

# Linear and Nonlinear Classifiers of Data with Support Vector Machines and Generalized Support Vector Machines

Talat Nazir, Xiaomin Qi and Sergei Silvestrov

**Abstract** The support vector machine for linear and nonlinear classification of data is studied. The notion of generalized support vector machine for data classifications is used. The problem of generalized support vector machine is shown to be equivalent to the problem of generalized variational inequality and various results for the existence of solutions are established. Moreover, examples supporting the results are provided.

**Keywords** Linear and nonlinear classification · Support vector machine · Generalized support vector machine · Kernel function

## 1 Support Vector Machine

Support vector machines (SVM) [2, 3, 13, 14, 18] were developed by Vapnik et al. (1995) and are gaining popularity due to many attractive features. As a very powerful tool for data classification and regression, it has been used in many fields, such as text classification [5], facial expression recognition [9], gene analysis [4] and many others [1, 6–8, 10–12, 17, 19–22]. Recently, it has been used for faults classification in a water level control system [15]. And a faults classifier based SVM is used to diagnose the faults for a water level control process [16].

---

T. Nazir (✉) · X. Qi · S. Silvestrov  
Division of Applied Mathematics, School of Education, Culture and Communication,  
Mälardalen University, Box 883, 721 23 Västerås, Sweden  
e-mail: talat@ciit.net.pk

X. Qi  
e-mail: xiaomin.qi@mdh.se

S. Silvestrov  
e-mail: sergei.silvestrov@mdh.se

T. Nazir  
Department of Mathematics, COMSATS Institute of Information Technology,  
Abbottabad 22060, Pakistan

The classification problems can be restricted to consideration of the two-class problems without loss of generality. The goal of support vector classification (SVC) is to separate the two classes by a hyperplane which can also work well on unseen examples. The method is to find the optimal hyperplane that maximizes the margin between two classes of data. The set of data is said to be optimally separated by the hyperplane if it is separated without error and the distance between the closest data is maximal. Support vector classification can be thought of a process using given data to find the decision plane which can guarantee good predictive performance on unseen data. And the process of finding the decision plane is a quadratic programming process.

In this paper, we study the problems of support vector machine and generalized support vector machine. We also show the sufficient conditions for the existence of solutions for problems of generalized support vector machine. We also present various examples to support these results.

Throughout this paper, by  $\mathbb{N}$ ,  $\mathbb{R}$ ,  $\mathbb{R}^n$  and  $\mathbb{R}_n^+$  we denote the set of all natural numbers, the set of all real numbers, the set of all  $n$ -tuples real numbers, the set of all  $n$ -tuples of nonnegative real numbers, respectively.

Also, we consider  $\|\cdot\|$  and  $\langle \cdot, \cdot \rangle$  as Euclidean norm and usual inner product on  $\mathbb{R}^n$ , respectively, such as,  $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 + \dots + x_ny_n$  for all  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ ,  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  in  $\mathbb{R}^n$ . Furthermore, for any two vectors  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ , we say that  $\mathbf{x} \leq \mathbf{y}$  if and only if  $x_i \leq y_i$  for all  $i \in \{1, 2, \dots, n\}$ , where  $x_i$  and  $y_i$  are the components of  $\mathbf{x}$  and  $\mathbf{y}$ , respectively.

## 1.1 Data Classification

Actually, complex real-world applications are always not linearly separable. Kernel representations offer an alternative solution by projecting the data into a higher dimensional feature space to increase the computational power of the linear learning machine.

In order to learn linear or non-linear relations with a linear machine, a set of non-linear features is selected. This is equivalent to applying a fixed non-linear mapping function  $\Phi$  that transforms data in input space  $X$  to data in feature space  $F$ , in which the linear machine can be used. For this classification, both spaces  $X$  and  $F$  need to be vector spaces, where dimension of these two spaces may or may not be same. When the given data is linearly separable, we consider  $\Phi$  as identity operator. For binary classification of data, we consider the decision function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , where the input  $\mathbf{x} = (x_1, \dots, x_n)$  is assigned to the positive class if,  $f(\mathbf{x}) \geq 0$  and otherwise to the negative class. The decision function is defined as

$$f(\mathbf{x}) = \langle \mathbf{w}, \Phi(\mathbf{x}) \rangle + b. \quad (1)$$

This means two steps will be built for non-linear machine: first a fixed non-linear mapping of the data to a feature space, and then a linear machine is used to classify them in the feature space.

In addition, the vector  $\mathbf{w}$  is a linear combination of the support vectors in the training data and can be written as

$$\mathbf{w} = \sum_i \alpha_i \Phi(\mathbf{x}_i), \tag{2}$$

where each  $\alpha_i$  is Lagrange multiplier of the support vectors.

So the decision function can be rewritten as

$$f(\mathbf{x}) = \sigma \left( \sum_i \alpha_i (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x})) + b \right), \tag{3}$$

where  $\sigma$  is a sign function.

The Kernel  $K$  has an associated feature with mapping  $\Phi$ , and it takes two inputs and give their similarity in feature space  $F$ , that is,  $K : F \times F \rightarrow \mathbb{R}$  is defined as

$$K(\mathbf{x}_i, \mathbf{x}) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}). \tag{4}$$

Thus, the decision function from (3) becomes

$$f(\mathbf{x}) = \sigma \left( \sum_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b \right). \tag{5}$$

Some useful kernels for real valued vectors are defined below:

(I) Linear kernel

$$K(\mathbf{x}_i, \mathbf{x}) = \mathbf{x}_i \cdot \mathbf{x}.$$

(II) Polynomial kernel (of degree  $p$ )

$$K(\mathbf{x}_i, \mathbf{x}) = (\mathbf{x}_i \cdot \mathbf{x})^p \text{ or } (\mathbf{x}_i \cdot \mathbf{x} + 1)^p,$$

where  $p$  is a tunable parameter.

(III) Radial Basis Function (RBF) kernel

$$K(\mathbf{x}_i, \mathbf{x}) = \exp[-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2],$$

where  $\gamma$  is a hyperparameter (also called kernel bandwidth). The RBF kernel corresponds to an infinite feature space.

(IV) Sigmoid Kernel

$$K(\mathbf{x}_i, \mathbf{x}) = \tanh(k\mathbf{x}_i \cdot \mathbf{x} + \theta),$$

where  $k$  is a scalar and  $\theta$  is the displacement.

(V) Inverse multi-quadratic kernel

$$K(\mathbf{x}_i, \mathbf{x}) = (\|\mathbf{x}_i - \mathbf{x}\|^2 + \gamma^{-2})^{-1/2},$$

where  $\gamma$  is a hyperparameter (also called kernel bandwidth).

Now, from (1), we define the functional margin of an example  $(\Phi(\mathbf{x}_i), y_i)$  with respect to a hyperplane  $(\mathbf{w}, b)$  to be the quantity

$$\gamma_i = y_i (\langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle + b),$$

where  $y_i \in \{-1, 1\}$ . Note that  $\gamma_i > 0$  implies correct classification of  $(\mathbf{x}_i, y_i)$ . If we replace functional margin by geometric margin we obtain the equivalent quantity for the normalized linear function  $(\frac{1}{\|\mathbf{w}\|} \mathbf{w}, \frac{1}{\|\mathbf{w}\|} b)$ , which therefore measures the Euclidean distances of the points from the decision boundary in the input space.

Actually geometric margin can be written as

$$\tilde{\gamma} = \frac{1}{\|\mathbf{w}\|} \gamma.$$

To find the hyperplane which has maximal geometric margin for a training set  $S$  means to find maximal  $\tilde{\gamma}$ . For convenience, we let  $\gamma = 1$ , the objective function can be written as

$$\max \frac{1}{\|\mathbf{w}\|}.$$

Of course, there are some constraints for the optimization problem. According to the definition of margin, we have  $y_i (\langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle + b) \geq 1, i = 1, \dots, l$ . We rewrite in the equivalent form the objective function with the constraints as

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \quad \text{such that} \quad y_i (\langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle + b) \geq 1, i = 1, \dots, l. \quad (6)$$

We denote this problem by SVM for data classification.

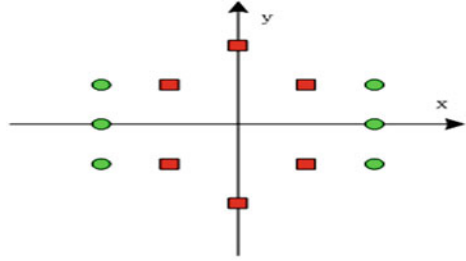
*Example 1* Let's take the group of points  $(0, 2), (0, -2), (1, 1), (1, -1), (-1, 1), (-1, -1)$  as positive class and the group of points  $(2, 0), (-2, 0), (2, 1), (2, -1), (-2, 1), (-2, -1)$  as negative class shown in Fig. 1.

By using the mapping function

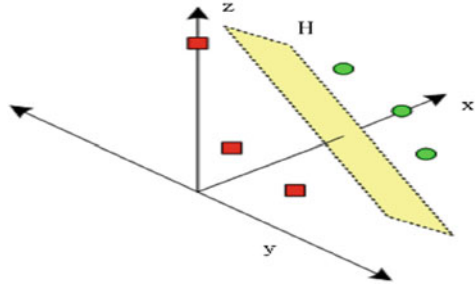
$$\Phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2),$$

which transforms data from two-dimensional input space to three-dimensional feature space, that is  $(1, \sqrt{2}, 1), (1, -\sqrt{2}, 1)$  and  $(0, 0, 4)$  as positive class and  $(4, 2\sqrt{2}, 1), (4, -2\sqrt{2}, 1)$  and  $(4, 0, 0)$  as negative data shown in Fig. 2.

**Fig. 1** The data points given in Example 1



**Fig. 2** The data separation in three dimensional feature space



Now by using this data in three dimensional feature space, we consider the following: For positive points, we have

$$\begin{aligned}
 (w_1, w_2, w_3) \begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix} + b &\geq 1, \\
 (w_1, w_2, w_3) \begin{bmatrix} 1 \\ -\sqrt{2} \\ 1 \end{bmatrix} + b &\geq 1, \\
 (w_1, w_2, w_3) \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix} + b &\geq 1,
 \end{aligned}$$

which implies

$$\begin{aligned}
 w_1 + \sqrt{2}w_2 + w_3 + b &\geq 1, \\
 w_1 - \sqrt{2}w_2 + w_3 + b &\geq 1, \\
 4w_3 + b &\geq 1.
 \end{aligned}$$

For negative points, we have

$$\begin{aligned}
 (w_1, w_2, w_3) \begin{bmatrix} 4 \\ 2\sqrt{2} \\ 1 \end{bmatrix} + b &\leq -1, \\
 (w_1, w_2, w_3) \begin{bmatrix} 4 \\ -2\sqrt{2} \\ 1 \end{bmatrix} + b &\leq -1, \\
 (w_1, w_2, w_3) \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} + b &\leq -1,
 \end{aligned}$$

implying that

$$\begin{aligned}
 4w_1 + 2\sqrt{2}w_2 + w_3 + b &\leq -1, \\
 4w_1 - 2\sqrt{2}w_2 + w_3 + b &\leq -1, \\
 4w_1 + b &\leq -1.
 \end{aligned}$$

From the equations, we get  $\mathbf{w} = (-0.6667, 0, 0)$  with  $\|\mathbf{w}\| = 0.6667$  and shown in Fig. 3.

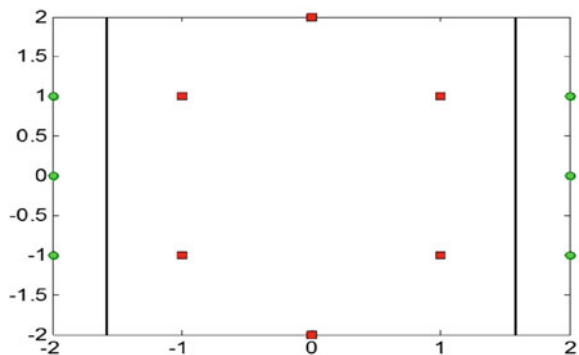
Further, if we use Radial Basis Function (RBF) Kernel  $K(\mathbf{x}_i, \mathbf{x}) = \exp[-\gamma\|\mathbf{x}_i - \mathbf{x}\|^2]$ , with  $\gamma = 1/3$ , we get  $w = (0.0031, 0.0012)$  which is shown in Fig. 4.

Also if we use Sigmoid Kernel  $K(\mathbf{x}_i, \mathbf{x}) = \tanh(k\mathbf{x}_i \cdot \mathbf{x} + \theta)$  with  $k = 1/3$  and  $\theta = 2.85$ , we get  $w = (0, 0)$  shown in Fig. 5.

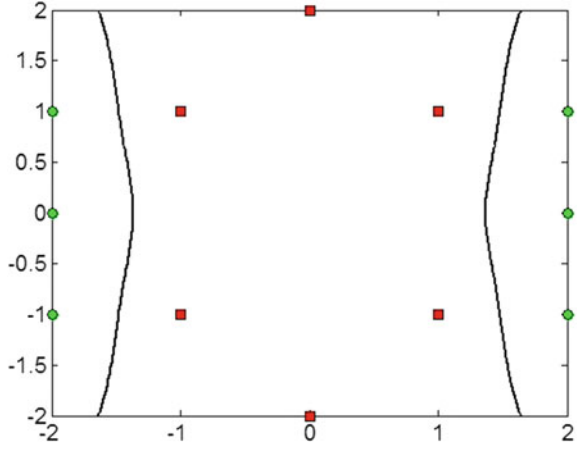
*Example 2* Let us look at another example. The positive data be shown as red square and the negative data be shown as blue circle respectively as shown in Fig. 6.

It is also a non-linear separable problem. Now, if we transfer the original data into the feature space by using the mapping function  $\Phi(\mathbf{x})$ , we can see that the data in the feature space is linear separable see Fig. 7.

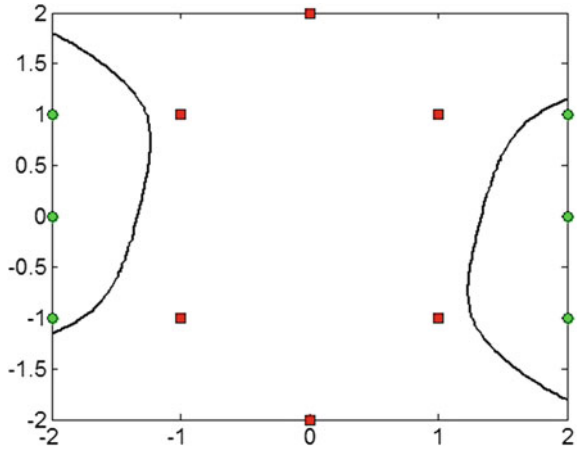
**Fig. 3** The data separation using Polynomial Kernel of degree 2



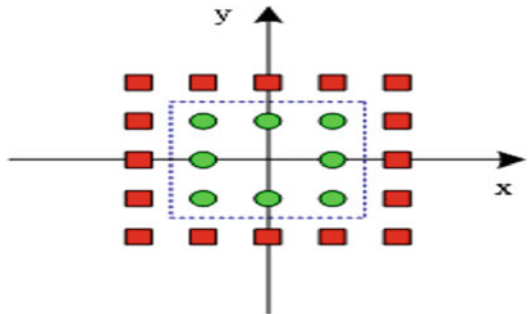
**Fig. 4** The data separation using Radial Basis Function (RBF)



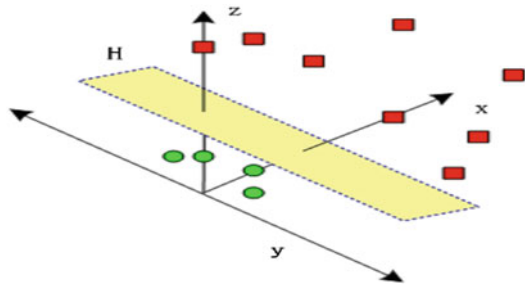
**Fig. 5** The data separation using Sigmoid Kernel



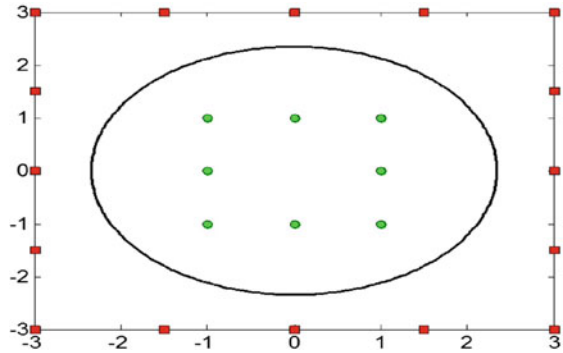
**Fig. 6** The data points given in Example 2



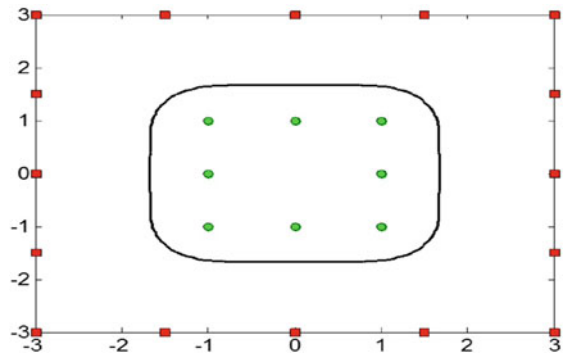
**Fig. 7** The data separation in feature space of Example 2



**Fig. 8** The data separation of Example 2 using Polynomial Kernel



**Fig. 9** The data separation of Example 2 using RBF Kernel



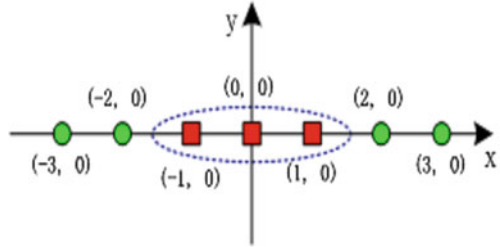
Using Polynomial Kernel with  $p = 2$ , we get  $w = (-0.4898, -0.1633)$  which is shown in Fig. 8.

Next if we use Radial Basis Function (RBF) Kernel  $K(x_i, x) = \exp[-\gamma||x_i - x||^2]$ , with  $\gamma = 2$ , we get  $w = (-0.0016, 0.0014)$  as shown in Fig. 9.

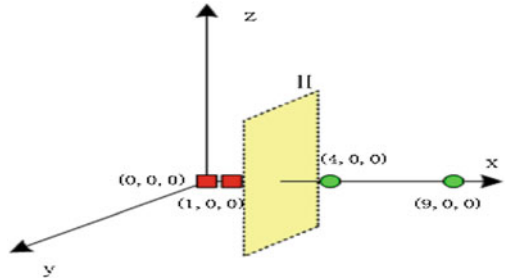
*Example 3* Consider the points  $(0, 0), (1, 0), (-1, 0)$  as positive class and points  $(2, 0), (3, 0), (-2, 0), (-3, 0)$  as negative class see in Fig. 10.



**Fig. 10** The data points given in Example 3



**Fig. 11** Data separation of Example 3 by using Polynomial Kernel of degree 2



Note that, no linear separator exists for this data in the input space. Now, if we use  $\Phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$ , then it transforms two-dimensional data into three-dimensional feature space, which can be separated by hyperplane  $H$  as shown in the Fig. 11.

## 2 Generalized Support Vector Machines

Consider a new control function  $F : \mathbb{R}^p \rightarrow \mathbb{R}^p$  defined as

$$F(\mathbf{x}) = W\Phi(\mathbf{x}) + B, \tag{7}$$

where  $W \in \mathbb{R}^{p \times p}$ ,  $B \in \mathbb{R}^p$  are parameters and  $p$  is the dimension of feature space. In addition,  $W$  contains the  $\mathbf{w}_i$  as a row, where each  $\mathbf{w}_i$  is the linear combination of the support vectors in the feature space and can be written as

$$\mathbf{w}_i = \sum_j \alpha_j^{(i)} \Phi(\mathbf{x}_j), \tag{8}$$

where  $\Phi$  is a mapping that transforms data in input space  $X$  to data in feature space  $F$ . From (7), we obtain

$$\begin{aligned}
 F(\mathbf{x}) &= \begin{bmatrix} \sum_j \alpha_j^{(1)} \Phi(\mathbf{x}_j) \\ \vdots \\ \sum_j \alpha_j^{(p)} \Phi(\mathbf{x}_j) \end{bmatrix} \Phi(\mathbf{x}) + B \\
 &= \begin{bmatrix} \sum_j \alpha_j^{(1)} \Phi(\mathbf{x}_j) \Phi(\mathbf{x}) \\ \vdots \\ \sum_j \alpha_j^{(p)} \Phi(\mathbf{x}_j) \Phi(\mathbf{x}) \end{bmatrix} + B \\
 &= \begin{bmatrix} \sum_j \alpha_j^{(1)} K(\mathbf{x}_j, \mathbf{x}) \\ \vdots \\ \sum_j \alpha_j^{(p)} K(\mathbf{x}_j, \mathbf{x}) \end{bmatrix} + B \\
 &= \begin{bmatrix} \sum_j \alpha_j^{(1)} \\ \vdots \\ \sum_j \alpha_j^{(p)} \end{bmatrix} K(\mathbf{x}_j, \mathbf{x}) + B,
 \end{aligned}$$

where  $K(\mathbf{x}_j, \mathbf{x})$  is the kernel having associated feature with mapping  $\Phi$ .  
 Define

$$\begin{aligned}
 \tilde{\gamma}_k^* &= \mathbf{y}_k (W \Phi(\mathbf{x}_k) + B) \\
 &= \mathbf{y}_k \left( \begin{bmatrix} \sum_j \alpha_j^{(1)} \\ \vdots \\ \sum_j \alpha_j^{(p)} \end{bmatrix} K(\mathbf{x}_j, \mathbf{x}) + B \right) \\
 &= \mathbf{y}_k (\zeta K(\mathbf{x}_j, \mathbf{x}) + B) \geq 1 \quad \text{for } k = 1, 2, \dots, l,
 \end{aligned}$$

where  $\mathbf{y}_k \in \{(-1, -1, \dots, -1), (1, 1, \dots, 1)\}$  is a  $p$ -dimensional vector,  $K(\mathbf{x}_j, \mathbf{x}) = \Phi(\mathbf{x}) \Phi(\mathbf{x}_k)$  and  $\zeta = \begin{bmatrix} \sum_j \alpha_j^{(1)} \\ \vdots \\ \sum_j \alpha_j^{(p)} \end{bmatrix}$ .

**Definition 1** We define a map  $G : \mathbb{R}^p \rightarrow \mathbb{R}_+^p$  by

$$G(\mathbf{w}_i) = (\|\mathbf{w}_i\|, \|\mathbf{w}_i\|, \dots, \|\mathbf{w}_i\|) \quad \text{for } i = 1, 2, \dots, p, \tag{9}$$

where  $\mathbf{w}_i$  are the rows of  $W_{p \times p}$  for  $i = 1, 2, \dots, p$ .

Now, the problem is to find  $\mathbf{w}_i \in \mathbb{R}^p$  that satisfy

$$\min_{\mathbf{w}_i \in W} G(\mathbf{w}_i) \quad \text{such that } \eta \geq 0, \tag{10}$$

where  $\eta = \mathbf{y}_k (\zeta K(\mathbf{x}_j, \mathbf{x}) + B) - 1$ .

We call this problem as the Generalized Support Vector Machine (GSVM).

Note that, if 
$$\begin{bmatrix} \sum_j \alpha_j^{(1)} \\ \vdots \\ \sum_j \alpha_j^{(p)} \end{bmatrix} K(\mathbf{x}_j, \mathbf{x}) = -B, \text{ then } \eta = -1 \text{ and we obtain no solution}$$
 of GSVM problem.

*Example 4* Consider the data of points for positive and negative class as given in Example 1. Then by using polynomial Kernel of degree two, we obtain  $(1, \sqrt{2}, 1)$ ,  $(1, -\sqrt{2}, 1)$ ,  $(0, 0, 4)$  the vectors of positive data and  $(4, 2\sqrt{2}, 1)$ ,  $(4, -2\sqrt{2}, 1)$ ,  $(4, 0, 0)$  the vector negative data in feature space. From positive data points, we have

$$\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \geq \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

$$\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \begin{bmatrix} 1 \\ -\sqrt{2} \\ 1 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \geq \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

$$\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \geq \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

which gives

$$\begin{aligned} w_{11} + \sqrt{2}w_{12} + w_{13} + b_1 &\geq 1, \\ w_{21} + \sqrt{2}w_{22} + w_{23} + b_2 &\geq 1, \\ w_{31} + \sqrt{2}w_{32} + w_{33} + b_3 &\geq 1, \end{aligned}$$

$$\begin{aligned} w_{11} - \sqrt{2}w_{12} + w_{13} + b_1 &\geq 1, \\ w_{21} - \sqrt{2}w_{22} + w_{23} + b_2 &\geq 1, \\ w_{31} - \sqrt{2}w_{32} + w_{33} + b_3 &\geq 1, \end{aligned}$$

$$\begin{aligned} 4w_{13} + b_1 &\geq 1, \\ 4w_{23} + b_2 &\geq 1, \\ 4w_{33} + b_3 &\geq 1. \end{aligned}$$

Also from negative data points,

$$\begin{aligned} & \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \begin{bmatrix} 4 \\ 2\sqrt{2} \\ 1 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \leq \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}, \\ & \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \begin{bmatrix} 4 \\ -2\sqrt{2} \\ 1 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \leq \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}, \\ & \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \leq \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}, \end{aligned}$$

which gives

$$\begin{aligned} 4w_{11} + 2\sqrt{2}w_{12} + w_{13} + b_1 &\leq -1, \\ 4w_{21} + 2\sqrt{2}w_{22} + w_{23} + b_2 &\leq -1, \\ 4w_{31} + 2\sqrt{2}w_{32} + w_{33} + b_3 &\leq -1, \\ \\ 4w_{11} - 2\sqrt{2}w_{12} + w_{13} + b_1 &\leq -1, \\ 4w_{21} - 2\sqrt{2}w_{22} + w_{23} + b_2 &\leq -1, \\ 4w_{31} - 2\sqrt{2}w_{32} + w_{33} + b_3 &\leq -1, \\ \\ 4w_{11} + b_1 &\leq -1, \\ 4w_{12} + b_2 &\leq -1, \\ 4w_{13} + b_3 &\leq -1. \end{aligned}$$

By solving these equations, we get

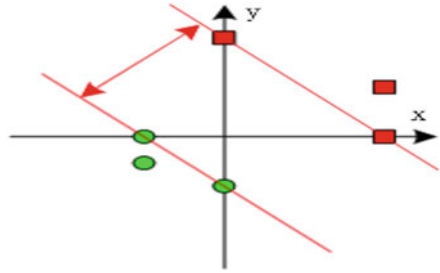
$$W = \begin{bmatrix} -1.39 & -0.512 & -0.627 \\ 0.667 & 0 & -0.667 \\ 0.667 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3.742 \\ 1.047 \\ 1.51 \end{bmatrix},$$

with smallest norm of  $w_i$

$$\min_{w_i \in W} G(w_i) = (0.667, 0.667, 0.667).$$

Hence we get  $w = (0.667, 0, 0)$  that minimize  $G(w_i)$  for  $i = 1, 2, 3$ .

**Fig. 12** Data for situation 1 in Example 5



If we are dealing with the data that can linearly separable, then in the process of G SVM, map  $\Phi$  deals as identity operator. The next example we show the situations for this case.

*Example 5* Let us consider the three categories of data:

**Situation 1** Suppose that we have data  $(2, 0)$ ,  $(0, 2)$ ,  $(2, 1)$  as positive class and data  $(-1, 0)$ ,  $(0, -1)$ ,  $(-1, -1/2)$  as negative class shown in Fig. 12.

For positive points, we have  $(2, 0)$ ,  $(0, 2)$ ,  $(2, 1)$ , so

$$\begin{aligned} \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\geq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \\ \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\geq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \\ \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\geq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \end{aligned}$$

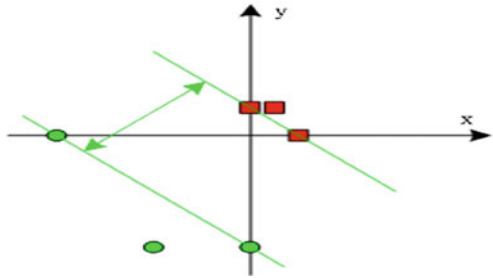
which implies

$$\begin{aligned} \begin{bmatrix} 2w_{11} \\ 2w_{21} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\geq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \\ \begin{bmatrix} 2w_{12} \\ 2w_{22} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\geq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \\ \begin{bmatrix} 2w_{11} + w_{12} \\ 2w_{21} + w_{22} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\geq \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \end{aligned}$$

Again, for the negative points, we have  $(-1, 0)$ ,  $(0, -1)$ ,  $(-1, -1/2)$  and

$$\begin{aligned} \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\leq \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \\ \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\leq \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \\ \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} -1 \\ -1/2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\leq \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \end{aligned}$$

**Fig. 13** The data separation for situation 2



which gives

$$\begin{aligned} \begin{bmatrix} -w_{11} \\ -w_{21} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\leq \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \\ \begin{bmatrix} -w_{12} \\ -w_{22} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\leq \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \\ \begin{bmatrix} -w_{11} - \frac{1}{2}w_{12} \\ -w_{21} - \frac{1}{2}w_{22} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\leq \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \end{aligned}$$

From above equations, we get

$$W = \begin{bmatrix} \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -\frac{1}{3} \\ -\frac{1}{3} \end{bmatrix}.$$

Thus we get

$$\min_{w_i \in W} G(\mathbf{w}_i) = \left( \frac{2\sqrt{2}}{3}, \frac{2\sqrt{2}}{3} \right).$$

Hence we get  $\mathbf{w} = \left( \frac{2}{3}, \frac{2}{3} \right)$  that minimizes  $G(\mathbf{w}_i)$  for  $i = 1, 2$ .

**Situation 2** We consider the data  $(1, 0), (0, 1), (1/2, 1)$  as positive class, data  $(-4, 0), (0, -4), (-2, -4)$  as negative class which is shown in Fig. 13.

Now, for positive points of **Situation 2**, we have  $(1, 0), (0, 1), (1/2, 1)$  and

$$\begin{aligned} \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\geq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \\ \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\geq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \\ \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\geq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \end{aligned}$$

which gives

$$\begin{aligned} \begin{bmatrix} w_{11} \\ w_{21} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\geq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \\ \begin{bmatrix} w_{12} \\ w_{22} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\geq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \\ \begin{bmatrix} \frac{1}{2}w_{11} + w_{12} \\ \frac{1}{2}w_{21} + w_{22} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\geq \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \end{aligned}$$

For negative points for this case, we have

$$\begin{aligned} \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} -4 \\ 0 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\leq \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \\ \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} 0 \\ -4 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\leq \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \\ \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} -2 \\ -4 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\leq \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \end{aligned}$$

which gives

$$\begin{aligned} \begin{bmatrix} -4w_{11} \\ -4w_{21} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\leq \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \\ \begin{bmatrix} -4w_{12} \\ -4w_{22} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\leq \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \\ \begin{bmatrix} -2w_{11} & -4w_{12} \\ -2w_{21} & -4w_{22} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &\leq \begin{bmatrix} -1 \\ -1 \end{bmatrix}. \end{aligned}$$

Thus, we obtain that

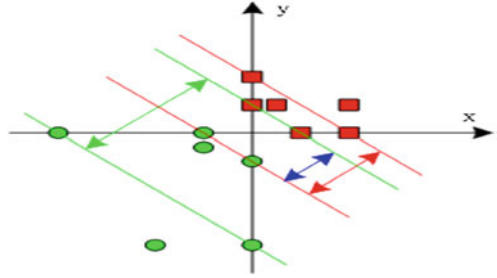
$$W = \begin{bmatrix} \frac{2}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{2}{5} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \frac{3}{5} \\ \frac{3}{5} \end{bmatrix}.$$

Thus we get

$$\min_{i \in \{1,2\}} G(\mathbf{w}_i) = \left( \frac{2\sqrt{2}}{5}, \frac{2\sqrt{2}}{5} \right).$$

Hence we get  $\mathbf{w} = (\frac{2}{5}, \frac{2}{5})$  that minimize  $G(\mathbf{w}_i)$  for  $i = 1, 2$ .

**Fig. 14** The data separation for situation 3



In the next **Situation 3**, we combine of this two groups of data. Now, we have data  $(2, 0), (0, 2), (2, 1), (1, 0), (0, 1), (1/2, 1)$  as positive class and  $(-1, 0), (0, -1), (-1, -1/2), (-4, 0), (0, -4), (-2, -4)$  as negative class see Fig. 14.

For the positive points of the combination, we have

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

and

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

which gives

$$\begin{bmatrix} w_{11} \\ w_{21} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} w_{12} \\ w_{22} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

For negative points for this case, we have

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix},$$

and

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix},$$

which gives

$$\begin{bmatrix} -w_{11} \\ -w_{21} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -w_{12} \\ -w_{22} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$



From this, we obtain that

$$W = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Thus we get

$$\min_{i \in \{1,2\}} G(\mathbf{w}_i) = (\sqrt{2}, \sqrt{2}).$$

Hence we get  $\mathbf{w} = (1, 1)$  that minimize  $G(\mathbf{w}_i)$  for  $i = 1, 2$ .

The problem of GSVM defined in (10) is equivalent to

$$\text{find } \mathbf{w}_i \in W : \langle G'(\mathbf{w}_i), \mathbf{v} - \mathbf{w}_i \rangle \geq 0 \quad \text{for all } \mathbf{v} \in \mathbb{R}^p \quad \text{with } \eta \geq 0. \quad (11)$$

Hence the problem of GSVM becomes to the problem of generalized variational inequality.

Note that if we take  $G'(\mathbf{w}_i) = \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}$ , then from (11), we obtain

$$\text{find } \mathbf{w}_i \in W : \langle \mathbf{w}_i, \mathbf{v} - \mathbf{w}_i \rangle \geq 0 \quad \text{for all } \mathbf{v} \in \mathbb{R}^p \quad \text{with } \eta \geq 0, \quad (12)$$

or

$$\text{find } \mathbf{w}_i \in W : \langle \mathbf{w}_i, \mathbf{v} \rangle \geq \|\mathbf{w}_i\|^2 \quad \text{for all } \mathbf{v} \in \mathbb{R}^p \quad \text{with } \eta \geq 0. \quad (13)$$

We study the sufficient conditions for the existence of solutions for GSVM problems.

**Proposition 1** *Let  $G : \mathbb{R}^p \rightarrow \mathbb{R}_+^p$  be a differentiable operator. An element  $\mathbf{w}^* \in \mathbb{R}^p$  minimizes  $G$  if and only if  $G'(\mathbf{w}^*) = \mathbf{0}$ , that is,  $\mathbf{w}^* \in \mathbb{R}^p$  solves GSVM if and only if  $G'(\mathbf{w}^*) = \mathbf{0}$ .*

*Proof* Let  $G'(\mathbf{w}^*) = \mathbf{0}$ , then for all  $\mathbf{v} \in \mathbb{R}^p$  with  $\eta = \mathbf{y}_k(\zeta K(\mathbf{x}_j, \mathbf{x}) + B) - 1 \geq 0$ ,

$$\langle G'(\mathbf{w}^*), \mathbf{v} - \mathbf{w}^* \rangle = \langle \mathbf{0}, \mathbf{v} - \mathbf{w}^* \rangle = 0,$$

and consequently, the inequality

$$\langle G'(\mathbf{w}^*), \mathbf{v} - \mathbf{w}^* \rangle \geq 0$$

holds for all  $\mathbf{v} \in \mathbb{R}^p$ . Hence  $\mathbf{w}^* \in \mathbb{R}^p$  solves problem of GSVM.

Conversely, assume that  $\mathbf{w}^* \in \mathbb{R}^p$  satisfies

$$\langle G'(\mathbf{w}^*), \mathbf{v} - \mathbf{w}^* \rangle \geq 0 \quad \forall \mathbf{v} \in \mathbb{R}^n \quad \text{such that } \eta \geq 0.$$

Taking  $\mathbf{v} = \mathbf{w}^* - G'(\mathbf{w}^*)$  in the above inequality implies that

$$\langle G'(\mathbf{w}^*), -G'(\mathbf{w}^*) \rangle \geq 0,$$

which further implies

$$-\|G'(\mathbf{w}^*)\|^2 \geq 0,$$

and we get  $G'(\mathbf{w}^*) = \mathbf{0}$ . □

*Remark 1* Note that if  $G'(\mathbf{w}^*) = \mathbf{0}$  at some  $\mathbf{w}^* \in \mathbb{R}^p$ , then we obtain  $\frac{\mathbf{w}^*}{\|\mathbf{w}^*\|} = \mathbf{0}$  which implies  $\mathbf{w}^* = \mathbf{0}$ . Thus it follows from Proposition 2.4 that if  $G'(\mathbf{w}^*) = \mathbf{0}$  at some  $\mathbf{w}^* \in \mathbb{R}^p$ , then  $\mathbf{w}^* = \mathbf{0}$  solves GSVM problem.

*Remark 2* If  $\mathbf{w}^* = \mathbf{0}$ , then from (8), we obtain

$$\sum_j \alpha_j^{(*)} \Phi(\mathbf{x}_j) = \mathbf{0},$$

which implies

$$\sum_j \alpha_j^{(*)} \Phi(\mathbf{x}_j) \Phi(\mathbf{x}) = 0,$$

that is

$$\sum_j \alpha_j^{(*)} K(\mathbf{x}_j, \mathbf{x}) = 0. \tag{14}$$

Since  $\alpha_j^{(*)} > 0$  for all  $j$ , so we have

$$K(\mathbf{x}_j, \mathbf{x}) = 0.$$

**Definition 2** Let  $K$  be a closed and convex subset of  $\mathbb{R}^n$ . Then, for every point  $\mathbf{x} \in \mathbb{R}^n$ , there exists a unique nearest point in  $K$ , denoted by  $P_K(\mathbf{x})$ , such that  $\|\mathbf{x} - P_K(\mathbf{x})\| \leq \|\mathbf{x} - \mathbf{y}\|$  for all  $\mathbf{y} \in K$  and also note that  $P_K(\mathbf{x}) = \mathbf{x}$  if  $\mathbf{x} \in K$ .  $P_K$  is called the metric projection of  $\mathbb{R}^n$  onto  $K$ . It is well known that  $P_K : \mathbb{R}^n \rightarrow K$  is characterized by the properties:

- (i)  $P_K(\mathbf{x}) = \mathbf{z}$  for  $\mathbf{x} \in \mathbb{R}^n$  if and only if  $\langle \mathbf{z} - \mathbf{x}, \mathbf{y} - \mathbf{z} \rangle \geq 0$  for all  $\mathbf{y} \in \mathbb{R}^n$ ;
- (ii) For every  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ,  $\|P_K(\mathbf{x}) - P_K(\mathbf{y})\|^2 \leq \langle \mathbf{x} - \mathbf{y}, P_K(\mathbf{x}) - P_K(\mathbf{y}) \rangle$ ;
- (iii)  $\|P_K(\mathbf{x}) - P_K(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|$ , for every  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ , that is,  $P_K$  is nonexpansive map.

**Proposition 2** Let  $G : \mathbb{R}^p \rightarrow \mathbb{R}_+^p$  be a differentiable operator. An element  $\mathbf{w}^* \in \mathbb{R}^p$  minimize mapping  $G$  defined in (11) if and only if  $\mathbf{w}^*$  is the fixed point of map

$$P_{\mathbb{R}_+^p}(I - \rho G') : \mathbb{R}^p \rightarrow \mathbb{R}_+^p \text{ for any } \rho > 0,$$

that is,

$$\begin{aligned} \mathbf{w}^* &= P_{\mathbb{R}_+^p} (I - \rho G') (\mathbf{w}^*) \\ &= P_{\mathbb{R}_+^p} (\mathbf{w}^* - \rho G' (\mathbf{w}^*)), \end{aligned}$$

where  $P_{\mathbb{R}_+^p}$  is a projection map from  $\mathbb{R}^p$  to  $\mathbb{R}_+^p$  and  $\eta = \mathbf{y}_k (\zeta K(\mathbf{x}_j, \mathbf{x}) + B) - 1 \geq 0$ .

*Proof* Suppose  $\mathbf{w}^* \in \mathbb{R}_+^p$  is solution of GSVM. Then for  $\eta = \mathbf{y}_k (\zeta K(\mathbf{x}_j, \mathbf{x}) + B) - 1 \geq 0$ , we have

$$\langle G' (\mathbf{w}^*), \mathbf{w} - \mathbf{w}^* \rangle \geq 0 \quad \text{for all } \mathbf{w} \in \mathbb{R}^p.$$

Adding  $\langle \mathbf{w}^*, \mathbf{w} - \mathbf{w}^* \rangle$  on both sides, we get

$$\langle \mathbf{w}^*, \mathbf{w} - \mathbf{w}^* \rangle + \langle G' (\mathbf{w}^*), \mathbf{w} - \mathbf{w}^* \rangle \geq \langle \mathbf{w}^*, \mathbf{w} - \mathbf{w}^* \rangle \quad \text{for all } \mathbf{w} \in \mathbb{R}^p,$$

which further implies that

$$\langle \mathbf{w}^* - (\mathbf{w}^* - G' (\mathbf{w}^*)), \mathbf{w} - \mathbf{w}^* \rangle \geq 0 \quad \text{for all } \mathbf{w} \in \mathbb{R}^p,$$

which is possible only if  $\mathbf{w}^* = P_{\mathbb{R}_+^p} (\mathbf{w}^* - \rho G' (\mathbf{w}^*))$ , that is,  $\mathbf{w}^*$  is the fixed point of  $G'$ .

Conversely, let  $\mathbf{w}^* = P_{\mathbb{R}_+^p} (\mathbf{w}^* - \rho G' (\mathbf{w}^*))$  with  $\eta = \mathbf{y}_k (\zeta K(\mathbf{x}_j, \mathbf{x}) + B) - 1 \geq 0$ , then we have

$$\langle \mathbf{w}^* - (\mathbf{w}^* - G' (\mathbf{w}^*)), \mathbf{w} - \mathbf{w}^* \rangle \geq 0 \quad \text{for all } \mathbf{w} \in \mathbb{R}^p,$$

which implies

$$\langle G' (\mathbf{w}^*), \mathbf{w} - \mathbf{w}^* \rangle \geq 0 \quad \text{for all } \mathbf{w} \in \mathbb{R}^p,$$

and so  $\mathbf{w}^* \in \mathbb{R}_+^p$  is the solution of GSVM. □

### 3 Conclusion

The linear and nonlinear data classifications by using support vector machine and generalized support vector machine have been studied. We also studied the sufficient conditions for existence of the solution of generalized support vector machine. Some examples are shown for supporting these results.

**Acknowledgements** Talat Nazir and Xiaomin Qi are grateful to the Erasmus Mundus project FUSION for supporting the research visit to Mälardalen University, Sweden, and to the Research environment MAM in Mathematics and Applied Mathematics, Division of Applied Mathematics, the School of Education, Culture and Communication of Mälardalen University for creating excellent research environment.

## References

1. Adankon, M.M., Cheriet, M.: Model selection for the LS-SVM. Application to handwriting recognition. *Pattern Recognit.* **42**(12), 3264–3270 (2009)
2. Cortes, C., Vapnik, V.N.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
3. Cristianini, N., Shawe-Taylor, J.: *An Introduction to Support Vector Machines and other Kernel Based Learning Methods*. Cambridge University Press, Cambridge (2000)
4. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. *Mach. Learn.* **46**(1–3), 389–422 (2002)
5. Joachims, T.: Text categorization with support vector machines: learning with many relevant features. In: *Proceedings of the European Conference on Machine Learning*. Springer, Heidelberg (1998)
6. Khan, N., Ksantini, R., Ahmad, I., Boufama, B.: A novel SVM+NDA model for classification with an application to face recognition. *Pattern Recognit.* **45**(1), 66–79 (2012)
7. Li, S., Kwok, J.T., Zhu, H., Wang, Y.: Texture classification using the support vector machines. *Pattern Recognit.* **36**(12), 2883–2893 (2003)
8. Liu, R., Wang, Y., Baba, T., Masumoto, D., Nagata, S.: SVM-based active feedback in image retrieval using clustering and unlabeled data. *Pattern Recognit.* **41**(8), 2645–2655 (2008)
9. Michel, P., Kaliouby, R.E.: Real time facial expression recognition in video using support vector machines. In: *Proceedings of ICMI'03*, pp. 258–264 (2003)
10. Noble, W.S.: *Support Vector Machine Applications in Computational Biology*. MIT Press, Cambridge (2004)
11. Shao, Y., Lunetta, R.S.: Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. *ISPRS J. Photogramm. Remote Sens.* **70**, 78–87 (2012)
12. Shao, Y.H., Chen, W.J., Deng, N.Y.: Nonparallel hyperplane support vector machine for binary classification problems. *Inf. Sci.* **263**, 22–35 (2014)
13. Vapnik, V.N.: *The Nature of Statistical Learning Theory*. Springer, New York (1996)
14. Vapnik, V.N.: *Statistical Learning Theory*. Wiley, New York (1998)
15. Wang, D., Qi, X., Wen, S., Deng, M.: SVM based fault classifier design for a water level control system. In: *Proceedings of 2013 International Conference on Advanced Mechatronic Systems*, pp. 152–157. Luoyang, China (2013)
16. Wang, D., Qi, X., Wen, S., Dan, Y., Ouyang, L., Deng, M.: Robust nonlinear control and SVM classifier based fault diagnosis for a water level process. *ICIC Express Lett.* **5**(1), 767–774 (2014)
17. Wang, X.Y., Wang, T., Bu, J.: Color image segmentation using pixel wise support vector machine classification. *Pattern Recognit.* **44**(4), 777–787 (2011)
18. Weston, J., Watkins, C.: *Multi-class support vector machines*. Technical report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London (1998)
19. Wu, Y.C., Lee, Y.-S., Yang, J.-C.: Robust and efficient multiclass SVM models for phrase pattern recognition. *Pattern Recognit.* **41**(9), 2874–2889 (2008)
20. Xue, Z., Ming, D., Song, W., Wan, B., Jin, S.: Infrared gait recognition based on wavelet transform and support vector machine. *Pattern Recognit.* **43**(8), 2904–2910 (2010)
21. Zhao, Z., Liu, J., Cox, J.: Safe and efficient screening for sparse support vector machine. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD 14, pp. 542–551, New York, NY, USA (2014)
22. Zuo, R., Carranza, E.J.M.: Support vector machine: a tool for mapping mineral prospectivity. *Comput. Geosci.* **37**(12), 1967–1975 (2011)