Optimization Methods for Solving Problems of Machine Learning

Prof Talat Nazir
Department of Mathematical Sciences,
University of South Africa

talatn@unisa.ac.za

MOOCTA 2026

Support Vector Machine:

Artificial intelligence data classification is transforming data management by sorting and analyzing data quickly and accurately, helping businesses stay ahead. It empowers organizations to identify their data types, locations, and handle sensitive information securely. This process also ensures compliance with regulations. Moving forward, AI's role in data analysis will grow, machine learning will become more common, and AI will incorporate technologies like cloud computing and big data analytics, elevating data classification further.

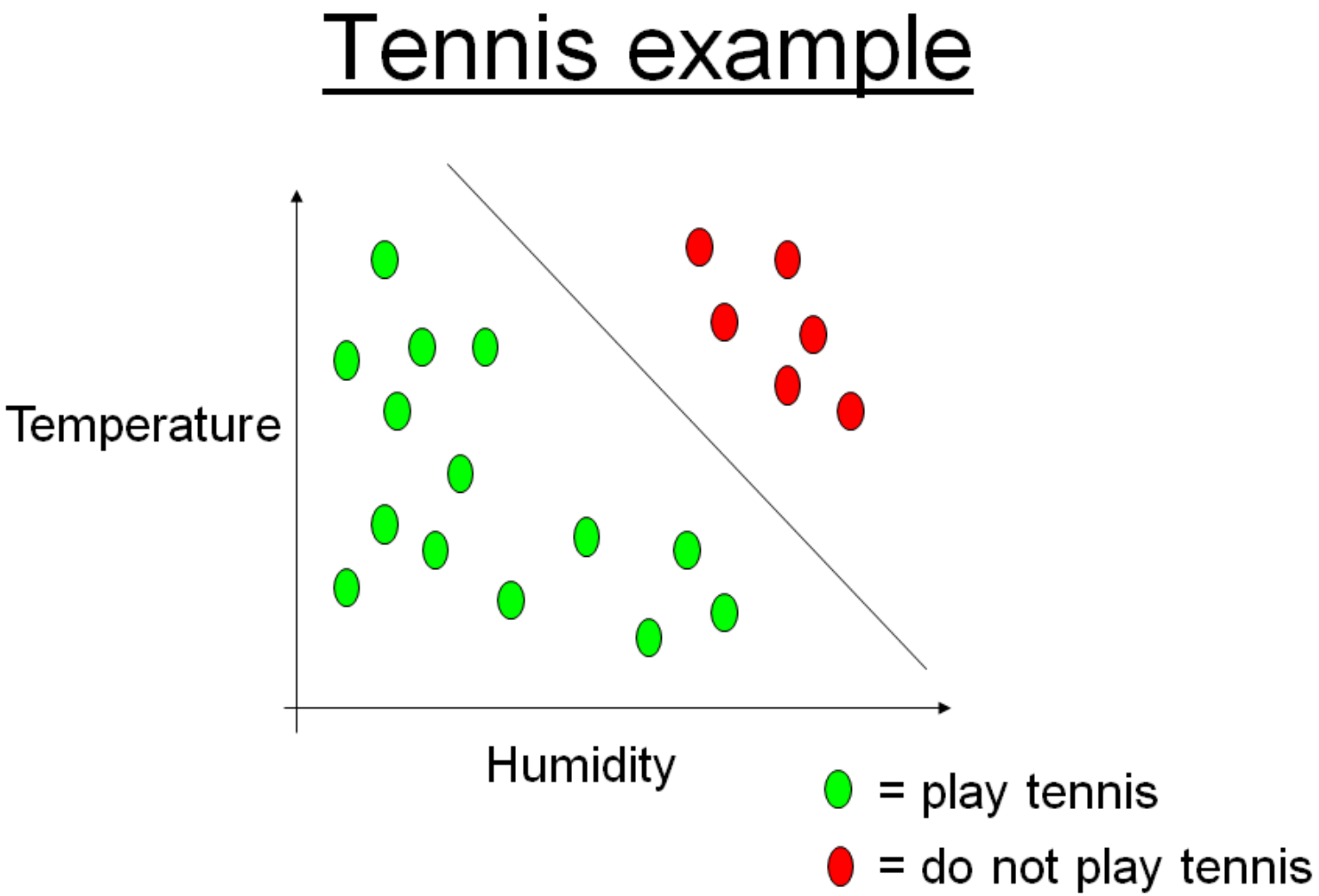
The aim of this talk is to describe the new methods for data separation by using support vector machine (SVM) and generalized support vector machine (GSVM). SVM is a powerful tool for data classification in machine learning.

We show that the problem of GSVM is equivalent to the problem of generalized variational inequality and establish various results for the existence of solutions. In this way, the GSVM problem is equivalent to optimization problem and precisely the fixed point problem.

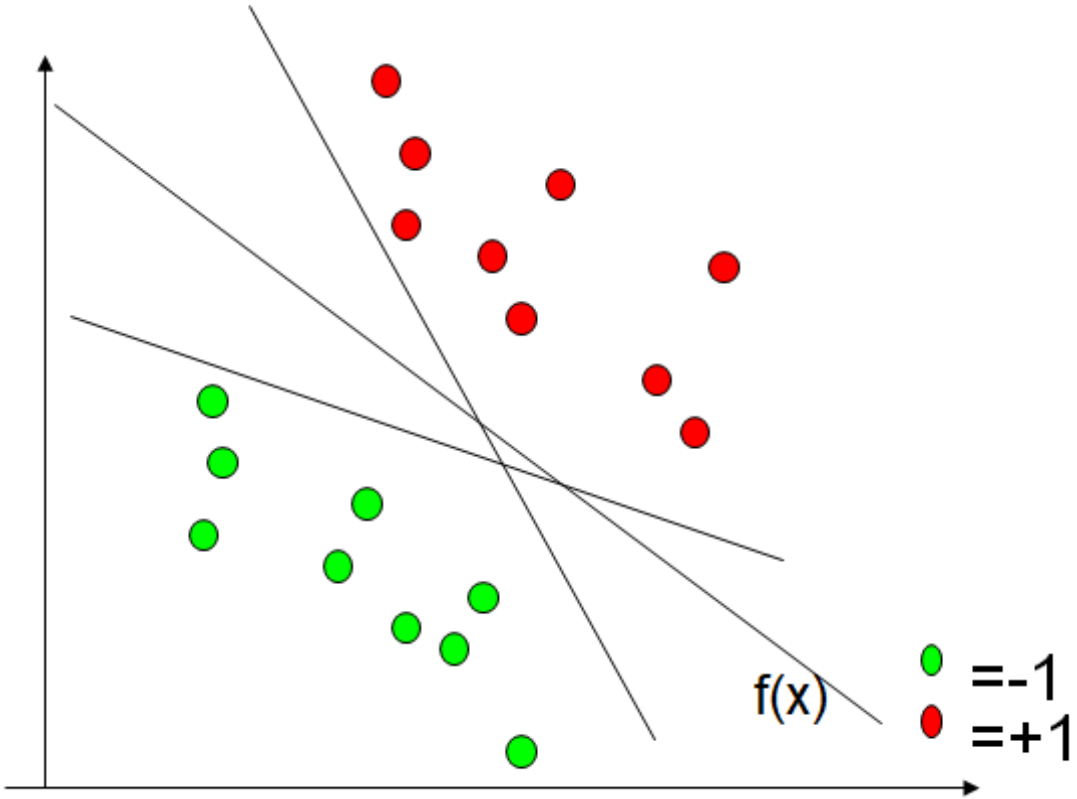
Support Vector Machine:

In machine learning, the support vector machines (SVM) are gaining popularity due to many attractive features. As very powerful tools for data classification and regression, it has been used in many kinds of fields, such as for text classification, sound deduction, software defect prediction, gene analysis and many others. Recently, it has been used by for faults classify in a water level control system.
The classification problems can be restricted to consideration of the two-class problems without loss of generality. The goal of support vector classification 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 plan which can guarantees of good predictive performance on unseen data.

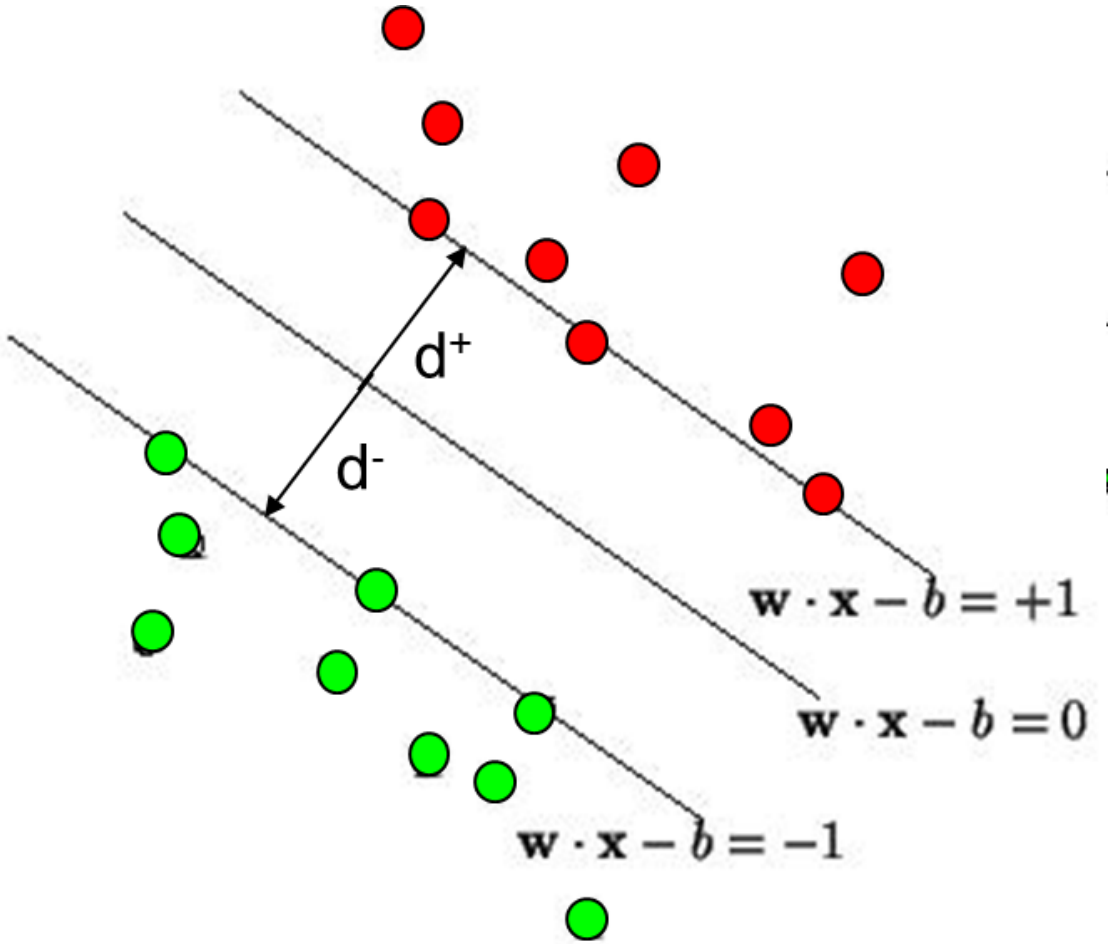
Classification Example



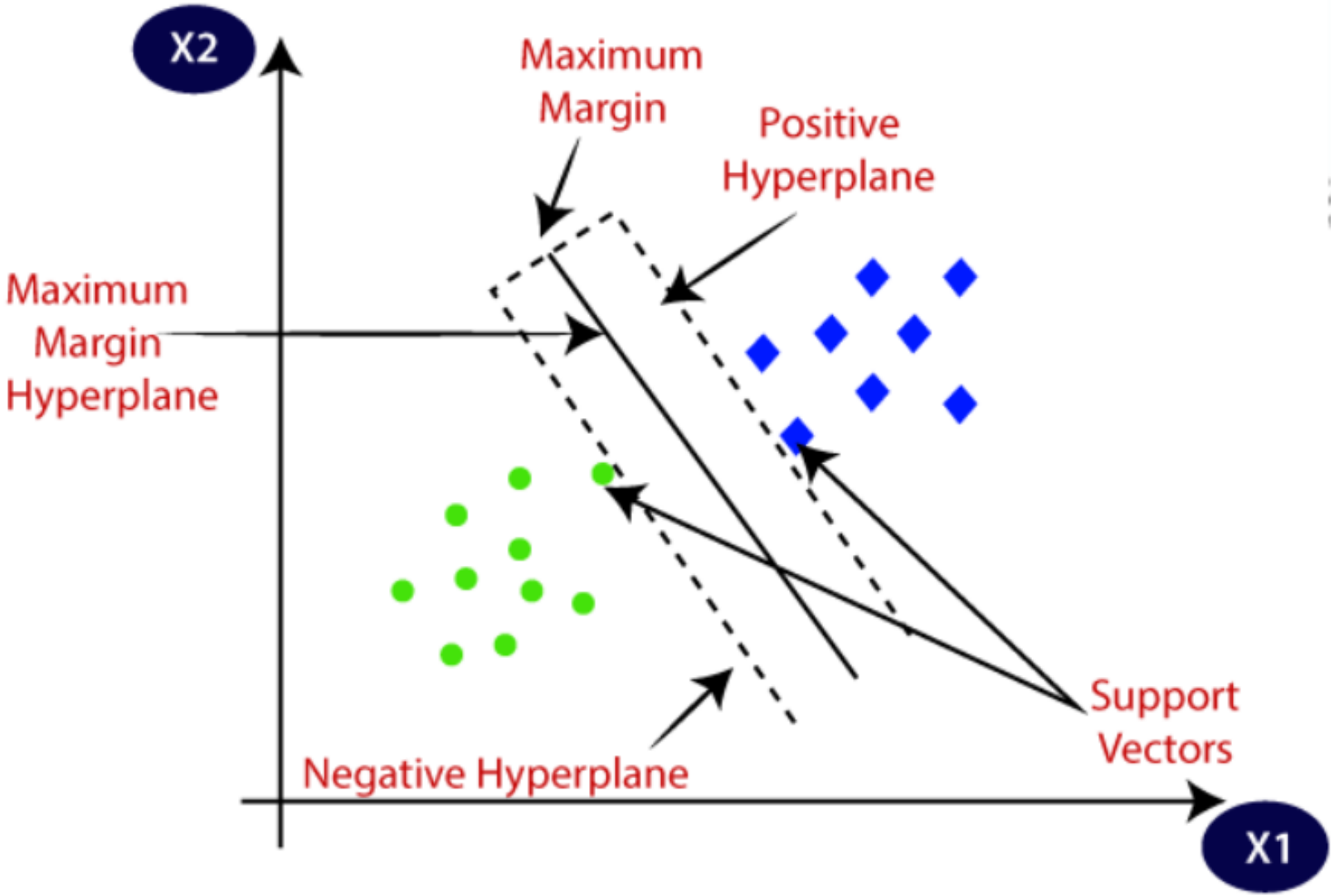
Data Classification



Data Classification



Data Classification



Linear Classification by Support Vector Machine:

Linear classification is frequently performed by using a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ in the following way:
the input $\mathbf{x} = (x_1, ..., x_n)$ is assigned to the positive class if, $f(\mathbf{x}) \geq 0$ and otherwise to the negative class.
We consider the case where $f(\mathbf{x})$ is a linear function of $\mathbf{x}$, so that it can be written as

$$\begin{aligned} f(\mathbf{x}) &= \langle \mathbf{w}, \mathbf{x} \rangle + b \\ &= \sum_{i=1}^{n} w_i x_i + b, \end{aligned}$$

where $\mathbf{w} \in \mathbb{R}^n, b \in \mathbb{R}$ are the parameters that control the functions and the decision rule is given by $sgn(f(\mathbf{x}))$.
The learning methodology implies that these parameters must be learned from the data.

Linear Classification by SVM

The functional margin of $(\mathbf{x}_k, y_k)$ with respect to a hyperplane $(\mathbf{w}, b)$ is defined by

$$\begin{aligned} \gamma_k &= y_k \left(\langle \mathbf{w}, \mathbf{x}_k \rangle + b \right) \\ &= y_k \left(f \left(\mathbf{x}_k \right) \right), \end{aligned}$$

where $y_k \in \{-1, 1\}$. Note that $\gamma_k > 0$ implies correct classification of $(\mathbf{x}_k, y_k)$.

If we replace functional margin by geometric margin, then we obtain the equivalent quantity for the normalized linear function $\left(\frac{1}{\|\mathbf{w}\|}\mathbf{w}, \frac{1}{\|\mathbf{w}\|} b \right)$, which therefore measures the Euclidean distances of the points from the decision boundary in the input space.

Actually geometric margin can be written as

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

Linear Classification by SVM

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

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

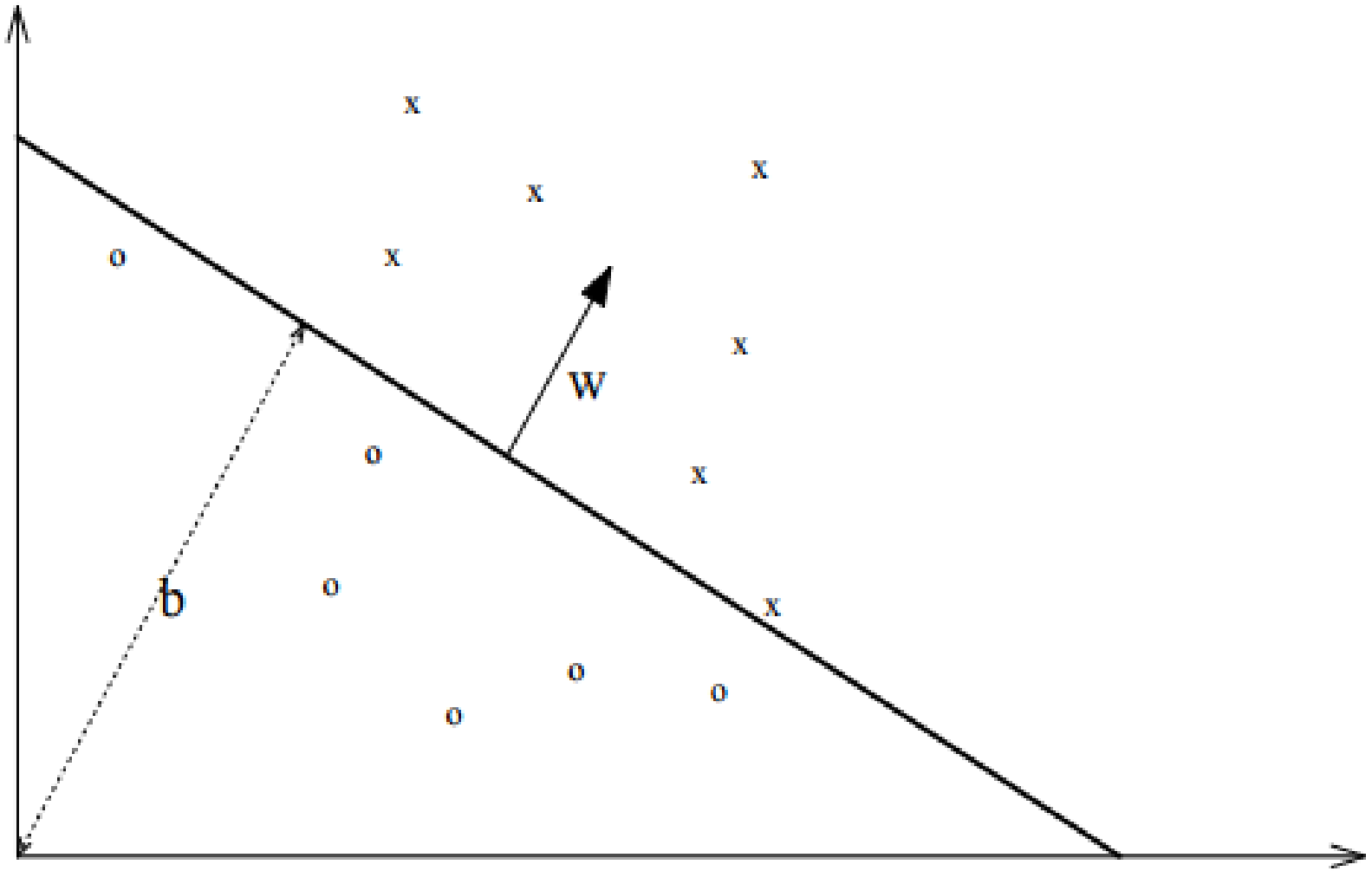
Of course, there have some constraints for the optimization problem. According to the definition of margin, we have $y_k \left(\langle \mathbf{w}, \mathbf{x}_k \rangle + b\right) \geq 1$, $k = 1, ..., m$.
We rewrite the equivalent formation of the objective function with the constraints as

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \text{ such that } y_k \left(\langle \mathbf{w}, \mathbf{x}_k \rangle + b\right) \geq 1, \text{ for } k = 1, ..., m.$$

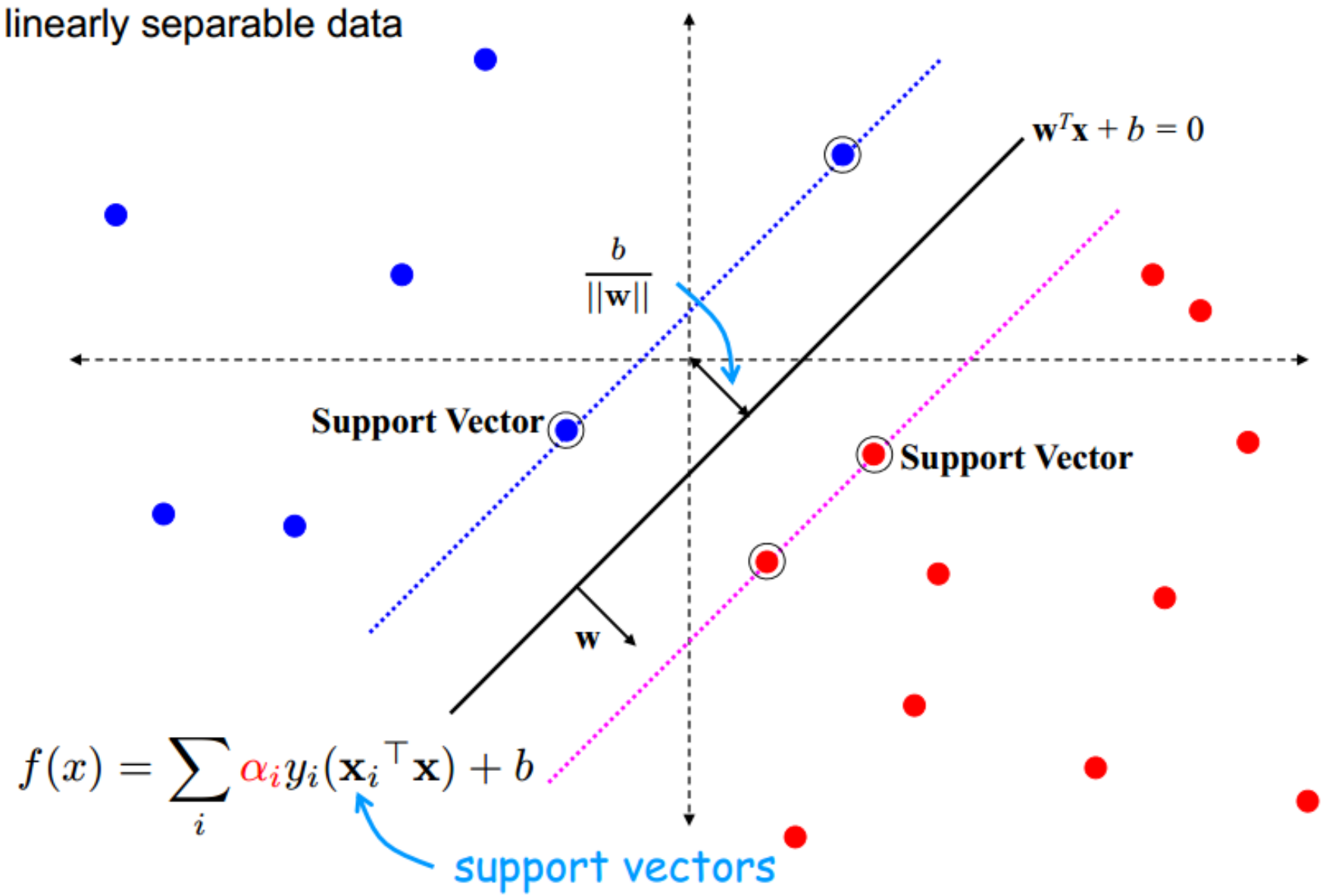
We denote this problem by SVM.

Linear Seperation of Input Space by SVM

$f(\mathbf{x}_k) = \langle \mathbf{w}, \mathbf{x}_k \rangle + b$ and
$\gamma_k = y_k (f(\mathbf{x}_k))$



Support Vector Machine



Support Vector Machine

- Since $\mathbf{wx} + b = 0$ define the same plane, we have the freedom to choose the normalization of $\mathbf{w}$.

Support Vector Machine

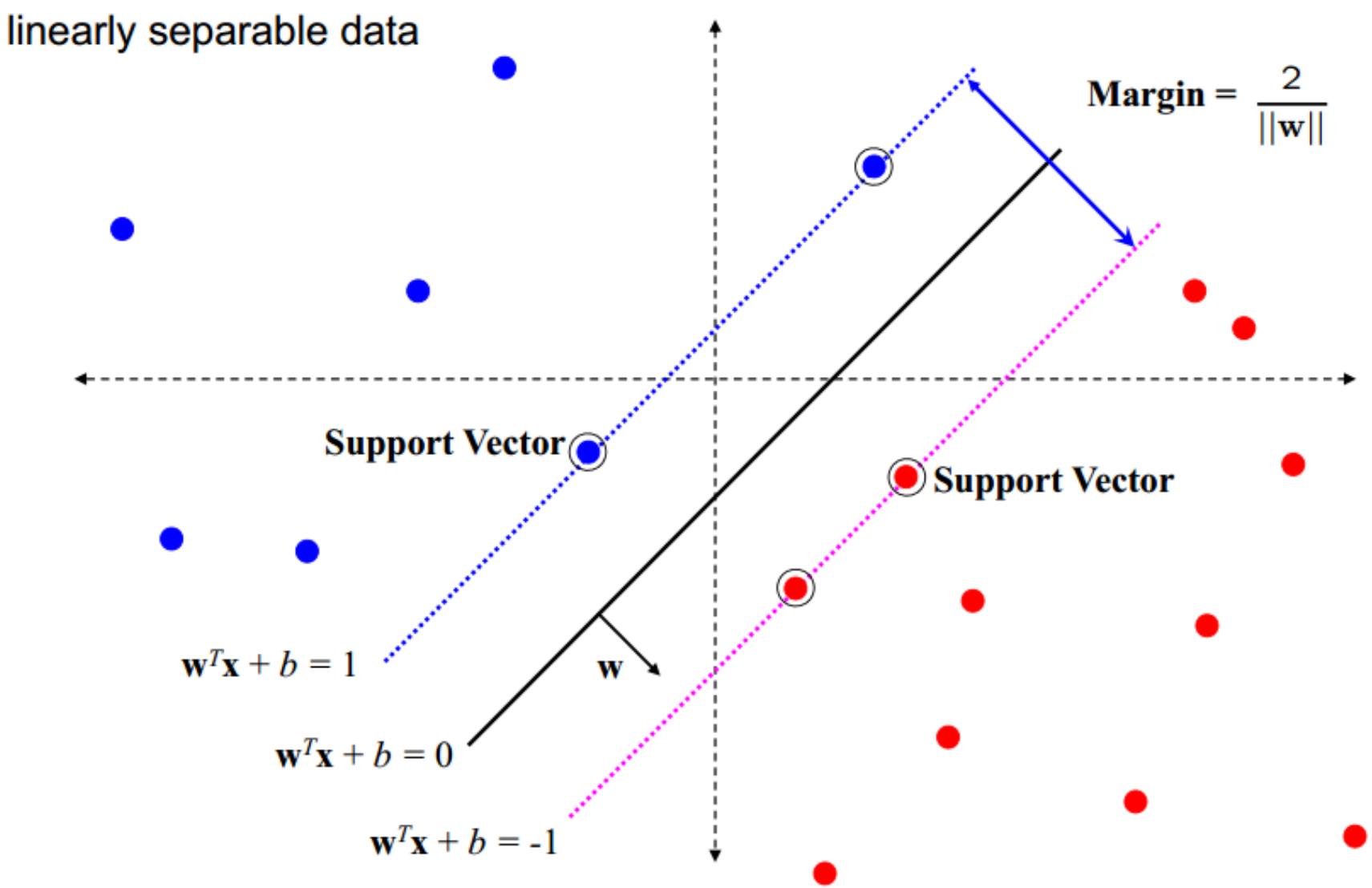
- Since $\mathbf{wx} + b = 0$ define the same plane, we have the freedom to choose the normalization of $\mathbf{w}$.
- Choose normalization such that $\mathbf{wx}_+ + b = 1$ and $\mathbf{wx}_- + b = -1$ for the positive and negative support vectors respectively.

Support Vector Machine

- Since $\mathbf{wx} + b = 0$ define the same plane, we have the freedom to choose the normalization of $\mathbf{w}$.
- Choose normalization such that $\mathbf{wx}_+ + b = 1$ and $\mathbf{wx}_- + b = -1$ for the positive and negative support vectors respectively.
- Then the marginis given by

$$\begin{aligned} \frac{\mathbf{w}}{\|\mathbf{w}\|}(\mathbf{x}_+ - \mathbf{x}_-) &= \frac{1}{\|\mathbf{w}\|}(\mathbf{wx}_+ - \mathbf{wx}_-) \\ &= \frac{1}{\|\mathbf{w}\|}(1 - b - (-1 - b)) \\ &= \frac{2}{\|\mathbf{w}\|}. \end{aligned}$$

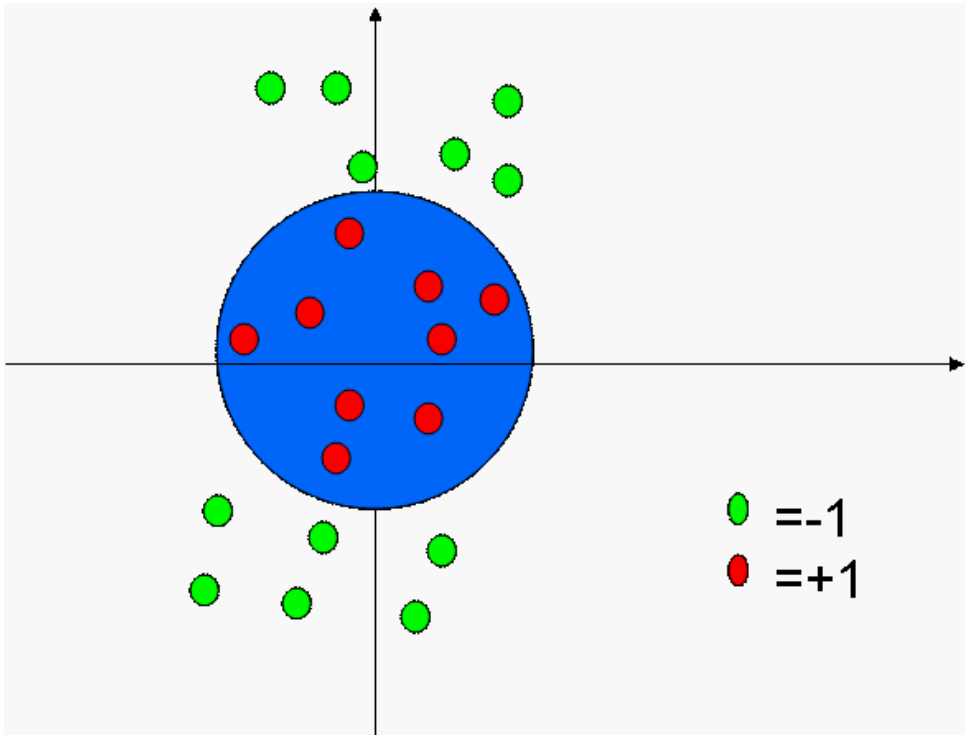
Support Vector Machine



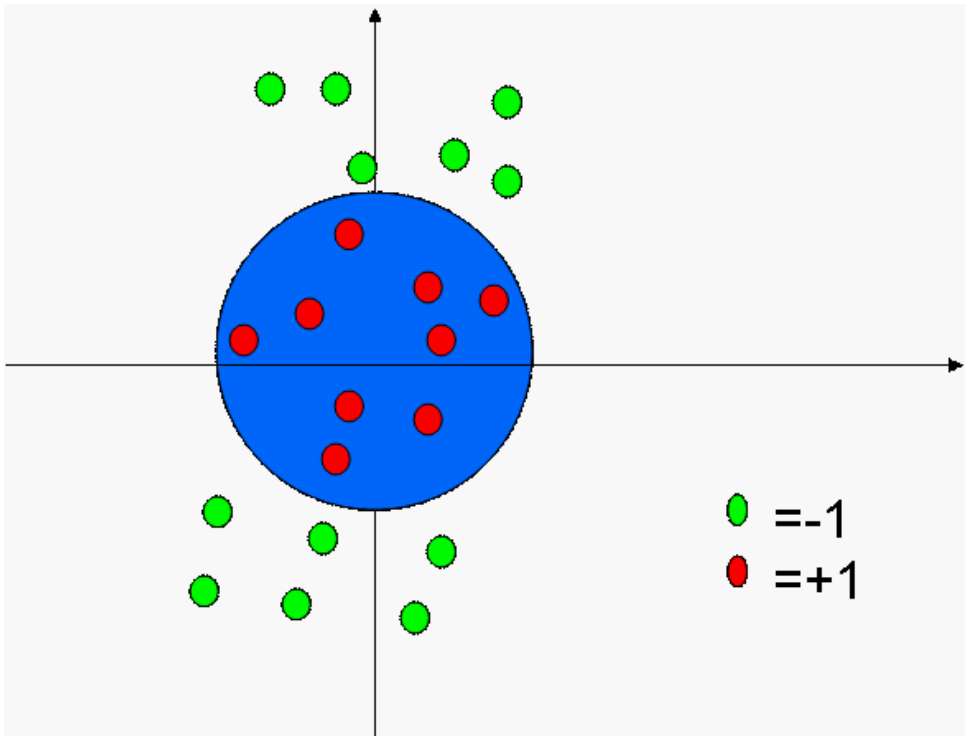
Non Linear Separable by SVM

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 apply a fixed non-linear mapping function Φ that transforms data in input space X to data in feature space F, in which the linear machine can be used.

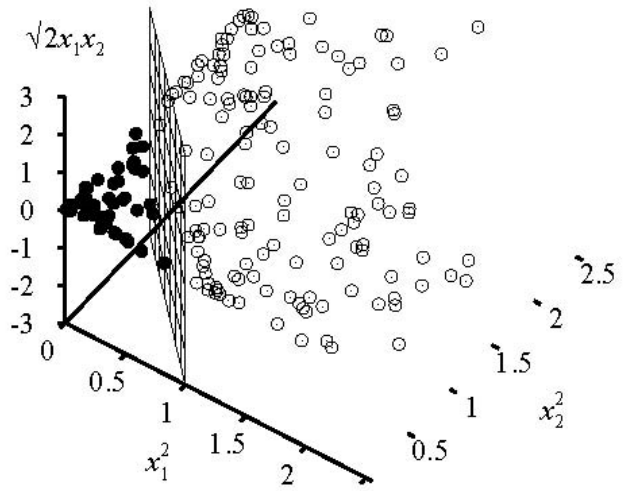
Non Linear Separable by SVM



Non Linear Separable by SVM



$\Rightarrow$



Non Linear Separable by SVM

For this nonlinear 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 Φ as identity operator. For binary classification of data, we consider the decision function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, where the input $\mathbf{x} = (x_1, ..., x_n)$ is assigned to the positive class if, $f(\mathbf{x}) \geq 0$ and otherwise to the negative class. The decision function is defined as

$$f(\mathbf{x}) = < \mathbf{w}, \Phi(\mathbf{x}) > + b. \tag{1}$$

Non Linear Separable by SVM

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

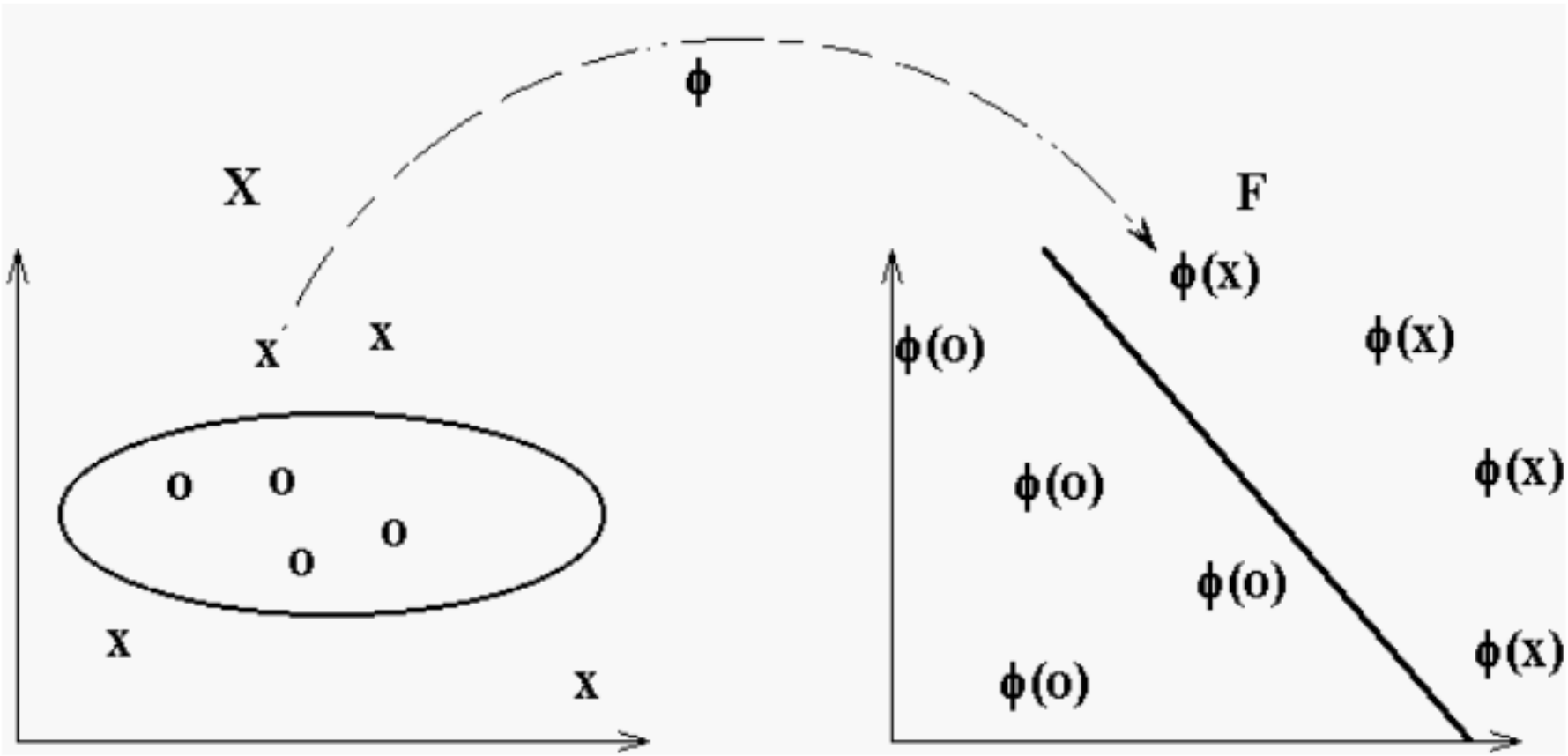
$$\mathbf{w} = \sum_i \alpha_i \Phi(\mathbf{x}_i), \tag{2}$$

where each α_i is Lagrange multipliers 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 σ is a sign function.

Non Linear Separable by SVM



Non Linear Separable by SVM

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

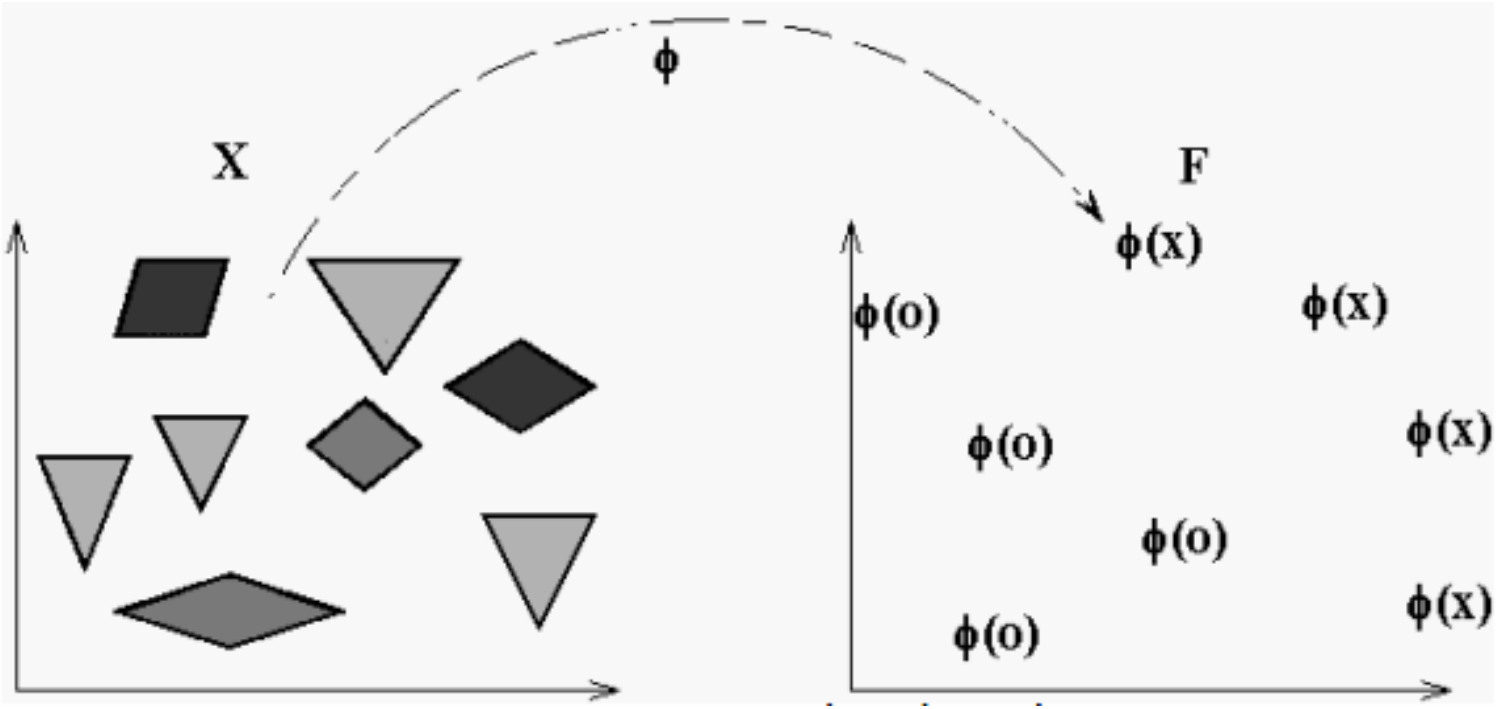
$$K(\mathbf{x}_i, \mathbf{x}) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}). \tag{4}$$

Thus, the decision function becomes

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

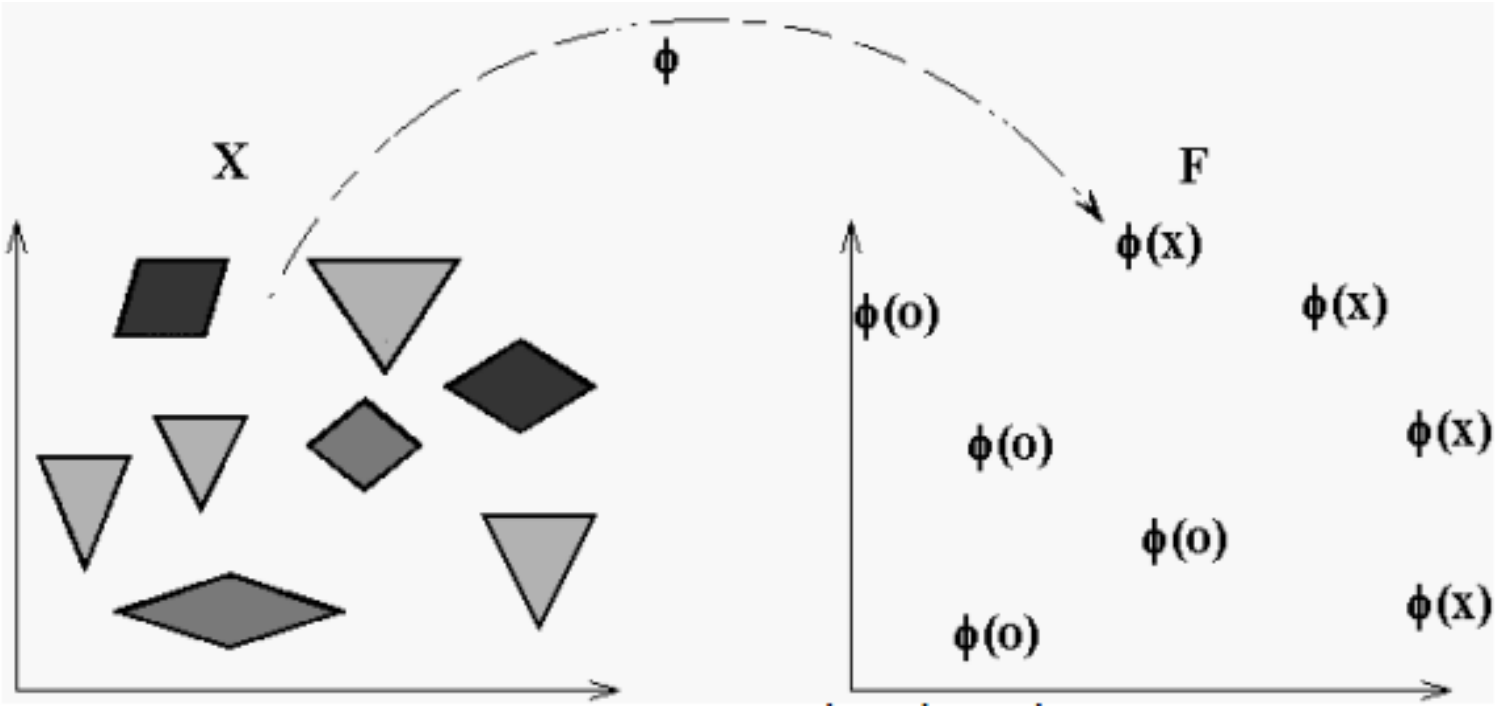
Kernels over General Structure

- Haussler, Watkins, etc: kernels over sets, over sequences, over trees, etc.

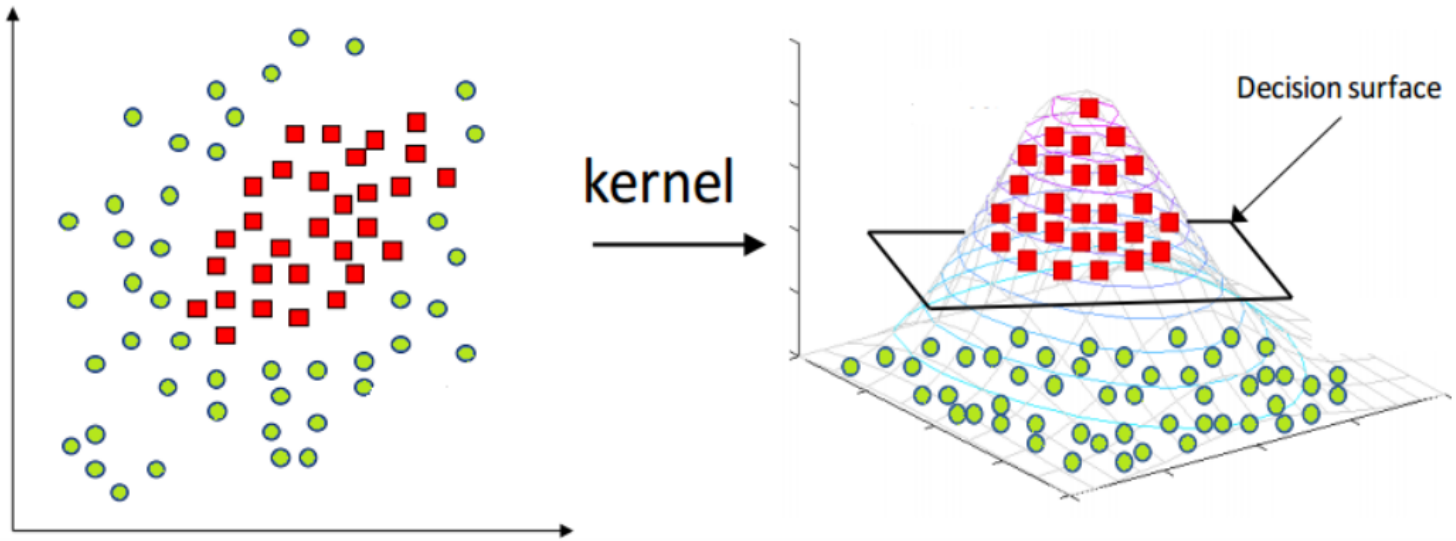


Kernels over General Structure

- Haussler, Watkins, etc: kernels over sets, over sequences, over trees, etc.
- Applied in text categorization, bioinformatics,



Kernels over General Structure



Kernels for SVM

Some useful kernels for real valued vectors 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 γ is a hyperparameter (also called kernel bandwidth). The RBF kernel corresponds to an infinite feature space.

Kernels for SVM (continued)

(IV) Sigmoid Kernel

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

where k is a scalar and θ is the displacement.

(V) Inverse multi-quadratic Kernel

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

where γ is a hyperparameter (also called kernel bandwidth).

Non Linear Separable by SVM

Now, from decision function $f(\mathbf{x}) = <\mathbf{w}, \Phi(\mathbf{x})> + b$, 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

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

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

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

Non Linear Separable by SVM

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

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

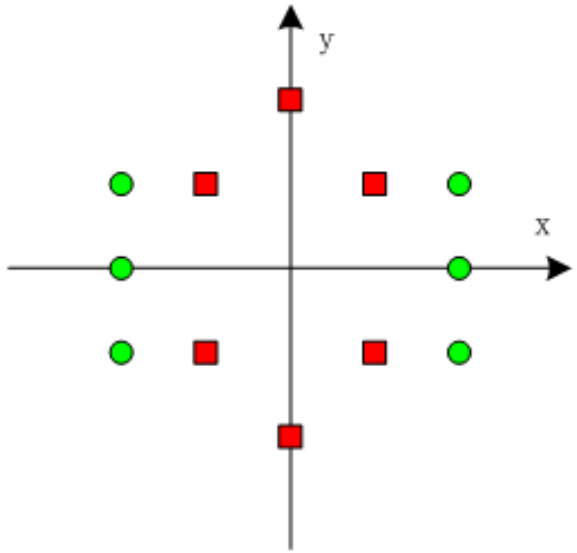
Of course, there have some constraints for the optimization problem. According to the definition of margin, we have $y_i \left(\langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle + b\right) \geq 1$, $i = 1, ..., l$. We rewrite the equivalent formation of the objective function with the constraints as

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

We denote this problem by SVM for data classification.

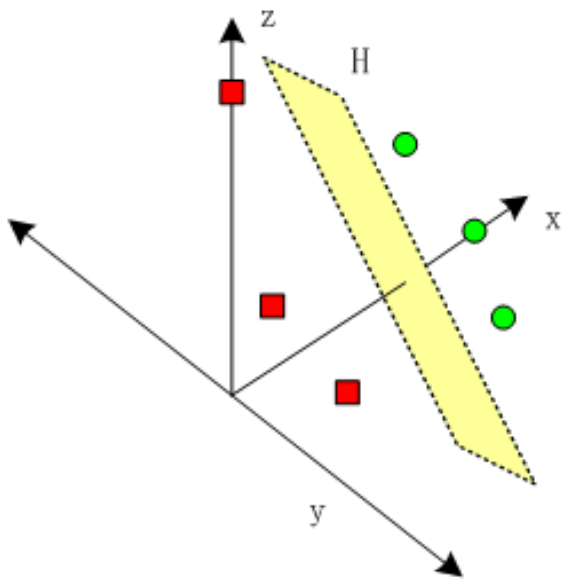
Examples

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.



Example (Continued)

By using the mapping function $\Phi(\mathbf{x}) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right)$, which transform our 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.



Example (Continued)

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

$$(w_1, w_2, w_3)\begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix} + b \geq 1,$$

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

$$(w_1, w_2, w_3)\begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix} + b \geq 1,$$

which implies

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

Example (Continued)

For negative points, we have

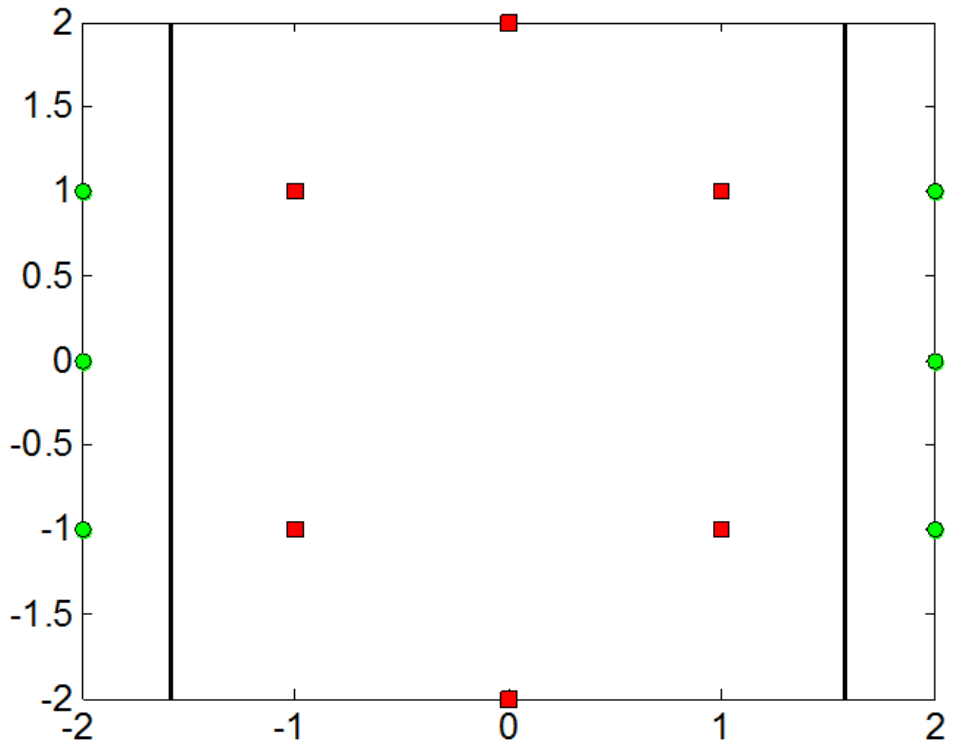
$$
\begin{aligned}
(w_1, w_2, w_3) \begin{bmatrix} 4 \\ 2\sqrt{2} \\ 1 \end{bmatrix} + b &\leq -1, \\
(w_1, w_2, w_3) \begin{bmatrix} 4 \\ -2\sqrt{2} \\ 1 \end{bmatrix} + b &\leq -1, \\
(w_1, w_2, w_3) \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} + b &\leq -1,
\end{aligned}
$$

implies that

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

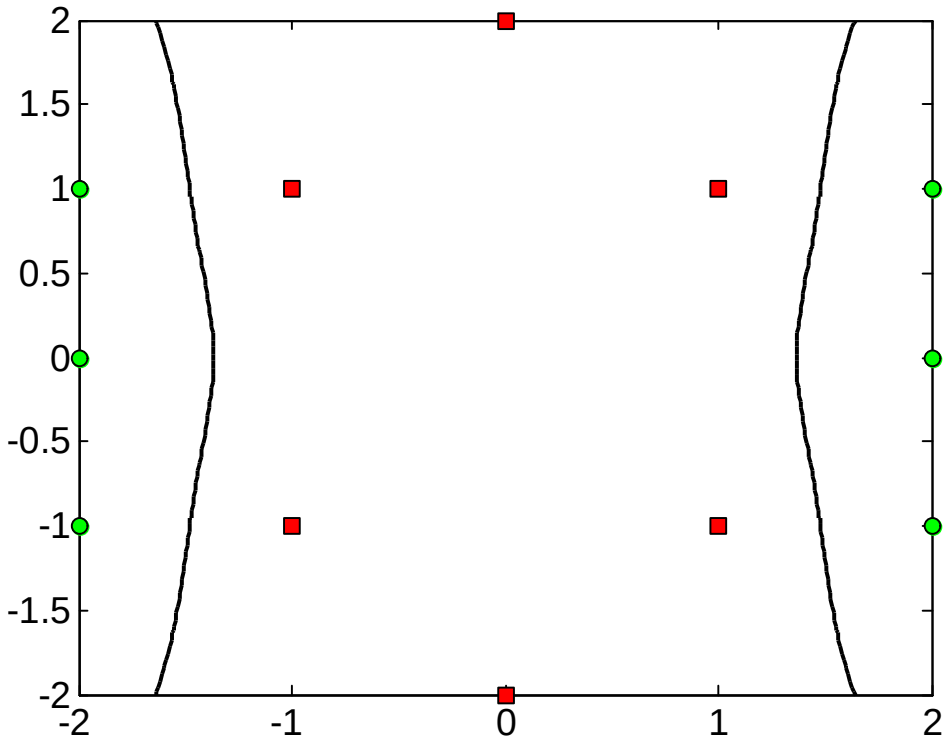
Example (Continued)

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



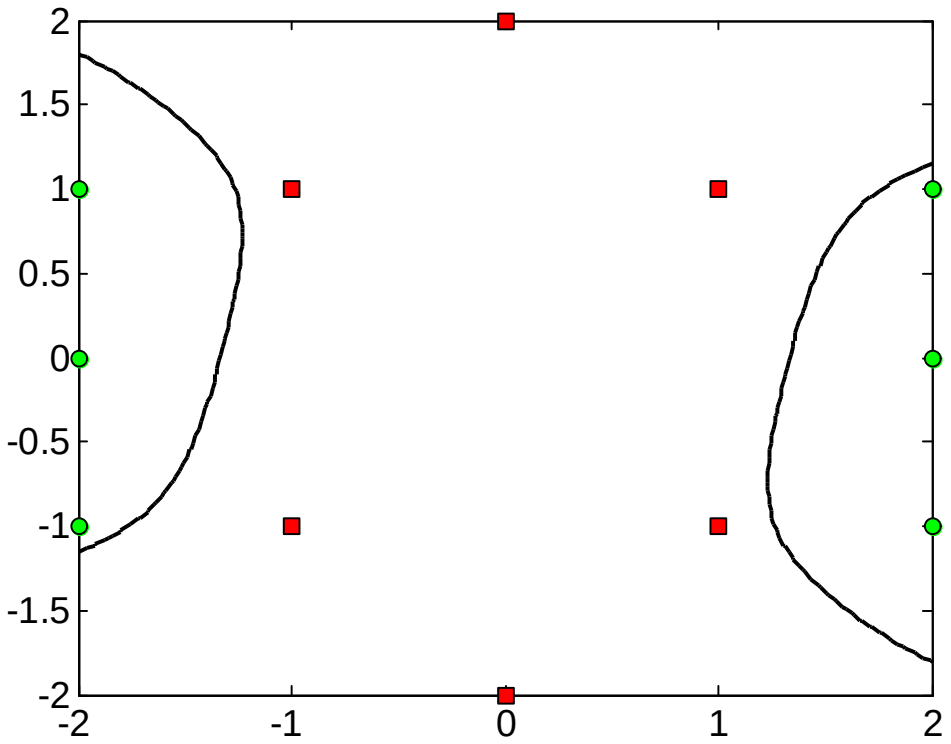
Example (Continued)

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 $w = (0.0031, 0.0012)$.



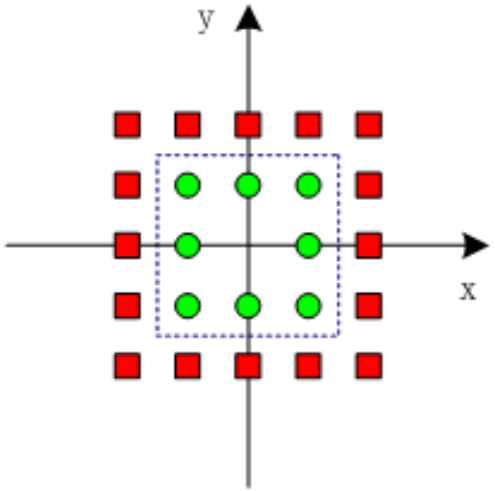
Example (Continued)

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 $w = (0, 0)$.



Examples

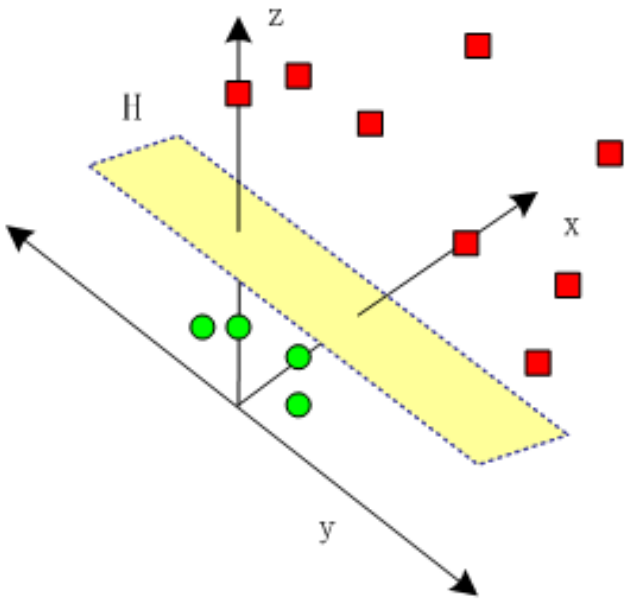
Example 2. Let us look at another example. The positive data be shown as red square and the negative data be shown as blue circle respectively:



It is also a non-linear separable problem.

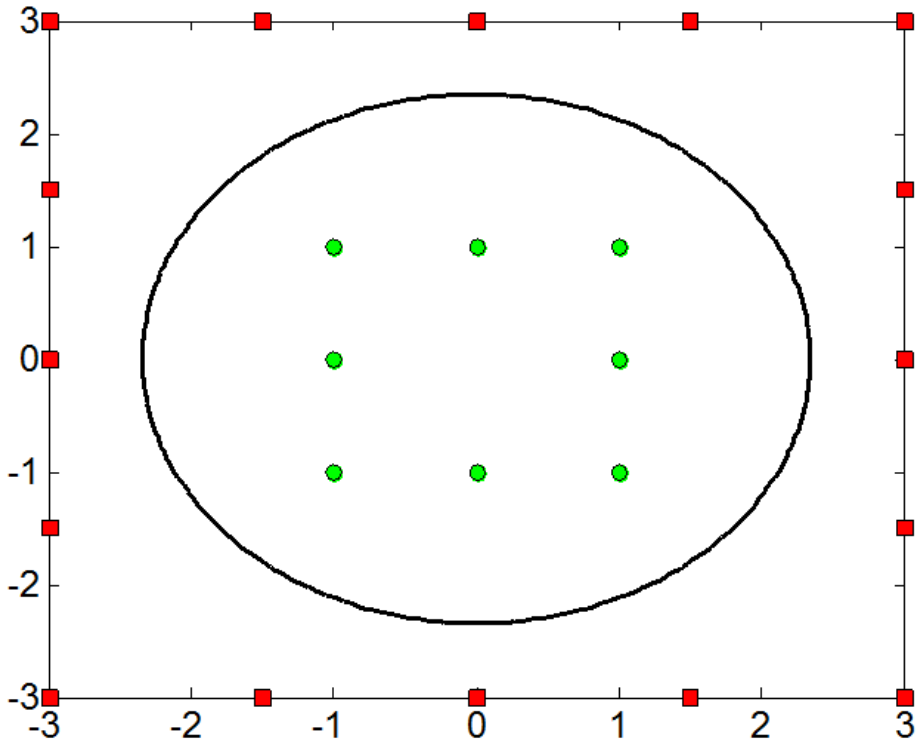
Example (Continued)

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.



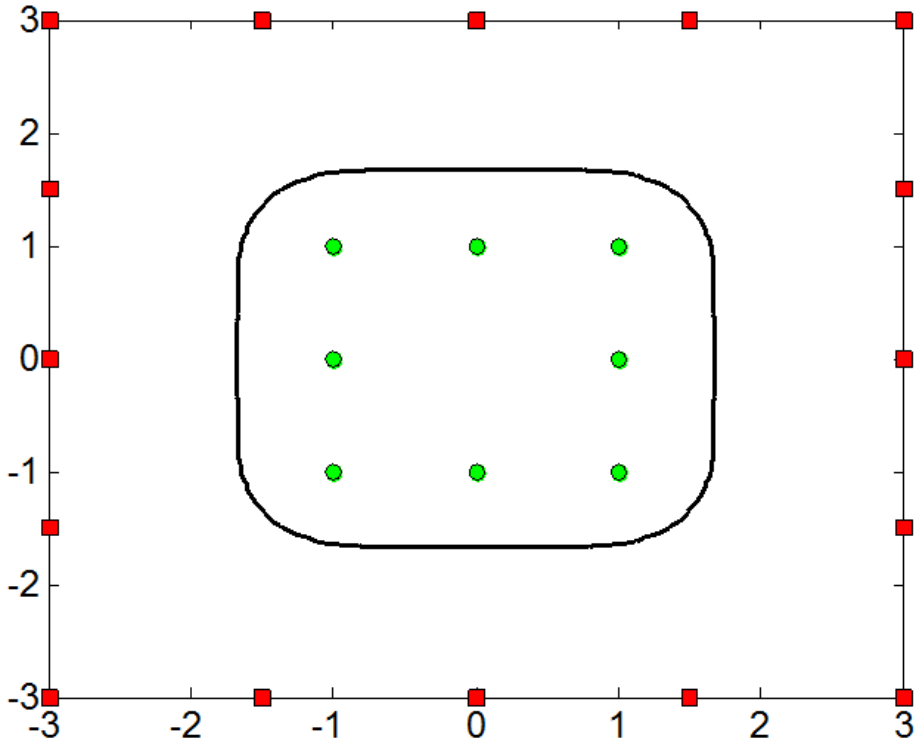
Example (Continued)

Using Polynomial Kernel with $p = 2$, we get $\mathbf{w} = (-0.4898, -0.1633)$.



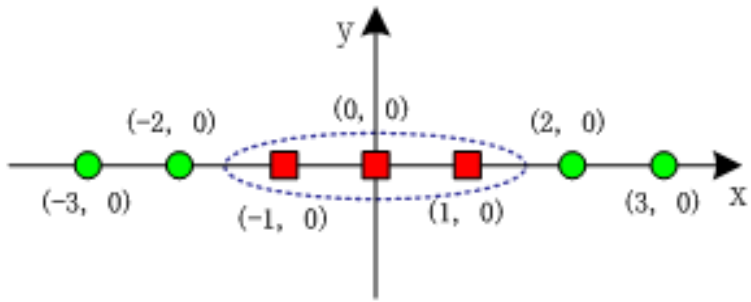
Example (Continued)

Next if we use Radial Basis Function (RBF) Kernel $K(\mathbf{x}_i, \mathbf{x}) = \exp[-\gamma||\mathbf{x}_i - \mathbf{x}||^2]$, with $\gamma = 2$, we get $w = (-0.0016, 0.0014)$ and



Examples

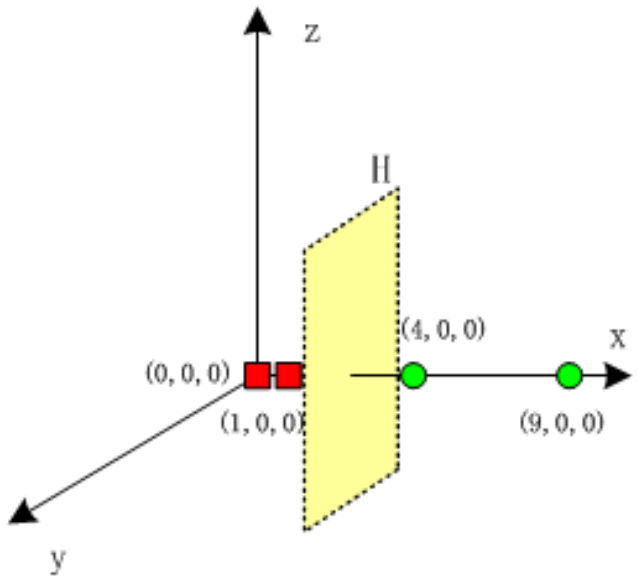
Example 3. Consider the points $(0,0),(1,0),(-1,0)$ as positive class and points $(2,0),(3,0),(-2,0),(-3,0)$ as negative class:



Note that, no linear separator exists for this data in the input space.

Example (Continued)

Now, if we use $\Phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$, then it transform two dimensional data into three dimensional feature space, which can be separated by hyperplane H as shown in the figure:



Generalized Support Vector Machines (GSVM)

We replace $\mathbf{w}, b$ by W, B respectively,
the control function $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ defined as

$$F(\mathbf{x}) = W.\mathbf{x} + B, \tag{1}$$

where $W \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^n$ are the parameters of control functions.
Define

$$\tilde{\gamma}_k^* = \mathbf{y}_k (W\mathbf{x}_k + B) > 1 \qquad for \qquad k = 1, 2, ..., m, \tag{2}$$

where $\mathbf{y}_k \in \{(-1, -1, ..., -1), (1, 1, ..., 1)\}$ is n^{th} dimensional vector.

GSVM

We define a map $G : \mathbb{R}^n \rightarrow \mathbb{R}^n_+$ by

$$G(\mathbf{w}_i) = (\|\mathbf{w}_i\|, \|\mathbf{w}_i\|, ..., \|\mathbf{w}_i\|) \qquad for \qquad i = 1, 2, ..., n, \tag{3}$$

where $\mathbf{w}_i$ be the row of $W_{n \times n}$ for $i = 1, 2, ..., n$.

The problem is to find $\mathbf{w}_i \in \mathbb{R}^n$ that satisfy

$$\min_{\mathbf{w}_i \in W} G(\mathbf{w}_i) \text{ such that } \eta > 0, \tag{4}$$

where $\eta = \mathbf{y}_k (W\mathbf{x}_k + B) - 1$.

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

GSVM

The GSVM is equivalent to

$$\text{find} \quad \mathbf{w}_i \in W: \quad \left\langle G'(\mathbf{w}_i), \mathbf{v} - \mathbf{w}_i \right\rangle \geq 0 \qquad \text{for all} \quad \mathbf{v} \in \mathbb{R}^n \quad \text{with} \quad \eta > 0,$$

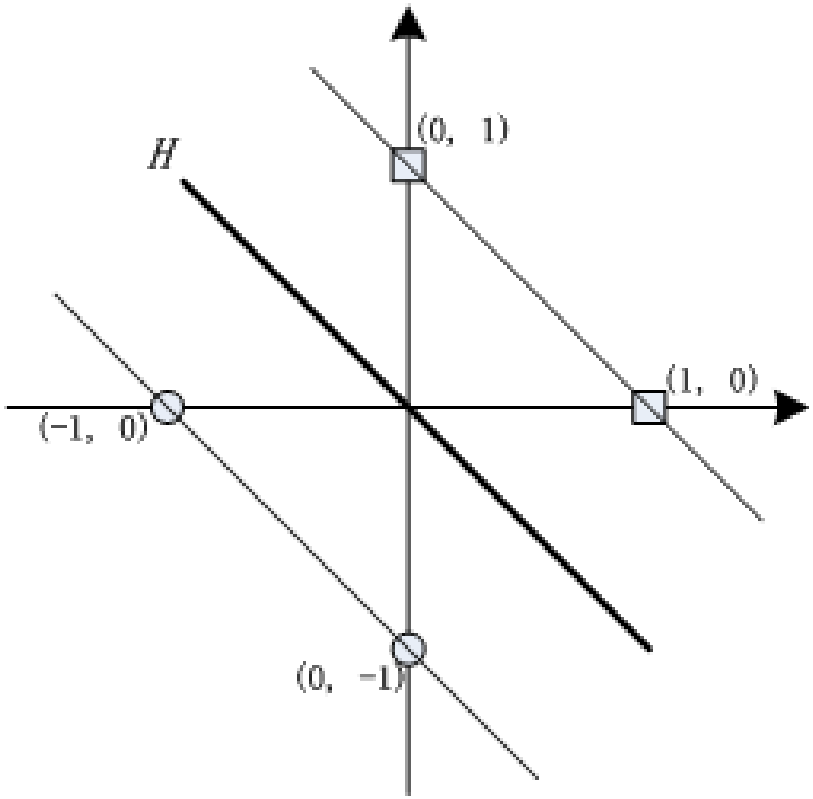
or more specifically

$$\text{find} \quad \mathbf{w}_i \in W: \quad \left\langle \eta G'(\mathbf{w}_i), \mathbf{v} - \mathbf{w}_i \right\rangle \geq 0 \quad \text{for all} \quad \mathbf{v} \in \mathbb{R}^n. \tag{5}$$

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

Example

Let us take the group of points positive class $(1, 0)$, $(0, 1)$ and negative class $(-1, 0)$, $(0, -1)$.
First we use SVM to solve this problem to find the hyperplane $< \mathbf{w}, \mathbf{x} > +b = 0$ that separate this two kinds of points.
The hyperplane is H which is shown in the Figure.



Example (Continued)

For two positive points, we have

$$\begin{aligned}(w_1, w_2)\begin{bmatrix}1\\0\end{bmatrix} + b &= 1\\ (w_1, w_2)\begin{bmatrix}0\\1\end{bmatrix} + b &= 1\end{aligned}$$

For two negative points, we have

$$\begin{aligned}(w_1, w_2)\begin{bmatrix}-1\\0\end{bmatrix} + b &= -1\\ (w_1, w_2)\begin{bmatrix}0\\-1\end{bmatrix} + b &= -1\end{aligned}$$

From these equations, we get $\mathbf{w} = (1, 1)$ and $b = 0$.
The result is $\|\mathbf{w}\| = \sqrt{2}$.

Example continued (for GSVM)

Now we apply GSVM for this data.
For two positive points, 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

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

For two negative points, 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

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

Example continued (for GSVM)

By solving these equations, we get

$$W = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad B = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Thus

$$\min G(\mathbf{w}_i) = \min \{G(\mathbf{w}_1), G(\mathbf{w}_2)\} = (\sqrt{2}, \sqrt{2}).$$

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

GSVM Result

Result. Let $G : \mathbb{R}^n \to \mathbb{R}^n_+$ be a differentiable operator. An element $\mathbf{w}^* \in \mathbb{R}^n$ minimize G if and only if $G'(\mathbf{w}^*) = 0$,
that is,
$\mathbf{w}^* \in \mathbb{R}^n$ solves GSVM if and only if $G'(\mathbf{w}^*) = 0$.

Proof. Let $G'(\mathbf{w}^*) = 0$, then for all $\mathbf{v} \in \mathbb{R}^n$,
$< \eta G'(\mathbf{w}^*), \mathbf{v} - \mathbf{w}^* > \; = \; < 0, \mathbf{v} - \mathbf{w}^* > \; = \; 0.$
Consequently, the inequality
$< \eta G'(\mathbf{w}^*), \mathbf{v} - \mathbf{w}^* > \; = \; < 0, \mathbf{v} - \mathbf{w}^* > \; \geq 0$ holds for all $\mathbf{v} \in \mathbb{R}^n$.
Hence $\mathbf{w}^* \in \mathbb{R}^n$ solves problem of GSVM.
Conversely, assume that $\mathbf{w}^* \in \mathbb{R}^n$ satisfies
$< \eta G'(\mathbf{w}^*), \mathbf{v} - \mathbf{w}^* > \; \geq 0 \;\; \forall \;\; \mathbf{v} \in \mathbb{R}^n.$
Take $\mathbf{v} = \mathbf{w}^* - G'(\mathbf{w}^*)$ in the above inequality implies that
$< \eta G'(\mathbf{w}^*), -G'(\mathbf{w}^*) > \; \geq \; 0,$
which further implies $-\eta ||G'(\mathbf{w}^*)||^2 \; \geq \; 0.$
Since $\eta > 0$, so we get $G'(\mathbf{w}^*) = 0$. □

Metric Projection

Definition

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 $P_K : \mathbb{R}^n \to K$ is characterized by the properties:

(i) $P_K(\mathbf{x}) = \mathbf{z}$ for $\mathbf{x} \in \mathbb{R}^n$ if and only if $< \mathbf{z}, \mathbf{y} - \mathbf{z} > \geq < \mathbf{x}, \mathbf{y} - \mathbf{z} >$ for all $\mathbf{y} \in \mathbb{R}^n$;

(ii) For every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\|P_K(\mathbf{x}) - P_K(\mathbf{y})\|^2 \leq < \mathbf{x} - \mathbf{y}, P_K(\mathbf{x}) - P_K(\mathbf{y}) >$;

(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.

Result of GSVM

Result. Let $G : \mathbb{R}^n \longrightarrow \mathbb{R}^n_+$ be a differentiable operator. An element $\mathbf{w}^* \in \mathbb{R}^n$ minimize mapping G if and only if $\mathbf{w}^*$ is the fixed point of map

$$P_{\mathbb{R}^n_+}\left(I - \rho G'\right) : \mathbb{R}^n \longrightarrow \mathbb{R}^n_+ \quad \text{for any} \quad \rho > 0.$$

that is,

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

where $P_{\mathbb{R}^n_+}$ is a projection map from $\mathbb{R}^n$ to $\mathbb{R}^n_+$.

Definitions:

A map $G : \mathbb{R}^n \to \mathbb{R}^n$ is said to be
(I) L-Lipschitz if for every $L > 0$,

$$\|G(\mathbf{x}) - G(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\| \quad \textit{for all} \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

(II) monotone if

$$< G(\mathbf{x}) - G(\mathbf{y}), \mathbf{x} - \mathbf{y} > \ \geq \ 0 \quad \textit{for all} \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

(III) strictly monotone if

$$< G(\mathbf{x}) - G(\mathbf{y}), \mathbf{x} - \mathbf{y} > \ > \ 0 \quad \textit{for all} \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \quad \textit{with} \quad \mathbf{x} \neq \mathbf{y}.$$

(IV) α-strongly monotone if

$$< G(\mathbf{x}) - G(\mathbf{y}), \mathbf{x} - \mathbf{y} > \ \geq \ \alpha\|\mathbf{x} - \mathbf{y}\|^2 \quad \textit{for all} \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Result of GSVM

Theorem

Let $K \subseteq \mathbb{R}^n$ be closed and convex and $G' : \mathbb{R}^n \to K$ is strictly monotone. If there exists a $\mathbf{w}^* \in K$ which is the solution of GSVM, then $\mathbf{w}^*$ is unique in K.

Theorem

Let $K \subseteq \mathbb{R}^n$ be closed and convex. If the map $G' : \mathbb{R}^n \to K$ is L-Lipchitz and α-strongly monotone then there exists a unique $\mathbf{w}^* \in K$ which is the solution of GSVM.

Example 2

Example

Let us take the group of data of positive class $(\alpha_1, \alpha_2, ..., \alpha_{n-1}, 0)$, $(\alpha_1, \alpha_2, ..., \alpha_{n-2}, 0, \alpha_n)$, ..., $(0, \alpha_2, \alpha_3, ..., \alpha_n)$ and negative class $(k\alpha_1, k\alpha_2, ..., k\alpha_{n-1}, 0)$, $(k\alpha_1, k\alpha_2, ..., k\alpha_{n-2}, 0, k\alpha_n)$, ..., $(0, k\alpha_2, k\alpha_3, ..., k\alpha_n)$ for $n \geq 2$, where each $\alpha_i \neq 0$ for $i \in \mathbb{N}$ and $k \neq 1$. A map $G : \mathbb{R}^n \to \mathbb{R}^n_+$ be given as

$$G(\mathbf{w}_i) = (\|\mathbf{w}_i\|, \|\mathbf{w}_i\|, ..., \|\mathbf{w}_i\|) \qquad for \quad i = 1, 2, ..., n,$$

where $\mathbf{w}_i$ are the row of $W_{n\times n}$ for $i = 1, 2, ..., n$. Then we have

$$G'(\mathbf{w}_i) = \frac{1}{\|\mathbf{w}_i\|}\mathbf{w}_i \qquad for \quad i = 1, 2, ..., n.$$

Example 2 (continued)

Example

Now from the given data, we get

$$W = \frac{2}{(n-1)(1-k)} \begin{bmatrix} \frac{1}{\alpha_1} & \frac{1}{\alpha_2} & \dots & \frac{1}{\alpha_n} \\ \frac{1}{\alpha_1} & \frac{1}{\alpha_2} & \dots & \frac{1}{\alpha_n} \\ . & . & . & . \\ . & . & . & . \\ \frac{1}{\alpha_1} & \frac{1}{\alpha_2} & \dots & \frac{1}{\alpha_n} \end{bmatrix}$$

and so we have

$G(\mathbf{w}_i) = \frac{2}{(n-1)(1-k)} \sqrt{\frac{1}{\alpha_1^2} + \frac{1}{\alpha_2^2} + \ldots + \frac{1}{\alpha_n^2}} (1, 1, \ldots, 1)$ *for* $i = 1, \ldots, n$ and

$$G'(\mathbf{w}_i) = \frac{1}{\sqrt{\frac{1}{\alpha_1^2} + \frac{1}{\alpha_2^2} + \ldots + \frac{1}{\alpha_n^2}}} \left(\frac{1}{\alpha_1}, \frac{1}{\alpha_2}, \ldots, \frac{1}{\alpha_n}\right) \quad \textit{for} \quad i = 1, 2, \ldots, n.$$

Example 2 (continued)

Example

Note that, for any $\mathbf{w}_1, \mathbf{w}_2 \in W$,

$$\left\| G'(\mathbf{w}_1) - G'(\mathbf{w}_2) \right\| \leq L \left\| \mathbf{w}_1 - \mathbf{w}_2 \right\|$$

is satisfied where L is any nonnegative real number. Also

$$< G'(\mathbf{w}_1) - G'(\mathbf{w}_2), \mathbf{w}_1 - \mathbf{w}_2 > \geq 0$$

which show that G' is monotone operator.
Moreover,

$$\mathbf{w} = \frac{2}{(n-1)(1-k)}\left(\frac{1}{\alpha_1}, \frac{1}{\alpha_2}, ..., \frac{1}{\alpha_n}\right)$$

is the solution of GSVM with $\|\mathbf{w}\| = \frac{2}{(n-1)(1-k)} \sqrt{\frac{1}{\alpha_1^2} + \frac{1}{\alpha_2^2} + ... + \frac{1}{\alpha_n^2}}$.

Example 3

Example

Let us take the group of data of positive class $(\alpha_1, \alpha_2, ..., \alpha_m, 0, 0, ..., 0)$, $(0, \alpha_2, \alpha_3, ..., \alpha_{m+1}, 0, 0, ..., 0)$, ..., $(\alpha_1, \alpha_2, ..., \alpha_{m-1}, 0, 0, ..., 0, \alpha_n)$ and negative class $(\kappa\alpha_1, \kappa\alpha_2, ..., \kappa\alpha_m, 0, 0, ..., 0)$, $(0, \kappa\alpha_2, \kappa\alpha_3, ..., \kappa\alpha_{m+1}, 0, 0, ..., 0)$, ..., $(\kappa\alpha_1, \kappa\alpha_2, ..., \kappa\alpha_{m-1}, 0, 0, ..., 0, \kappa\alpha_n)$ for $n > m \geq 1$, where each $\alpha_i \neq 0$ for $i \in \mathbb{N}$ and $\kappa \neq 1$.
A map $G : \mathbb{R}^n \to \mathbb{R}^n_+$ be given as

$$G(\mathbf{w}_i) = (\|\mathbf{w}_i\|, \|\mathbf{w}_i\|, ..., \|\mathbf{w}_i\|) \qquad for \quad i = 1, 2, ..., n,$$

where $\mathbf{w}_i$ are the row of $W_{n\times n}$ for $i = 1, 2, ..., n$. Then we have

$$G'(\mathbf{w}_i) = \frac{1}{\|\mathbf{w}_i\|}\mathbf{w}_i \qquad for \quad i = 1, 2, ..., n.$$

Example 3 (continued)

Example

Now from the given data, we get

$$W = \frac{2}{m\left(1-k\right)}\begin{bmatrix} \frac{1}{\alpha_1} & \frac{1}{\alpha_2} & \dots & \frac{1}{\alpha_n} \\ \frac{1}{\alpha_1} & \frac{1}{\alpha_2} & \dots & \frac{1}{\alpha_n} \\ . & . & . & . \\ . & . & . & . \\ \frac{1}{\alpha_1} & \frac{1}{\alpha_2} & \dots & \frac{1}{\alpha_n} \end{bmatrix}$$

and so we have

$G\left(\mathbf{w}_i\right) = \frac{2}{m(1-k)}\sqrt{\frac{1}{\alpha_1^2}+\frac{1}{\alpha_2^2}+...+\frac{1}{\alpha_n^2}}(1,1,...,1) \quad for\ i=1,2,...,n$

and

$G'\left(\mathbf{w}_i\right) = \frac{1}{\sqrt{\frac{1}{\alpha_1^2}+\frac{1}{\alpha_2^2}+...+\frac{1}{\alpha_n^2}}}(\frac{1}{\alpha_1},\frac{1}{\alpha_2},...,\frac{1}{\alpha_n}) \quad for\ i=1,2,...,n.$

Example 3 (continued)

Example

It is easy to verify that G' is monotone and Lipchitz continuous operator. The vector

$$\mathbf{w} = \frac{2}{m\,(1-k)}\left(\frac{1}{\alpha_1}, \frac{1}{\alpha_2}, ..., \frac{1}{\alpha_n}\right)$$

solves GSVM and

$$\|\mathbf{w}\| = \frac{2}{m\,(1-k)}\sqrt{\frac{1}{\alpha_1^2} + \frac{1}{\alpha_2^2} + ... + \frac{1}{\alpha_n^2}}.$$

Example 4

Example

Consider $(\alpha_1, 0, 0)$, $(0, \alpha_2, 0)$, $(0, 0, \alpha_3)$, $(\beta_1, 0, 0)$, $(0, \beta_2, 0)$, $(0, 0, \beta_3)$ as data of positive class and $(k\alpha_1, 0, 0)$, $(0, k\alpha_2, 0)$, $(0, 0, k\alpha_3)$, $(k\beta_1, 0, 0)$, $(0, k\beta_2, 0)$, $(0, 0, k\beta_3)$ as negative class of data, where α_i, β_i and k are positive real numbers with each $\alpha_i \leq \beta_i$ for $i = 1, 2, 3$ and $k \neq 1$.

The map $G : \mathbb{R}^n \to \mathbb{R}^n_+$ is given as

$$G(\mathbf{w}_i) = (\|\mathbf{w}_i\|, \|\mathbf{w}_i\|, ..., \|\mathbf{w}_i\|) \qquad for \qquad i = 1, 2, ..., n,$$

where $\mathbf{w}_i$ are the row of $W_{3\times 3}$ for $i = 1, 2, 3$. Then we have

$$G'(\mathbf{w}_i) = \frac{1}{\|\mathbf{w}_i\|}\mathbf{w}_i \qquad for \qquad i = 1, 2, 3.$$

Example 4 (continued)

Example

Now from the given data, we get

$$W = \frac{2}{(1-k)} \begin{bmatrix} \frac{1}{\alpha_1} & \frac{1}{\alpha_2} & \frac{1}{\alpha_3} \\ \frac{1}{\alpha_1} & \frac{1}{\alpha 2} & \frac{1}{\alpha_3} \\ \frac{1}{\alpha_1} & \frac{1}{\alpha_2} & \frac{1}{\alpha_3} \end{bmatrix}$$

and so we have

$$G(\mathbf{w}_i) = \frac{2}{(1-k)}\left(\sqrt{\frac{1}{\alpha_1^2}+\frac{1}{\alpha_2^2}+\frac{1}{\alpha_3^2}}, \sqrt{\frac{1}{\alpha_1^2}+\frac{1}{\alpha_2^2}+\frac{1}{\alpha_3^2}}, \sqrt{\frac{1}{\alpha_1^2}+\frac{1}{\alpha_2^2}+\frac{1}{\alpha_3^2}}\right)$$

and

$$G'(\mathbf{w}_i) = \frac{1}{\sqrt{\frac{1}{\alpha_1^2}+\frac{1}{\alpha_2^2}+\frac{1}{\alpha_3^2}}}\left(\frac{1}{\alpha_1}, \frac{1}{\alpha_2}, \frac{1}{\alpha_3}\right).$$

Example 4 (continued)

Example

Note that, for any $\mathbf{w}_1, \mathbf{w}_2 \in W$,

$$\left\| G'(\mathbf{w}_1) - G'(\mathbf{w}_2) \right\| \leq L \left\| \mathbf{w}_1 - \mathbf{w}_2 \right\|$$

is satisfied for $L > 0$. Also

$$< G'(\mathbf{w}_1) - G'(\mathbf{w}_2), \mathbf{w}_1 - \mathbf{w}_2 > \geq 0$$

is satisfied which show that G' is monotone operator. Moreover,

$$\mathbf{w} = \frac{2}{(1-k)}\left(\frac{1}{\alpha_1}, \frac{1}{\alpha_2}, \frac{1}{\alpha 3}\right)$$

is the solution of GSVM with $\|\mathbf{w}\| = \frac{2}{(1-k)}\sqrt{\frac{1}{\alpha_1^2} + \frac{1}{\alpha_2^2} + \frac{1}{\alpha_3^2}}$.

Example 5

Example

Let us take the group of data of positive class $(1,0,0)$, $(1,1,0)$, $(0,1,1)$ and negative class $(-\frac{1}{2},0,0)$, $(-\frac{1}{2},-\frac{1}{2},0)$, $(0,-\frac{1}{2},-\frac{1}{2})$.
Now from the given data, we have

$$W = \begin{bmatrix} \frac{4}{3} & 0 & \frac{4}{3} \\ \frac{4}{3} & 0 & \frac{4}{3} \\ \frac{4}{3} & 0 & \frac{4}{3} \end{bmatrix}$$

with

$$G(\mathbf{w}_i) = \frac{4}{3}(\sqrt{2},\sqrt{2},\sqrt{2}) \quad for \quad i = 1,2,3 \text{ and}$$

$$G'(\mathbf{w}_i) = \frac{1}{\sqrt{2}}(1,0,1) \quad for \quad i = 1,2,3.$$

It is easy to verify that G' is monotone operator and Lipchitz continuous. Moreover, $\mathbf{w} = (\frac{4}{3},0,\frac{4}{3})$ is the solution of GSVM with $\|\mathbf{w}\| = \frac{4}{3}\sqrt{2}$.

Open Problems

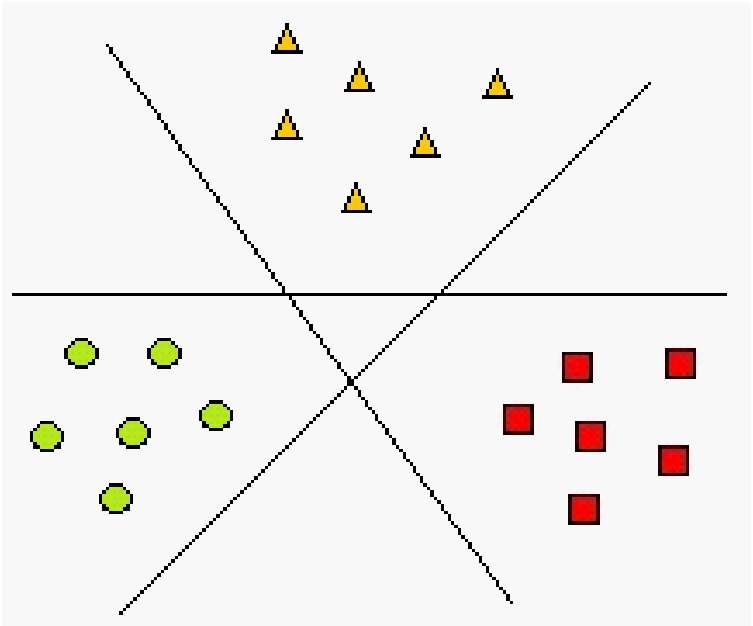
The following questions of interest are still open and will be addressed in this part of our research work.

Problem 1.
To prove new schemes and algorithm for data classification from any general vector space. In this way, we can not only the separate the data of $\mathbb{R}^n$ but also establish the classification of general vector spaces particularly in the framework of Hilbert space and Banach space. With the help of variational inequality problem (VIP), we can also modify the general support vector machine problem to variational inequalities based optimization problem along with certain constraints.

Open Problems

Problem 2.

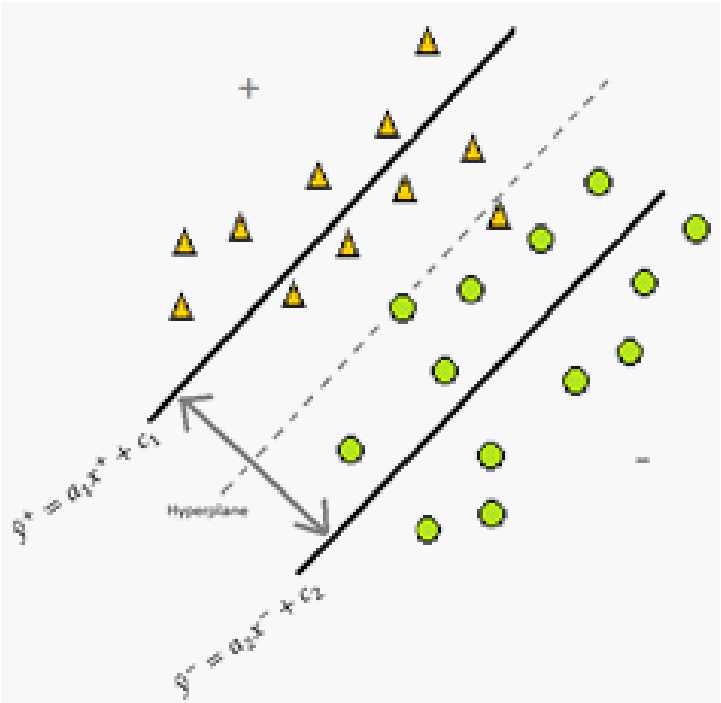
To establish new data classification scheme which is useful to classify the data not only into two categories but also split the data into various parts. In current methods, we can only classify them into two parts such as into positive data and negative data. New scheme can split and arrange the given input data into many parts. See for example the case of data splits into three part in Figure 1.



Open Problems

Problem 3.

Develop the approximate data separation technique. As some times, the data cannot be separable by employing even any Kernel function. Then we can separable this data by using SVM approximation method. In this method, use least square fitting and then split the given data into desired parts. See for example the case of approximate data grouping into two groups by least square fitting method in Figure.



References 1

- T. Nazir, X. Qi and S. Silvestrov, Linear Classification of Data with Support Vector Machines and Generalized Support Vector Machines, Engineering Mathematics II, Springer Proceedings in Mathematics & Statistics, Volume 179 (2017), Pages 355-375.
- T. Nazir, X. Qi and S. Silvestrov, Linear and Nonlinear Classifiers of Data with Support Vector Machines and Generalized Support Vector Machines, Engineering Mathematics II, Springer Proceedings in Mathematics & Statistics, Volume 179 (2017), Pages 377-396.
- X. Qi, S. Silvestrov and T. Nazir, Data classification with support vector machine and generalized support vector machine, AIP Conference Proceedings 1798, 020126 (2017), 10 page.
- D. Wang, Xiaomin Qi, S. Wen, and M. Deng, SVM based fault classifier design for a water level control system, Proceedings of 2013 International Conference on Advanced Mechatronic Systems, 2013/9/25-27.

References 2

- T. Joachims, Text categorization with support vector machines: Learning with many relevant features, In Proceedings of the European Conference on Machine Learning, Springer, 1998.
- W. S, Noble, Support vector machine applications in computational Biology, MIT Press, 2004.
- X.Y. Wang, T. Wang, and J. Bu, Color image segmentation using pixel wise support vector machine classification, Pattern Recognition 44 (4), 777-787 (2011).
- N. Cristianini, and J. Shawe-Taylor, An Introduction to support vector machines and other kernel based learning methods. Cambridge University Press, 2000.
- I. Guyon, J. Weston, S. Barnhill and V. Vapnik, Gene selection for cancer classification using support vector machines. Machine Learning 46 (1-3), 389-422 (2002).

References 3

- N. Khan, R. Ksantini, I. Ahmad and B. Boufama, A novel SVM+NDA model for classification with an application to face recognition. Pattern Recognition 45 (1), 66-79 (2012).
- D. Kinderlehrer and G. Stampacchia, An Introduction to Variational Inequalities and Their Applications, Society for Industrial and Applied Mathematics Publications, Phildelphia, PA, USA, 2000.
- M. A. Noor, Some Development in General Variational Inequalities, Applied Mathematics and computation, 152 (2004), 199-277.
- Can H, Jianchun X, Ruide Z, Juelong L, Qiliang Y, Liqiang X. A new model for software defect prediction using particle swarm optimization and support vector machine. In: Control and decision conference (CCDC), 2013 25th Chinese. IEEE; 2013. p. 4106–10.

Thanks