中国农业大学学报　2006,11(6):92-95
Journal of China Agricultural University

线性支持向量分类机优化问题解的二阶充分条件

蔡 春[1,3]　刘宝光[2]　邓乃扬[3]
(1. 北京联合大学 应用文理学院,北京 100083;2. 北京理工大学 理学院,北京 100086;
3. 中国农业大学 理学院,北京 100083)

摘 要　优化问题的二阶充分条件是研究灵敏度分析的基础,支持向量机是数据挖掘的新方法。针对线性支持向量分类机优化问题,研究了其解的二阶充分条件,给出了二阶充分条件成立的假设条件。研究表明,该假设条件很弱,用支持向量机算法求解实际问题时,通常假定这一条件成立,特别地,对线性可分支持向量分类机优化问题,其解一定满足这一条件,满足二阶充分条件成为当然成立的事实。

关键词　支持向量分类机;二阶充分条件;数据挖掘;起作用约束
中图分类号　TP 13　　**文章编号**　1007-4333(2006)06-0092-04　　**文献标识码**　A

Second order sufficient condition for optimization problem of a linear support vector classifier

Cai Chun[1,3], Liu Baoguang[2], Deng Naiyang[3]
(1. College of Arts and Science, Beijing Union University, Beijing 100083, China; 2. College of Science, Beijing Technology University, Beijing 100086, China; 3. College of Science, China Agricultural University, Beijing 100083, China)

Abstract　A second-order-sufficient-condition is one of basis for sensitivity analysis of optimization problems. A hypothesis to ensure the tenability of a second-order-sufficient-condition of a linear support vector classifier is presented in the paper, which is a weak one for the solution of a linearly separable support vector classifier must meet it. In addition, the problem of a linear support vector classifier is solved under such a hypothesis.

Key words　support vector classifier; second order sufficient condition; data mining; active constraint

二阶充分条件成立是进行非线性规划解灵敏度分析的重要前提。非线性规划的解不一定满足二阶充分条件,但二阶充分条件除保证最优性外,还常常是解的某种正则性的构成部分,因而判定非线性规划解是否满足二阶充分条件,是进行非线性规划解灵敏度分析的一个重要步骤。验证非线性规划解的二阶充分条件是否满足的工作量很大且非常烦琐。支持向量机(support vector machines, SVM)是数据挖掘的新方法,支持向量机分类方法是在统计学习理论[1]这一框架下产生的,在应用中已初步表现出很多优于已有方法的性能,成为一种新的通用机器学习方法[2]。SVM 正在成为继人工神经网络(ANN)之后新的研究热点,并将有力地推动机器学习理论和技术的发展[3-7]。

对于支持向量机优化问题,近几年国内外研究大多集中在支持向量机优化问题的求解和数据挖掘功能方面,通过求解支持向量机优化问题的 Wolfe 对偶问题的算法(除大型问题)已趋于成熟,且取得了广泛的应用;进行支持向量分类机优化问题解的灵敏度分析是新的研究方向。支持向量分类机优化问题是一类特殊的非线性规划(目标函数是凸函数,约束条件是线性函数),关于此特殊凸二次规划问题二阶充分条件是否有别于一般非线性规划情形的研究未见报道。本研究对支持向量分类机优化问题解的二阶充分条件进行研究,以期为非线性规划解灵敏度分析奠定基础。

收稿日期:2006-09-15
基金项目:国家自然科学基金资助项目(10371131)
作者简介:蔡春,博士,主要从事运筹学、模式识别的研究,E-mail:caihun@ygi.edu.cn;邓乃扬,教授,博士生导师,通讯作者,主要从事最优化计算方法、数据挖掘算法的研究,E-mail:dengnaiyang@vip.163.com

线性支持向量分类机是支持向量分类机问题的特殊形式,但线性支持向量分类机很好地体现了支持向量分类机的基本思想[8-9],因此本研究以线性支持向量分类机优化问题为研究对象,揭示支持向量分类机优化问题解的二阶充分条件性质。

1　引理及定理

考虑线性支持向量分类机优化问题

$$\min_{w,b,\xi}\tau(w,b,\xi)=\frac{1}{2}\|w\|^2+\sum_{i=1}^{m}C_i\xi_i$$
$$\text{s.t}\ \ g_i(w,b,\xi)=-y_i(x_i\cdot w+b)-\xi_i+1\le 0\quad i=1,2,\cdots,m$$
$$g_{m+i}(w,b,\xi)=-\xi_i\le 0,\quad i=1,2,\cdots,m \tag{1}$$

其中$(x_i,y_i)\in\mathbf{R}^n\times\{-1,+1\}$。

在问题(1)中去掉$\xi_i(i=1,2,\cdots,m)$,从而去掉目标函数中的加权项$\sum_{i=1}^{m}C_i\xi_i$和约束函数$g_1,g_2,\cdots,g_m$中含项$\xi_i(i=1,2,\cdots,m)$,并去掉约束$g_{m+1},g_{m+2},\cdots,g_{2m}$,优化问题(1)就化为线性可分支持向量分类机优化问题

$$\min_{w,b}\tau(w,b)=\frac{1}{2}\|w\|^2$$
$$\text{s.t}\ \ g_i(w,b)=-y_i(x_i\cdot w+b)+1\le 0\quad i=1,2,\cdots,m \tag{2}$$

设(w^*,b^*,ξ^*)是优化问题(1)的最优解,相应于(w^*,b^*,ξ^*)训练数据$(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)$分为A、B、C 3类。

A类:超平面$w^*\cdot x+b^*=1$上$y_i=+1$的点和超平面$w^*\cdot x+b^*=-1$上$y_i=-1$的点,为方便记此类点为$(x_1,y_1),(x_2,y_2),\cdots,(x_t,y_t)$。

B类:开半空间$w^*\cdot x+b^*>1$中$y_i=+1$的点和开半空间$w^*\cdot x+b^*<-1$中$y_i=-1$的点,记为$(x_{t+1},y_{t+1}),(x_{t+2},y_{t+2}),\cdots,(x_s,y_s)$。

C类:开半空间$w^*\cdot x+b^*<1$中$y_i=+1$的点和开半空间$w^*\cdot x+b^*>-1$中$y_i=-1$的点,记为$(x_{s+1},y_{s+1}),(x_{s+2},y_{s+2}),\cdots,(x_m,y_m)$。

对于线性可分支持向量分类机优化问题(2),由于训练点(x_i,y_i),$y_i=+1$只能满足$w^*\cdot x+b^*=1$或者$w^*\cdot x+b^*>1$之一情况,同理(x_i,y_i),$y_i=-1$只能满足$w^*\cdot x+b^*<-1$或$w^*\cdot x+b^*<-1$之一情况。亦即训练数据$(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)$对于线性可分支持向量分类机优化问题(2)只分为A、B 2类。

对于定义了的3类点,以A、B、C记这3类数据点的下标的集合。

$$A=\{i|\ w\cdot x_i+b=1,y_i=+1\ \text{或}\ w\cdot x_i+b=-1,y_i=-1\}=\{1,2,\cdots,t\} \tag{3}$$

$$B=\{i|\ w\cdot x_i+b>1,y_i=+1\ \text{或}\ w\cdot x_i+b<-1,y_i=-1\}=\{t+1,t+2,\cdots,s\} \tag{4}$$

$$C=\{i|\ w\cdot x_i+b>1,y_i=+1\ \text{或}\ w\cdot x_i+b>-1,y_i=-1\}=\{s+1,s+2,\cdots,m\} \tag{5}$$

设(w^*,b^*,ξ^*)是线性支持向量分类机优化问题(1)的最优解,有了类数据点的概念,则有下面引理成立。

引理　设(w^*,b^*,ξ^*)为优化问题(1)的最优解,则起作用约束指标集为

$$I(w^*,b^*,\xi^*)=A\cup(m+A)\cup(m+B)\cup C=\{1,2,\cdots,t,m+1,\cdots,m+t,m+t+1,\cdots,m+s,s+1,\cdots,m\} \tag{6}$$

而对于优化问题(2),起作用约束指标集为

$$I(w^*,b^*)=A=\{1,2,\cdots,t\} \tag{7}$$

证明:由起作用约束的定义[10]可得式(6)和(7)。

定理　设(w^*,b^*,ξ^*)为优化问题(1)的最优解,对应的KT乘子为$\alpha_1^*,\alpha_2^*,\cdots,\alpha_{2m}^*$,若存在$i\in A$,使得$0<\alpha_i^*<C_i$,则$(w^*,b^*,\xi^*)$满足二阶充分条件。

证明:因为(w^*,b^*,ξ^*)为优化问题(1)的最优解,对应的KT乘子为$\alpha_i^*,\alpha_2^*,\cdots,\alpha_{2m}^*$,优化问题(1)为凸二次规划,所以$(w^*,b^*,\xi^*)$为优化问题(1)的KT点[11],即$\alpha^*=(a_1^*,\alpha_2^*,\cdots,\alpha_{2m}^*)^{\mathrm T}\ge 0$,使得KT条件式(8)～(12)成立

$$w^*-\sum_{i=1}^{m}\alpha_i^*y_ix_i=0 \tag{8}$$

$$\sum_{i=1}^{m}\alpha_i^*y_i=0 \tag{9}$$

$$C_i-\alpha_i^*-\alpha_{m+i}^*=0 \tag{10}$$

$$\alpha_i^*(-y_i(w^*\cdot x_i+b^*)-\xi_i+1)=0 \tag{11}$$

$$\alpha_{m+i}^*(-\xi_i^*)=0 \tag{12}$$

其中$i=1,2,\cdots,m$。对于优化问题(1)的拉格朗日函数

$$L(w,b,\xi)=\tau(w,b,\xi)+\sum_{i=1}^{2m}\alpha_i^*g_i(w,b,\xi) \tag{13}$$

可得

$$\nabla^2 L(w^*, b^*, \xi^*) = \nabla^2 \tau(w^*, b^*, \xi^*) + \sum_{i=1}^{2m} \alpha_i^* \nabla^2 g_i(w^*, b^*, \xi^*) \qquad (14)$$

将目标函数和约束函数代入式(14)整理得

$$\nabla^2 L(w^*, b^*, \xi^*) = \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}_{(n+1+m)\times(n+1+m)}$$

令

$$Z = \{ z \in \mathbf{R}^{n+1+m} \mid z^{\mathrm{T}} \nabla g_i(w^*, b^*, \xi^*) \le 0,\ i \in I;\ z^{\mathrm{T}} \nabla g_i(w^*, b^*, \xi^*) = 0,\ i \in I_+ \} \qquad (15)$$

其中 $I_+ = \{ i \in I \mid \alpha_i^* > 0 \}$。对任意的 $z = (z_1, z_2, \cdots, z_{n+1}, z_{n+2}, \cdots, z_{n+1+m})^{\mathrm{T}}$ 都有

$$z^{\mathrm{T}} \nabla^2 L(w^*, b^*, \xi^*) z = z_1^2 + z_2^2 + \cdots + z_n^2 \qquad (16)$$

可见,只要 $z_1, z_2, \cdots, z_n$ 不全为0,就有

$$z^{\mathrm{T}} \nabla^2 L(w^*, b^*, \xi^*) z > 0 \qquad (17)$$

采用反证法证明对任意的 $z = (z_1, z_2, \cdots, z_{n+1}, z_{n+2}, \cdots, z_{n+1+m})^{\mathrm{T}} \in Z$, $z \ne 0$,式(17)成立,即假设 $z_1, z_2, \cdots, z_n$ 全为0,就可以得到 $z_{n+1} = z_{n+2} = \cdots = z_{n+1+m} = 0$ 与 $z \ne 0$ 矛盾。证明如下:

由引理得到所有起作用约束对应的梯度为

$$\nabla g_i(w, b, \xi) = (-y_i x_i, -y_i, -1, 0 \cdots 0)^{\mathrm{T}} (i = 1, 2, \cdots, t, s+1, m) \qquad (18)$$

↑

第$(n+1+i)$位置

$$\nabla g_{m+i}(w, b, \xi) = (0, 0, -1, 0 \cdots 0)^{\mathrm{T}} (i = 1, 2, \cdots, s) \qquad (19)$$

↑

第$(n+1+i)$位置

对于 $i \in A$,在 $z_1, z_2, \cdots, z_n$ 全为0的假设下,满足

$$z^{\mathrm{T}} \nabla g_i = -z_{n+1} y_i + z_{n+1+i}(-1) \le 0 \qquad (20)$$

及

$$z^{\mathrm{T}} \nabla g_{m+i} = z_{n+1+i}(-1) \le 0 \qquad (21)$$

由假设知存在 $i_0 \in A$,有 $0 < a_{i_0}^* < C_{i_0}$,由(10)式有 $a_{m+i_0}^* = C_{i_0} - a_{i_0}^* > 0$,因而有 $i_0 \in I_+$ 和 $m + i_0 \in I_+$,对此 i_0,式(20)和(21)都应取等号,由此得到 $z_{n+1+i_0} = 0$,以及

$$z_{n+1} = 0 \qquad (22)$$

对于 $i = 1, 2, \cdots, t_1 \le t$,不妨设有 $\alpha_i^* > 0$,即 x_i 为支持向量,由假设知 $t_1 \ge 1$,因$\{1, 2, \cdots, t_1\} \subset I_+$,对任意的 $i \in \{1, 2, \cdots, t_1\}$,式(20)等号成立,结合式(22)得 $z_{n+1+i} = 0 (i = 1, 2, \cdots, t_1)$,对于 $i = t_1 + 1, t_1 + 2, \cdots, t$ 有 $\alpha_i^* = 0$,由式(10)得 $\alpha_{i+m}^* = C_i$, $\forall i \in I_+$,式(21)取等号,由此得到 $z_{n+1+i} = 0$,此时可以得到 $z_{n+1+i} = 0 (i = t_1 + 1, t_1 + 2, \cdots, t)$。这样便得到了

$$z_{n+1+1} = z_{n+1+2} = \cdots = z_{n+1+t} = 0 \qquad (23)$$

对于 $i \in B$, $i = t+1, t+2, \cdots, s$ 因为 $\alpha_i^* = 0$, $\xi_i^* = 0$, $\alpha_{i+m}^* = C_i > 0$,知 $m + i \in I_+$, $\forall i \in I_+$,都有

$$z^{\mathrm{T}} \nabla g_{m+i} = z_{n+1+i}(-1) = 0$$

此时可以得到 $z_{n+1+i} = 0$,这样便得到了

$$z_{n+1+t+1} = z_{n+1+t+2} = \cdots = z_{n+1+s} = 0 \qquad (24)$$

对于 $i \in C$, $i = s+1, s+2, \cdots, m$,由于 $\xi_i^* > 0$, $\alpha_{i+m}^* = 0$, $\alpha_i^* = C_i > 0$,知 $i \in I_+$, $\forall i \in I_+$,都有

$$z^{\mathrm{T}} \nabla g_i = -z_{n+1} y_i + z_{n+1+i}(-1) = 0$$

由式(22)可以得到 $z_{n+1+i} = 0$。这样便得到了

$$z_{n+1+s+1} = z_{n+1+s+2} = \cdots = z_{n+1+m} = 0 \qquad (25)$$

综合式(22)~(25),可以得到若 $z \in Z$, $z_1 = z_2 = \cdots = z_n = 0$ 则, $z_{n+1} = z_{n+2} = \cdots = z_{n+1+m} = 0$,与 $z \ne 0$ 矛盾。而当 $z \ne 0$ 时,必有 $z_1, z_2, \cdots, z_n$ 不全为0,有 $z^{\mathrm{T}} \nabla^2 L(w^*, b^*) z = z_1^2 + z_2^2 + \cdots + z_n^2 > 0$,即$(w^*, b^*, \xi^*)$满足二阶充分条件。证明完毕。

定理显示,对于线性支持向量分类机问题(1)的最优解,只要作很弱的假设,即存在一个 A 类的支持向量对应的乘子 $\alpha_i^* < C_i$,其解一定满足二阶充分条件。这一重要性质对于作为其特例的线性可分问题,成为当然成立的事实。见如下推论。

推论 设(w^*, b^*)为线性可分支持向量分类机问题(2)的最优解,则(w^*, b^*)满足二阶充分条件。

证明:对应于定理的证明,此时有$\nabla^2 L(w^*, b^*) = \nabla^2 f(w^*, b^*) + \sum_{i=1}^{m} \alpha_i^* \nabla^2 g_i(w^*, b^*)$,整理得

$$\nabla^2 L(w^*, b^*) = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}_{(n+1)\times(n+1)} \qquad (26)$$

约束函数的梯度为:

$$\nabla g_i(w, b) = \begin{pmatrix} -y_i x_i \\ -y_i \end{pmatrix} \quad i = 1, 2, \cdots m \qquad (27)$$

由引理知,此时起作用约束集 $I = A$,而且一定存在支持向量,即 $I_+ \neq \phi$,设 $i_0 \in I_+$,在 $z_1, z_2, \cdots, z_n$ 全为 0 的假设下,有

$$z^T \nabla g_i = (z_i, z_2, \cdots, z_{n+1}) \cdot \begin{pmatrix} -y_i[x_i]_1 \\ \cdots \\ -y_i[x_i]_n \\ -y_i \end{pmatrix} = -z_{n+1} y_i = 0 \qquad (28)$$

得

$$z_{n+1} = 0 \qquad (29)$$

可见,若 $z_1, z_2, \cdots, z_n$ 全为 0,则必有 $z = 0$。当 $z \in Z, z \neq 0$,必有 $z_1, z_2, \cdots, z_n$ 不全为 0,即有 $z^T \nabla^2 L(w^*, b^*) z = z_1^2 + z_2^2 + \cdots + z_n^2 > 0$,则 (w^*, b^*) 满足二阶充分条件。证明完毕。

2 结　论

线性支持向量分类机优化问题解的二阶充分条件满足要求至少存在 1 个 A 类点,且对应的乘子满足 $0 < \alpha_i^* < C_i$。这个条件对于作为其特例的线性可分支持向量机问题自然成立。如此看来,线性支持向量分类机优化问题解的二阶充分条件很容易验证。这个条件很弱,很弱的条件为研究线性支持向量分类机解的灵敏度分析奠定很好的基础。线性支持向量分类机优化问题的解 (w^*, b^*, ξ^*) 在很大程度上满足二阶充分条件性质。

参　考　文　献

[1] Vapnik V. 统计学习理论的本质[M]. 张学工,译. 北京:清华大学出版社, 2000: 1-118

[2] Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernelbased Learning Methods[M]. Cambridge: Cambridge University Press, 2000:1-161

[3] Scholkopf B, Burges C J C, Smola A. Advances in Kernel Methods-support Vector Learning[M]. Cambridge: MIT Press,1999: 327-352

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

[5] Scholkopf B, Smola A. Learning With Kernels[M]. Cambridge: MIT Press,2002:123-157

[6] Deng Naiyang, Liu Guangli, Zhang Chunhua. A new version of support vector classification and its application to early warning of food security[J]. OR Transactions, 2003,7(2): 1-8

[7] Hua S J, Sun Z R. Support vector machine approach for protein subcellular localization prediction[J]. Bioinformatics, 2001:721-728

[8] Smola A, Bartlett P, Scholkopf B, et al. Advances in Large Margin Classifiers[M]. Cambridge: MIT Press, 2000:235-267

[9] 邓乃扬, 田英杰. 数据挖掘中的新方法——支持向量分类机[M]. 北京:科学出版社, 2004:49-272

[10] 邓乃扬,诸梅芳. 最优化方法[M]. 沈阳:辽宁教育出版社,1987:142-144

[11] 刘宝光. 非线性规划[M]. 北京:北京理工大学出版社, 1988:46-56