

# Data Perturbation Analysis of the Support Vector Classifier Dual Model

Chun Cai<sup>1</sup>, Xikui Wang<sup>2</sup>

<sup>1</sup>College of Arts and Science, Beijing Union University, Beijing, China

<sup>2</sup>Department of Statistics, University of Manitoba, Winnipeg, Manitoba, Canada

Email: caichun@buu.edu.cn, xikui.wang@umanitoba.ca

**How to cite this paper:** Cai, C. and Wang, X.K. (2018) Data Perturbation Analysis of the Support Vector Classifier Dual Model. *Journal of Software Engineering and Applications*, 11, 459-466.  
<https://doi.org/10.4236/jsea.2018.1110027>

**Received:** April 10, 2018

**Accepted:** October 20, 2018

**Published:** October 23, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

The paper establishes a theorem of data perturbation analysis for the support vector classifier dual problem, from which the data perturbation analysis of the corresponding primary problem may be performed through standard results. This theorem derives the partial derivatives of the optimal solution and its corresponding optimal decision function with respect to data parameters, and provides the basis of quantitative analysis of the influence of data errors on the optimal solution and its corresponding optimal decision function. The theorem provides the foundation for analyzing the stability and sensitivity of the support vector classifier.

## Keywords

Support Vector Classifier, Partial Derivative, Sensitivity, Stability

---

## 1. Introduction

Many methods of data mining exist, including machine learning which is a major research direction of artificial intelligence. The theory of statistics plays a fundamental role in machine learning. However the classic theory of statistics is often based on large sample properties, while in reality we often face small samples, sometimes with a very limited number of observations due to resource or other constraints. Consequently the performance of some large sample methods may be unsatisfactory in certain real applications. Vapnik and collaborators pioneered in the 1960s machine learning for finite samples and developed the statistical learning theory [1] [2]. To deal with the inconsistency between minimizing the empirical risk and minimizing the expected risk in statistical learning, they proposed the principle of structural risk minimization to investigate the

consistency in the machine learning process. Later on, Vapnik and his colleagues at AT & T Bell Laboratory proposed the method of support vector machine [3]. This has then been further developed into the method of support vector classifier (SVC) within statistical learning theory, which has shown satisfactory performance and is becoming a general method of machine learning. We focus our paper on SVC.

Let the training data set be  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\} \in (X \times Y)^m$ , where  $x_i \in X = R^n$  is the training input and  $y_i \in Y = \{-1, +1\}$  is the training output,  $i = 1, 2, \dots, m$ . The essence of the SVC problem for data mining is to seek a real-valued function  $f(x)$  on the training input set  $R^n$  such that the decision function  $sgn f(x)$  infers the corresponding category in the set  $\{-1, +1\}$  of an arbitrary data input  $x \in X = R^n$ , where  $sgn f(x) = -1$  if  $f(x) < 0$  and  $sgn f(x) = 1$  if  $f(x) > 0$  [4] [5]. Because it is not guaranteed that a linear space over the original input space  $R^n$  can be found to separate the training input set, a transformation  $\phi: x \rightarrow \mathbf{x} = \phi(x)$  is often introduced from the input space  $R^n$  to a high-dimensional Hilbert space  $\mathcal{H}$  such that the training input set corresponds to a training set in  $\mathcal{H}$ . Then a super-plane is sought after in  $\mathcal{H}$  to separate the input space to solve the data mining classification problems.

The primary problem of the standard support vector classifier (SVC) is to

$$\begin{aligned} \min_{w \in \mathcal{H}, b \in R, \psi \in R^m} \tau(w, b, \psi) &= \frac{1}{2} w^T w + \sum_{i=1}^m C_i \psi_i \\ \text{subject to } g_i(w, b, \psi) &= -y_i \left( (\mathbf{x}_i)^T w + b \right) - \psi_i + 1 \leq 0, i = 1, 2, \dots, m, \\ g_{m+i}(w, b, \psi) &= -\psi_i \leq 0, i = 1, 2, \dots, m, \end{aligned} \tag{1}$$

where  $T$  stands for transpose,  $C_i > 0, i = 1, \dots, m$  and  $\psi = (\psi_1, \dots, \psi_m)^T$  are penalty parameters,  $w$  consists of the slopes of the super-plane,  $b \in R$  is the intercept of the super-plane, and  $\mathbf{x}_i = \phi(x_i)$ .

The quadratic Wolfe dual problem of the above primary problem (1) is to

$$\begin{aligned} \min_{\alpha} W(\alpha) &= \frac{1}{2} \alpha^T H \alpha - e^T \alpha \\ \text{subject to } h(\alpha) &= \alpha^T y = 0, \\ C_i &\geq h_i(\alpha) = \alpha_i \geq 0, i = 1, 2, \dots, m, \end{aligned} \tag{2}$$

where  $\alpha = (\alpha_1, \dots, \alpha_m)^T$ ,  $y = (y_1, \dots, y_m)^T$ ,  $e = (1, \dots, 1)^T$ ,  $K(x_i, x_j) = (\phi(x_i))^T \phi(x_j)$  is the kernel function, and  $H_{ij} = y_i y_j K(x_i, x_j)$ .

The relationships between the optimal solution  $\alpha_i^*, i = 1, \dots, m$  of the dual problem (2) and the optimal solution  $(w^*, b^*)$  of the primary problem (1) is given by

$$\begin{aligned} w^* &= \sum_{i=1}^m \alpha_i^* y_i \mathbf{x}_i, \\ b^* &= y_i - (\phi(x_i))^T w^* = y_i - \sum_{j=1}^m \alpha_j^* y_j K(x_j, x_i), \text{ for any } \alpha_i^* \in (0, C_i). \end{aligned}$$

The corresponding optimal decision function is given by  $\text{sgn} f(x)$ , where  $f(x) = \sum_{i=1}^m \alpha_i^* y_i K(x_i, x) + b^*$ . Furthermore, each component  $\alpha_i^*$  of the optimal solution  $\alpha^*$  corresponds to a training input, and we call a training input  $x_i$  a support vector if its corresponding  $\alpha_i^* > 0$ .

In data mining, the training input data  $x_i, i=1, 2, \dots, m$  of the support vector classifier (SVC) model are approximation of the true values. When using the approximate values to establish the SVC model, errors in data will inevitably impact on the optimal solution and the corresponding optimal decision function of the SVC model. When the upper bound of data errors is known, the method of data perturbation analysis is used to derive the upper bounds of the optimal solution and its corresponding optimal decision function.

We are interested in the stability and sensitivity analysis of the optimal solution and its corresponding optimal decision function. Our analysis is based on the Fiacco's Theorem on the sensitivity analysis of a general class of nonlinear programming problem [6] [7] [8]. The second order sufficiency condition required for the Fiacco Theorem was further studied in [9]. On the other hand, kernel functions have been used in machine learning [10] [11] [12]. In this paper, we establish a Theorem on data perturbation analysis for a general class of SVC problems with kernel functions. The paper is organized as follows. The main Theorem and its Lemmas are introduced in the next section. Section 3 concludes the paper.

## 2. Data Perturbation Analysis of the SVC Dual Problem

Suppose that  $(w^*, b^*, \psi^*)$  is the optimal solution of the primary problem (1). Corresponding to the solution  $(w^*, b^*, \psi^*)$ , we divide, for convenience, the training data  $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$  into categories  $A, B, C$  as follows:

1)  $A$  category: points satisfying  $(\mathbf{x}_i)^T w^* + b^* = 1$  and  $y_i = +1$ , or  $(\mathbf{x}_i)^T w^* + b^* = -1$  and  $y_i = -1$ . These points are denoted as  $(x_1, y_1), (x_2, y_2), \dots, (x_t, y_t)$  for convenience. Denote  $A = \{1, \dots, t\}$ .

2)  $B$  category: points satisfying  $(\mathbf{x}_i)^T w^* + b^* > 1$  and  $y_i = +1$ , or  $(\mathbf{x}_i)^T w^* + b^* < -1$  and  $y_i = -1$ . These points are denoted as  $(x_{t+1}, y_{t+1}), (x_{t+2}, y_{t+2}), \dots, (x_s, y_s)$  for convenience. Denote  $B = \{t+1, \dots, s\}$ .

3)  $C$  category: points satisfying  $(\mathbf{x}_i)^T w^* + b^* < 1$  and  $y_i = +1$ , or  $(\mathbf{x}_i)^T w^* + b^* > -1$  and  $y_i = -1$ . These points are denoted as  $(x_{s+1}, y_{s+1}), (x_{s+2}, y_{s+2}), \dots, (x_m, y_m)$  for convenience. Denote  $C = \{s+1, \dots, m\}$ .

We call those indices  $i$  such that either  $g_i(w, b, \psi) = 0$  or  $g_{m+i}(w, b, \psi) = 0$  as working constraint indices. Let's start with a lemma.

**Lemma 1.** Suppose that  $(w^*, b^*, \psi^*)$  is the optimal solution of the primary problem (1). Then the set of all working constraint indices is give by

$$\begin{aligned} I(w^*, b^*, \psi^*) &= \{1, \dots, t, m+1, \dots, m+t, m+t+1, \dots, m+s, s+1, \dots, m\} \\ &= A \cup \{m+i : i \in A\} \cup \{m+j : j \in B\} \cup C. \end{aligned}$$

**Proof.** The proof is straightforward from the definitions of working constraint index and the observation that points in A and B categories imply that  $\psi_i = 0$  and points in the C category imply that  $\psi_i > 0$ .  $\square$

Then we offer a sufficient condition for the linear independence of the gradients indexed by the set  $I(w^*, b^*, \psi^*)$ .

**Lemma 2.** Let  $\alpha^* = (\alpha_1^*, \dots, \alpha_m^*)^T$  be the optimal solution of the dual problem (2). If there exists any component  $\alpha_j^*$  of  $\alpha^*$  such that  $0 < \alpha_j^* < C_j$  where  $j \in 1, 2, \dots, m$ , then the set of gradients  $\left\{ \left( \frac{\partial h}{\partial \alpha} \right)^T, \left( \frac{\partial h_i}{\partial \alpha} \right)^T, i \in I(w^*, b^*, \psi^*) \right\}$  is linearly independent.

**Proof.** For  $i \in A$ , suppose there exists  $1 \leq t_1 \leq t$  such that  $\alpha_i^* > 0$  for  $i = 1, 2, \dots, t_1$  but  $\alpha_i^* = 0$  for  $i = t_1 + 1, \dots, t$ . For  $i \in B$ , we always have  $\alpha_i^* = 0$ . For  $i \in C$ , we always have  $\alpha_i^* = C_i$ . The set of gradients

$\left\{ \left( \frac{\partial h}{\partial \alpha} \right)^T, \left( \frac{\partial h_i}{\partial \alpha} \right)^T, i \in I(w^*, b^*, \psi^*) \right\}$  becomes

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ \vdots \\ \vdots \\ y_m \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ -1 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

Since it is assumed that there is a support vector multiplier  $\alpha_i^*$  such that  $0 < \alpha_i^* < C_i$ , the number of gradients in  $\left\{ \left( \frac{\partial h_i}{\partial \alpha} \right)^T, i \in I(w^*, b^*, \psi^*) \right\}$  must be smaller than the dimension  $m$ , which is the dimension of the vector  $y = \left( \frac{\partial h}{\partial \alpha} \right)^T$ .

Therefore the vector  $\left( \frac{\partial h}{\partial \alpha} \right)^T = y$  whose  $m$  components are either  $-1$  or  $+1$  cannot be expressed as a linear combination of the set of gradients  $\left\{ \left( \frac{\partial h_i}{\partial \alpha} \right)^T, i \in I(w^*, b^*, \psi^*) \right\}$ . Hence the set of constrained gradients is linearly independent.  $\square$

We use the lower case letter  $p$  to denote the training input variable and  $p_0$  to denote the training input data. When the form of the kernel function  $\phi(x)$  is known, the following main Theorem investigates how the errors in the training input data impact on the optimal solution and the corresponding optimal decision function of the SVC model.

**Theorem 3.** Suppose that  $\alpha^* = (\alpha_1^*, \dots, \alpha_m^*)^T$  is the optimal solution of the dual problem (2) when  $p = p_0$ , and the corresponding Lagrange multiplier is

$b^*$ . Furthermore,  $g^* = (g_1^*, \dots, g_m^*)^T$  and  $\psi^* = (\psi_1^*, \dots, \psi_m^*)^T = (g_{m+1}^*, \dots, g_{2m}^*)^T$ . Suppose that  $A$  category input vectors  $x_1, x_2, \dots, x_t$  are all support vectors with  $0 < \alpha_i^* < C_i, i = 1, 2, \dots, t$ , and the set of vectors

$$\left( \begin{matrix} y_1 \phi(x_1) \\ y_1 \end{matrix} \right), \left( \begin{matrix} y_2 \phi(x_2) \\ y_2 \end{matrix} \right), \dots, \left( \begin{matrix} y_t \phi(x_t) \\ y_t \end{matrix} \right)$$

corresponding to points in category  $A$  is linearly independent. Then the following results hold.

1) The optimal solution  $\alpha^* = (\alpha_1^*, \dots, \alpha_m^*)^T$  is unique, and the corresponding multiplier  $b^*$  as well as  $g^* = (g_1^*, \dots, g_m^*)^T$  and  $\psi^* = (\psi_1^*, \dots, \psi_m^*)^T = (g_{m+1}^*, \dots, g_{2m}^*)^T$  are all unique.

2) There is a neighbourhood  $N(p_0)$  of  $p_0$  on which there is a unique continuously differentiable function  $y(p) = (\alpha(p), b(p), g(p), \psi(p))$  such that

- a)  $y(p_0) = (\alpha^*, b^*, g^*, \psi^*)$ ,
- b)  $\alpha(p)$  is a feasible solution to the dual problem (2) for any  $p \in N(p_0)$ ,
- c) the partial derivatives of  $y(p) = (\alpha(p), b(p), g(p), \psi(p))$  with respect to data parameters satisfy

$$M(p) \begin{pmatrix} \left( \frac{\partial \alpha}{\partial p} \right)^T \\ \left( \frac{\partial b}{\partial p} \right)^T \\ \left( \frac{\partial g}{\partial p} \right)^T \\ \left( \frac{\partial \psi}{\partial p} \right)^T \end{pmatrix} = M_1(p),$$

where  $M_1(p) = - \left[ \frac{\partial (\nabla_\alpha L)}{\partial p}, g_1 \nabla_p \alpha_1, \dots, g_m \nabla_p \alpha_m, \nabla_p (y^T \alpha) \right]^T$ ,  $\nabla_\alpha L$  is the

gradient of  $L$  with respect to  $\alpha$ ,  $\nabla_p \alpha_i, i = 1, \dots, m$  is the gradient of  $\alpha_i$  with respect to  $p$ ,  $\nabla_p (y^T \alpha)$  is the gradient of  $y^T \alpha$  with respect to  $p$ , and  $L$  is the

Lagrange function of the primary problem (1),  $M(p) = \begin{pmatrix} E & F \\ U & V \end{pmatrix}$ , and

$$E = \begin{pmatrix} y_1^2 K(x_1, x_1) & y_1 y_2 K(x_1, x_2) & \dots & y_1 y_m K(x_1, x_m) \\ y_2 y_1 K(x_2, x_1) & y_2^2 K(x_2, x_2) & \dots & y_2 y_m K(x_2, x_m) \\ \vdots & \vdots & \ddots & \vdots \\ y_m y_1 K(x_m, x_1) & y_m y_2 K(x_m, x_2) & \dots & y_m^2 K(x_m, x_m) \end{pmatrix},$$

$$F = \begin{pmatrix} -1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & y_1 \\ 0 & -1 & \dots & 0 & 0 & 1 & \dots & 0 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & 0 & 0 & \dots & 1 & y_m \end{pmatrix},$$

$$U = \begin{pmatrix} -g_1^* & 0 & \cdots & 0 \\ 0 & -g_2^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -g_m^* \\ \psi_1^* & 0 & \cdots & 0 \\ 0 & \psi_2^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_m^* \\ y_1 & y_2 & \cdots & y_m \end{pmatrix},$$

and

$$V = \begin{pmatrix} -\alpha_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -\alpha_2 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -\alpha_m & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \alpha_1 - C_1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & \alpha_2 - C_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \alpha_m - C_m & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

**Proof.**

Our results follow directly from the Fiacco Theorem after checking the following three conditions required in the Fiacco Theorem:

- The Second Order Sufficiency Condition for  $\alpha^*$ ,*
- Linear Independence of Gradients over the Working Constraint Set,*
- The Strict Complementarity Property.*

**Proof of the Second Order Sufficiency Condition.** Suppose that  $\alpha^* = (\alpha_1^*, \dots, \alpha_m^*)^T$  is the optimal solution to the dual problem. Then there exist multipliers  $b^* \in R, g^* = (g_1^*, \dots, g_m^*)^T \in R^m$  and  $\psi^* = (\psi_1^*, \dots, \psi_m^*)^T = (g_{m+1}^*, \dots, g_{2m}^*)^T \in R^m$  satisfying the Kurash-Kuhn-Tucker Condition:

$$\begin{aligned} H\alpha^* - e + b^*y - g^* + \psi^* &= 0, \\ g_i^* &\geq 0 \text{ for any } i, \\ \psi_i^* &\geq 0 \text{ for any } i, \\ (g^*)^T \alpha^* &= 0, \\ (\psi^*)^T (\alpha^* - C) &= 0, \end{aligned}$$

where  $H = (H_{ij})$  and  $C = (C_1, \dots, C_m)^T$ .

The Hessian matrix corresponding to the Lagrangian function

$$\begin{aligned} L(\alpha^*, b^*, g^*, \psi^*) \\ = \frac{1}{2}(\alpha^*)^T H\alpha^* - e^T \alpha^* + b^* \left( (\alpha^*)^T y \right) - (g^*)^T \alpha^* + (\psi^*)^T (\alpha^* - C) \end{aligned}$$

becomes the matrix  $H$ .

For  $i \in A$ , we have assumed that  $\alpha_i^* > 0$ . For  $i \in B$ , we have  $g_i^* > 0, \alpha_i^* = 0$  and  $\psi_i^* = 0$ . For  $i \in C$ , we have  $g_i^* = 0, \alpha_i^* = C_i$  and  $\psi_i^* > 0$ .

Define the set  $Z = \{z : z^T y = 0; z_i \leq 0, i = 1, \dots, t; z_i = 0, i = t + 1, \dots, m\}$ . Then for any  $z = (z_1, \dots, z_t, 0, \dots, 0) \in Z, z \neq 0$ , we have

$$z^T \nabla^2 L(\alpha^*, b^*, g^*, \psi^*) z = (z_1, \dots, z_t, 0, \dots, 0) H \begin{pmatrix} z_1 \\ \vdots \\ z_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} = (z_1, \dots, z_t) H^* \begin{pmatrix} z_1 \\ \vdots \\ z_t \end{pmatrix} \\ = \left( \sum_{i=1}^t z_i y_i \phi(x_i) \right)^T \left( \sum_{i=1}^t z_i y_i \phi(x_i) \right),$$

where  $\nabla^2 L(\alpha^*, b^*, g^*, \psi^*) = H$  is the Hessian matrix of  $L(\alpha^*, b^*, g^*, \psi^*)$  and  $H^*$  is the upper left  $t \times t$  sub-matrix of  $H$ .

Suppose that  $z_1 y_1 \phi(x_1) + \dots + z_t y_t \phi(x_t) = 0$ . Then because the set of vectors

$$\begin{pmatrix} y_1 \phi(x_1) \\ y_1 \end{pmatrix}, \begin{pmatrix} y_2 \phi(x_2) \\ y_2 \end{pmatrix}, \dots, \begin{pmatrix} y_t \phi(x_t) \\ y_t \end{pmatrix}$$

corresponding to points in category  $A$  is linearly independent and that  $z_1 y_1 + \dots + z_t y_t = 0$ , we must have  $z_1 = \dots = z_t = 0$ . This contradicts with the assumption that  $z \neq 0$ . Hence we must have  $z_1 y_1 \phi(x_1) + \dots + z_t y_t \phi(x_t) \neq 0$ . This implied that  $z^T \nabla^2 L(\alpha^*, b^*, g^*, \psi^*) z > 0$  and therefore the Second Order Sufficiency Condition is satisfied by  $\alpha^*$ .

**Proof of the Linear Independence of Gradients over the Working Constraint Set.** This is Lemma 2.

**Proof of the Strict Complementarity Property.** We need to show that  $g^*$  and  $\alpha^*$  cannot be 0 simultaneously, and  $\psi^*$  and  $\alpha^* - C$  cannot be 0 simultaneously. For  $i \in A$ , we have  $\alpha_i^* > 0$  and hence the intersection between  $A$  and the working constraint set is empty because of the assumption  $\alpha_i^* < C_i, i \in A$ . For  $i \in B$ , we have  $\alpha_i^* = 0$  and multiplier  $g_i^* > 0$ . For  $i \in C$ , we have  $\alpha_i^* = C_i$  and multiplier  $\psi_i^* > 0$ . Therefore the Strict Complementarity Property holds.  $\square$

### 3. Conclusion

Support vector classifier plays an important role in machine learning and data mining. Due to the standard results that connect the primary problem and its dual problem, analysis of the primary problem can be achieved by working on its dual problem. Our main result establishes the equation for solving the partial derivatives of the optimal solution and the corresponding optimal decision function with respect to data parameters. Because the derivative measures the rate of change and a large value of the derivative often implies a large rate of change and hence a sensitive and unstable solution, our main result provides the found-

dition of quantitative analysis based on the derivatives.

### Acknowledgements

This work is funded with grants from the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions CIT & TCD201404080 and New Start Academic Research Projects of Beijing Union University. Xikui Wang acknowledges research support from the Natural Sciences and Engineering Research Council (NSERC) of Canada and from the University of Manitoba.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Vapnik, V.N. and Lerner, A. (1963) Pattern Recognition Using Generalized Portrait Method. *Automation and Remote Control*, **24**, 774-780.
- [2] Vapnik, V.N. (1982) Estimation of Dependence Based on Empirical Data. Springer-Verlag, New York.
- [3] Cortes, C. and Vapnik, V.N. (1995) Support Vector Networks. *Machine Learning* **20**, 273-297. <https://doi.org/10.1007/BF00994018>
- [4] Deng, N. and Tian, Y. (2004) New Data Mining Method—Support Vector Machine. Science Press, Beijing.
- [5] Liu, J., Li, S.C. and Luo, X. (2012) Classification Algorithm of Support Vector Machine via p-Norm Regularization. *Acta Automatica Sinica*, **38**, 76-87. <https://doi.org/10.3724/SP.J.1004.2012.00076>
- [6] Fiacco, A. (1976) Sensitivity Analysis for Nonlinear Programming Using Penalty Methods. *Mathematical Programming*, **10**, 287-331. <https://doi.org/10.1007/BF01580677>
- [7] Liu, B. (1988) Nonlinear Programming. Beijing Institute of Technology Press, Beijing.
- [8] Wu, Q. and Yu, Z. (1989) Sensitivity Analysis for Nonlinear Programming Using Generalized Reduced Gradient Method. *Journal of Beijing Institute of Technology*, **9**, 88-96.
- [9] Cai, C., Liu, B. and Deng, N. (2006) Second Order Sufficient Conditions Property for Linear Support Vector Classifier. *Journal of China Agricultural University*, **11**, 92-95.
- [10] Scholkopf, B., Burges, C.J.C. and Smola, A.J. (1999) Advances in Kernel Methods-Support Vector Learning. MIT Press, Cambridge, MA, 327-352.
- [11] Cristianin, N. and Shawe-Taylor, J. (2000) An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511801389>
- [12] Li, H. and Zhong, B. (2009) Modified Kernel Function for Support Vector Machines Classifier. *Computer Engineering and Applications*, **45**, 53-55.