DOI: 10.13656/j.cnki.gxkx.2008.03.023
广西科学 Guangxi Sciences 2008, 15(3): 282～284

支持向量机回归与分类解的关系

The Relation of Solution Between Support Vector Machine Regression and Classification

周德强
ZHOU De-qiang

(长江大学信息与数学学院,湖北荆州 434023)
(School of Information and Mathematics, Yangtze University, Jingzhou, Hubei, 434023, China)

摘要: 将线性 $L_p(p=1)$ 和 L_ε 损失函数下的支持向量机回归与分类解的关系及支持向量机回归与 ν-支持向量机分类解的关系,推广到非线性 $L_p(p\geqslant 1)$ 和 L_ε 损失函数上,得到这些解关系更一般的形式.

关键词: 支持向量机 回归 分类 损失函数
中图法分类号: O177.91, TP18 文献标识码: A 文章编号: 1005-9164(2008)03-0282-03

Abstract The relation of solution between SVMR and SVMC, and the relation of solution between SVMR and ν-SVMC based on the linear L_p and L_ε loss function to the non-linear L_p and L_ε loss function are generalized, and the general form of original theory is obtained.

Key words support vector machine, regression, classification, loss function

支持向量机(SVM)由 Vapnik. V[1]在 1995年提出来,是近年来机器学习研究的一项重大成果. SVM 方法最初用于模式识别(分类)问题,近而成功地推广到函数回归与逼近问题.随之,支持向量机回归与分类的关系也得到广泛研究. Pontil等人在 L_p 损失函数 $p=1$的情况下研究支持向量机回归(SVMR)与分类(SVMC)解的关系[2].本文将线性 $L_p(p=1)$ 和 L_ε 损失函数下的支持向量机回归与分类解的关系及支持向量机回归与 ν-支持向量机分类解的关系,推广到非线性 $L_p(p\geqslant 1)$和 L_ε损失函数上,得到这些解关系更一般的形式.

1 预备知识

在支持向量机分类(SVMC)[1,3]问题中,给定 l 个样本 $(x_1,y_1),\cdots,(x_l,y_l)$,其中 $x_i\in R^n$, $y_i\in\{-1,+1\}$,目标是寻找一个超平面 (ω,b),使它能以最大间隔将给定的正样本和负样本分开.为了形式化表述这一目标,引入非负变量 $\xi_i\geqslant 0, i=1,2,\cdots,l$,根据这些变量, SVMC问题可描述为:

$$\min_{\omega,b,\xi}\frac{1}{2}\|\omega\|^2+CF\Big(\sum_{i=1}^{l}\xi_i^{\sigma}\Big), \qquad (1)$$

约束条件是

$y_i(\omega\cdot x_i+b)\geqslant 1-\xi_i;\ \xi_i\geqslant 0, i=1,2,\cdots,l$,

其中 $F(u)$是一个单调的凸范函 $(p\geqslant 1)$,称为损失函数, $C\geqslant 0$为惩罚因子[1,2].在本文所讨论的问题中,考虑 L_p 损失函数[2,3],则此时的 SVMC问题等价于问题(1)中取凸泛函 $F(\xi)=\sum_{i=1}^{l}\xi_i$,那么上述 SVMC问题变为:

$$\min_{\omega,b,\xi}\frac{1}{2}\|\omega\|^2+\frac{C}{p}\sum_{i=1}^{l}\xi_i^{p}, \qquad (2)$$

约束条件是

$y_i(\omega\cdot x_i+b)\geqslant 1-\xi_i;\ \xi_i\geqslant 0, i=1,2,\cdots,l$.

在支持向量机回归(SVMR)[1,3]问题中,目标是构造一个超平面使之尽可能接近大多数样本点.给定 l个样本 $(x_1,y_1),\cdots,(x_l,y_l)$,从这些有限样本中,需要找到回归函数的一个有效估计.这可表述为:

$$\min_{\omega,b,\xi,\xi^*}\frac{1}{2}\|\omega\|^2+\frac{C}{p}\sum_{i=1}^{l}(\xi_i^{p}+(\xi_i^*)^{p}), \qquad (3)$$

约束条件是

$y_i-(\omega\cdot x_i+b)\leqslant\varepsilon+\xi_i;\ -y_i+(\omega\cdot x_i+b)\leqslant\varepsilon+\xi_i^*;\ \xi_i,\xi_i^*\geqslant 0, i=1,2,\cdots,l$,

其中

收稿日期: 2007-08-31
作者简介: 周德强(1975-),男,硕士,讲师,主要从事机器学习理论研究工作。

$$L_\varepsilon = |y-(w\cdot x+b)|_\varepsilon^p = \begin{cases} 0, \text{当} |y-(w\cdot x+b)|\leqslant \varepsilon, \\ (|y-(w\cdot x+b)|-\varepsilon)^p, \text{其它}, \end{cases}$$

是 Vapnik的 ε-不敏感损失函数[1,3].对所有 i,若 $y_i\in\{-1,1\}$,可以在相同的样本集上研究 SVMR与 SVMC之间的关系.

定理 1[2] 假定分类问题(2)由正则化参数 C求解,最优解为 (w,b),那么,存在一个数 $a\in(0,1)$,使得 $\forall\varepsilon\in[a,1)$,若回归问题(3)由正则化参数 $(1-\varepsilon)C$ 求解,则最优解为 $(1-\varepsilon)(w,b)$.

2 SVMR与 SVMC解的关系

通常情况下,SVMC与 SVMR在求解的过程中使用一个非线性核 $K(x_i,y_i)$[2,4].为了表述简单,选择超平面的一个线性表达形式,所得的结论可应用到非线性情况.

注意,当在 $y_i\in\{-1,1\}$上讨论 SVMR问题时,若 $\varepsilon\geqslant 1$, $w=0$,则 $\xi=0$, $\xi^*=0$是对于问题(3)的一个最优解.因此,限制在 $\varepsilon<1$的条件下讨论 SVMR与 SVMC的关系,得到定理 1的推广形式.

定理 2 假定分类问题(2)由正则化参数 C求解,最优解为 (w,b),那么,存在一个数 $a\in(0,1)$,使得 $\forall\varepsilon\in[a,1)$.若回归问题(3)由正则化参数 $(1-\varepsilon)^{2-p}C$ $(p\geqslant 1)$ 求解,则最优解为 $(1-\varepsilon)(w,b)$.

证明 作变量替换:

$$\eta_i=\begin{cases}\xi_i, \text{当 } y_i=1,\\ \xi_i^*, \text{当 } y_i=-1,\end{cases}\quad \eta_i^*=\begin{cases}\xi_i^*, \text{当 } y_i=1,\\ \xi_i, \text{当 } y_i=-1.\end{cases}$$

结合替换以及 $y_i\in\{-1,1\}$产生问题(3)的修正:

$$\min_{w,b,\eta,\eta^*}\frac{1}{2}\|w\|^2+\frac{C}{p}\left(\sum_{i=1}^{l}(\eta_i^p+(\eta_i^*)^p)\right), \tag{4}$$

约束条件是

$y_i(w\cdot x_i+b)\geqslant 1-\varepsilon-\eta_i$; $y_i(w\cdot x_i+b)\leqslant 1+\varepsilon+\eta_i^*$; $\eta_i,\eta_i^*\geqslant 0$, $i=1,2,\cdots,l$.

进一步,在每一个约束条件两边同除以 $1-\varepsilon$,并作变量替换:

$$w'=\frac{w}{1-\varepsilon}, b'=\frac{b}{1-\varepsilon}, e=\frac{\eta}{1-\varepsilon}, e^*=\frac{\eta^*}{1-\varepsilon}$$

得到

$$\min_{w',b',e,e^*}\frac{1}{2}\|w'\|^2+\frac{C(1-\varepsilon)^{p-2}}{p}\left(\sum_{i=1}^{l}(e_i^p+(e_i^*)^p)\right). \tag{5}$$

约束条件是

$y_i(w'\cdot x_i+b')\geqslant 1-e_i$; $y_i(w'\cdot x_i+b')\leqslant \frac{1+\varepsilon}{1-\varepsilon}+e_i^*$; $e_i,e_i^*\geqslant 0$, $i=1,2,\cdots,l$.

观察(5)式,可假设 $\varepsilon\to 1$,约束条件的第 2个集合将自动满足 $e^*=0$,在此假定下构作拉格朗日对偶可以得到:

$$\min_{\alpha,\beta}\frac{1}{2}\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)D_{ij}(\alpha_j-\alpha_j^*)+\sum_{i=1}^{l}(\alpha_i+\beta_i)\left(\frac{\alpha_i+\beta_i}{C(1-\varepsilon)^{p-2}}\right)^{\frac{1}{p-1}}-\sum_{i=1}^{l}\alpha_i-\frac{C(1-\varepsilon)^{p-2}}{p}\sum_{i=1}^{l}\left(\frac{\alpha_i+\beta_i}{C(1-\varepsilon)^{p-2}}\right)^{\frac{p}{p-1}}+\frac{1+\varepsilon}{1-\varepsilon}\sum_{i=1}^{l}\alpha_i^*, \tag{6}$$

约束条件是

$\sum_{i=1}^{l}\alpha_i y_i=\sum_{i=1}^{l}\alpha_i^* y_i$; $\beta_i\geqslant 0$, $\alpha_i,\alpha_i^*\geqslant 0$; $\alpha_i,\alpha_i^*\leqslant C(1-\varepsilon)^{p-2}$, $i=1,2,\cdots,l$,

其中 D是对称半正定矩阵,其元素 $D_{ij}=y_iy_jx_ix_j$.对所有的 ε充分接近 1时,e^* 将全为零.事实上,当 $e=0$, $e^*=0$时,它们是(6)式中损失函数为零时的可行解,如果 e^* 是正的,对 ε充分接近 1时,(6)式的解 α, α^* 也都是正的.因此,假定 ε充分接近 1,可以从(5)式中排除 e^*,从(6)式中排除 α^*,移走这些项后,得到一个与问题(2)的对偶问题本质上一致的优化问题:

$$\min_{\alpha,\beta}\frac{1}{2}\sum_{i=1}^{l}\alpha_iD_{ij}\alpha_j+\sum_{i=1}^{l}(\alpha_i+\beta_i)\left(\frac{\alpha_i+\beta_i}{C(1-\varepsilon)^{p-2}}\right)^{\frac{1}{p-1}}-\sum_{i=1}^{l}\alpha_i-\frac{C(1-\varepsilon)^{p-2}}{p}\sum_{i=1}^{l}\left(\frac{\alpha_i+\beta_i}{C(1-\varepsilon)^{p-2}}\right)^{\frac{p}{p-1}}, \tag{7}$$

约束条件是

$\sum_{i=1}^{l}\alpha_i y_i=0$; $\beta_i\geqslant 0$, $\alpha_i\geqslant 0$; $\alpha_i\leqslant C(1-\varepsilon)^{p-2}$, $i=1,2,\cdots,l$.

通过这一对偶形式,可恢复稍有修正的 SVMC问题:

$$\min_{w,b,\xi}\frac{1}{2}\|w\|^2+C(1-\varepsilon)^{(p-2)}\left(\sum_{i=1}^{l}\xi_i\right), \tag{8}$$

约束条件是

$y_i(w\cdot x_i+b)\geqslant 1-\xi_i$; $\xi_i\geqslant 0$, $i=1,2,\cdots,l$.

因而定理得证.

注 (I)定理 2是对定理 1在非线性损失函数 L_p 上的直接推广,它包含了 $p=1$的线性情况.

(II)对于 $p=2$的情况,在从(4)式到(5)式的过程中,$(1-\varepsilon)$这一因子自动消除,因此(5)式中的目标函数简化为 $\frac{1}{2}\|w'\|^2+\frac{C}{p}\sum_{i=1}^{l}(e_i^2+(e_i^*)^2))$.定理变为对充分大的 ε,如果求解问题(2)用正则化参数 C得到解 (w,b),那么,同样用正则化参数 C求解问题(3)得到解 $(1-\varepsilon)(w,b)$.

(III)因 L_1与 L_2及相应的 L_ε损失函数具有独特之处(使优化问题较为简单),通常在实际中应用广

泛.但由定理 2可知,对于基于更广范围的 L_p 与 L_X型的凸损失函数,也能将 SVMC问题视为 SVMR问题的一种特殊情况.

3 SVMR与 υ-SVMC解的关系

g-SVMC是 SVMC的推广,所谓 g-SVMC,必须最小化问题:

$$\min_{k,b,d,Y_i} \frac{1}{2}||k||^2 - gd + \frac{C}{p}\sum_{i=1}^{l} Y, \quad (9)$$

约束条件是

$y_i(k\cdot x_i + b) \geqslant d - Y; Y \geqslant 0, i = 1,2,\cdots,l.$

特别地,在 g-SVMC问题中当参数 $g = 0$时,d可取值 $d = 1$[4],此时 g-SVMC问题就是 SVMC问题.当取 $p = 1$,并考虑线性可分情况时,有

定理 3[5] 设带有正则化参数 C 且 $d \geqslant 1$的 g-SVMC问题(9)的解为 (w,b,d),于是存在一个常数 a,当 $a \in (0,1)$,使对任意的 $X \in (a,1)$,存在带有正则化参数为 $(1 - X)C$的 SVMR问题(3)的解为 $(1 - X)(k,b)$.

对更一般的非线性损失函数 L_p 与 $L_X(p > 1)$,有

定理 4 设带有正则化参数 C 且 $d \geqslant 1$的 g-SVMC问题(9)的解为 (w,b,d),于是存在一个常数 a,当 $a \in (0,1)$,使对任意的 $X \in (a,1)$,存在带有正则化参数为 $(1 - X)^{2-p}C$的 SVMR问题(3)的解为 $(1 - X)(k,b)$.

可采用类似定理 3的证明方法证明定理 4,这里只给出要证明问题(3)所对应的修正后的 g-SVMC问题:

$$\min_{k,b,Y} \frac{1}{2}||k||^2 - gd + \frac{C}{p}(1 - X)^{p-2}\sum_{i=1}^{l} Y, \quad (10)$$

约束条件是

$y_i(k\cdot x_i + b) \geqslant 1 + d - Y; Y \geqslant 0, d \geqslant 1, i = 1,2,\cdots,l.$

在 g-SVMC问题中,当 $g = 0, d = 1$时,g-SVMC问题就是 SVMC问题,由此可得到

推论 1 设带有正则化参数 C且 SVMC问题(2)的解为 (w,b),于是存在一个常数 a,当 $a \in (0,1)$,使对任意的 $X \in (a,1)$,存在带有正则化参数为 $(1 - X)^{2-p}C$的 SVMR问题(3)的解为 $(1 - X)(k,b)$.

推论 1恰好是定理 2中所反映出的 SVMR与 SVMC解之间的关系,而当 $p = 1$时即为定理 1.

由上述定理可见,在 SVM技术中,凸损失函数的不同选择,定义了不同结构的学习问题.但不论损失函数是线性还是非线性形式,对一个给定的 SVMC问题,总存在某一个具有特定参数的 SVMR问题,使得两者的解等价.

参考文献:

[1] Vapnik V. The nature of statistical learning theory[M]. Heidellberg, Germany: Springer Verlag, 1995.

[2] Pontil M. Ryan rifkin and theodoros evgeniou, from regression to classification in support vector machines, Proceedings-European Symposium on artifical new works [C]. Bruges(Belgium): D-Facto Public, 1999: 21-23.

[3] Vapnik V. Statistical learning theory[M]. New York: Jhon Wiley, 1998.

[4] Burges C J C. A tutorial on support vector machines for pattern recognition [J]. Kata Mining and Knowledge Kiscovery, 1998, 2(2): 121-167.

[5] 邹斌,李落清.支持向量机回归与 υ支持向量机分类解的关系 [J].湖北大学学报:自然科学版, 2004, 26(1): 12-15.

(责任编辑:尹 闯)

科学家发现调控恐惧记忆的关键因素

5-羟色胺是神经系统内一个非常重要的神经递质,广泛参与各种行为和生理过程。一般认为,5-羟色胺功能低下可以导致多种精神类疾病,尤其是焦虑、抑郁和创伤后应激综合征等,而这些疾病都伴有学习和记忆的障碍。中国科学院的科学家们利用条件性基因敲除等遗传学手段,得到中枢神经系统中的 5-羟色胺神经元缺乏的小鼠,比较分析 5-羟色胺敲除小鼠与正常小鼠在焦虑和恐惧行为学测试中的表现。这种敲除手段不影响神经系统其他神经元包括表达 5-羟色胺受体的神经元的产生。研究结果发现,5-羟色胺敲除小鼠焦虑水平低于对照小鼠,在特定环境给予足部电刺激的恐惧记忆测试中,5-羟色胺敲除小鼠的恐惧记忆不仅增强而且长时间保持,而正常小鼠的恐惧记忆在短期内消退。科学家们再进行脑内直接给予 5-羟色胺至小鼠 5-羟色胺敲除小鼠后的分析测试,发现这种增强的恐惧记忆能够像对照小鼠一样消退,这进一步说明增强的恐惧记忆是由 5-羟色胺缺乏引起的。这个研究结果将有助于建立对人类创伤后应激综合征的治疗策略。

(据科学时报)