准确可靠，通常对所有的标准支持向量分别计算 b 的值，再求其平均值

$$b = \frac{1}{N}\{\sum_{0<\alpha_i<C}[y_i - \sum_{x_j \in SV}(\alpha_j - \alpha_j^*)x_j \cdot x_i - \varepsilon] + \sum_{0<\alpha_i^*<C}[y_i - \sum_{x_j \in SV}(\alpha_j - \alpha_j^*)x_j \cdot x_i + \varepsilon]\}, \tag{22}$$

其中，N 为标准支持向量的个数.

由(7)、(17)、(22) 式最终可得计算估计函数为

$$\hat{f}(x) = \sum_{x_i \in SVs}(\alpha_i - \alpha_i^*)x_i \cdot x + b, \tag{23}$$

其中，SVs 表示支持向量机.

3.2 非线性回归

对于训练集为非线性情况，首先通过非线性变换 $x \rightarrow \varphi(x)$，将输入空间映射成高维的特征空间(Hilbert 空间)，然后在特征空间中进行线性逼近。这样(7) 式及目标函数(16) 式分别变为

$$f(x) = \langle w, \varphi(x) \rangle + b, \tag{24}$$

$$\max -\frac{1}{2}\sum_{i,j=1}^{r}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle \varphi(x_i), \varphi(x_j) \rangle - \varepsilon\sum_{i=1}^{r}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{r} y_i(\alpha_i - \alpha_i^*). \tag{25}$$

约束条件不变，从而得到

$$w = \sum_{i=1}^{r}(\alpha_i - \alpha_i^*)\varphi(x_i). \tag{26}$$

在支持向量机中，引入核函数 Kernel Function[3] 来简化非线性逼近. 核函数 $k(x, x')$ 满足

$$k(x, x') = \langle \varphi(x), \varphi(x') \rangle,$$

这样(24)式变为

$$\max -\frac{1}{2}\sum_{i,j=1}^{r}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) - \varepsilon\sum_{i=1}^{r}(\alpha_i^* + \alpha_i) + \sum_{i=1}^{r} y_i(\alpha_i^* - \alpha_i), \tag{27}$$

而(22)式变为

$$b = \frac{1}{N}\{\sum_{0<\alpha_i<C}[y_i - \sum_{x_j \in SV}(\alpha_j - \alpha_j^*)k(x_j \cdot x_i) - \varepsilon] + \sum_{0<\alpha_i^*<C}[y_i - \sum_{x_j \in SV}(\alpha_j - \alpha_j^*)k(x_j \cdot x_i) + \varepsilon]\}. \tag{28}$$

由(24)、(26)、(28)式可得计算回归估计函数为

$$\hat{f}(x) = \sum_{x_i \in SVs}(\alpha_i - \alpha_i^*)k(x_i \cdot x) + b. \tag{29}$$

由(27)式可知，尽管通过非线性函数将样本数据映射到具有高维甚至为无穷维的特征空间，但在计算回归估计函数时并不需要显式计算该非线性函数，而只需计算核函数，从而避免高维特征空间引起的维数灾难问题. 核函数 $k(x, x')$ 是对称正实数函数，且必须满足 Merce 条件.

4 讨论

由于统计学习理论和支持向量机建立了一套较好的有限样本机器学习的理论框架和通用方法，具有严格的理论基础，能够较好地解决小样本、非线性、高维数和局部最小点等问题，因此成为 20 世纪 90 年代末发展最快的研究方向之一. 本文深入推导了用于解决回归估计问题的SVM 方法，与其它方法相比，SVM 具有泛化性强、效率高等特点. 但由于 SVM 是一种新技术，其发展仅有十多年时间，还有一些问题需要深入研究，如受核矩阵存贮空间约束的支持向量机适应大样本的情况、核函数类型及其相应参数的选择、输入样本映射到高维空间后是否为线性可分的判定等问题.

参考文献：

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

[2] Cherkassky V, Mulier F. Learning Form Data: Concepts, Theory and Methods[M]. New York: John Viley&. Sons, 1997.

[3] Nello Cristianinni, John Shawe-Taylor[英]. 支持向量机导论[M]. 李国正，王 蒙，曾华军译. 北京：电子工业出版社，2005. 100-103, 70-80, 24-45.

[4] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报，2000, 26(1): 32-42.

[5] 边肇祺，张学工. 模式识别[M]. 第 2 版. 北京：清华大学出版社，2000. 284-303.

[6] 杜树新，吴铁军. 用于回归估计的支持向量机方法[J]. 系统仿真学报，2003, 15(11): 1580-1585.

（责任编辑：黎贞崇）

（上接第 211 页）

3 结束语

目前，我国的电子商务正在蓬勃发展，正是建立和发展电子商务物流体系的黄金时期，可靠的、高效的物流配送系统是电子商务应用研究中的重要组成部分。本文提出的一种基于遗传算法的物流配送车的优化调度算法，在研制物流配送软件过程中，通过将遗传算法应用于配送车的调度问题，可以实现快速、合理地安排运输路线和运输车次，能够取得较好的应用效果。

参考文献：

[1] 文 岗. 电子商务时期的第三方物流管理[M]. 北京：中国商业出版社，2000.

[2] 王海龙，王行愚. 一种基于配送体系物流信息平台的研究[J]. 计算机应用研究，2001, 18(7): 32-34.

[3] 蔡希贤，夏士智编译. 物流合理化的数量方法[M]. 武汉：华中工学院出版社，1985.

[4] 陈 龙. 基于遗产算法的约束性多 TSP 问题及其应用[J]. 重庆邮电学院学报，2000, 12(2): 67-69, 74.

[5] 李陶深. 人工智能[M]. 重庆：重庆大学出版社，2002.

[6] Rong Yang. Solving large travelling salesman problems with small populations [EB/OL]. http://citeseer.nj.nec.com/145347.html.

[7] Hong Tuangpei, Wang Hongshuang, Chen Weichou. Simultaneously applying multiple mutation operators in genetic algorithms[J]. Journal of Heuristics, 2000, 6: 439-455.

（责任编辑：邓大玉）