SUPPORT VECTOR MACHINE

常琳

Content

- Review--SVM
- Multi-class classification

REVIEW

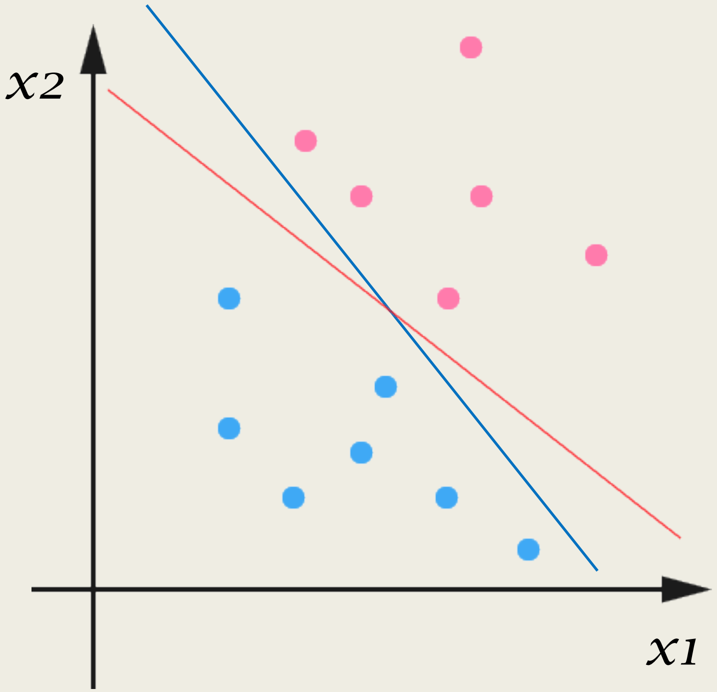
Margin and support vector

Dual problem

Kernel function

Soft margin and regularization

Margin and support vector 间隔与支持向量

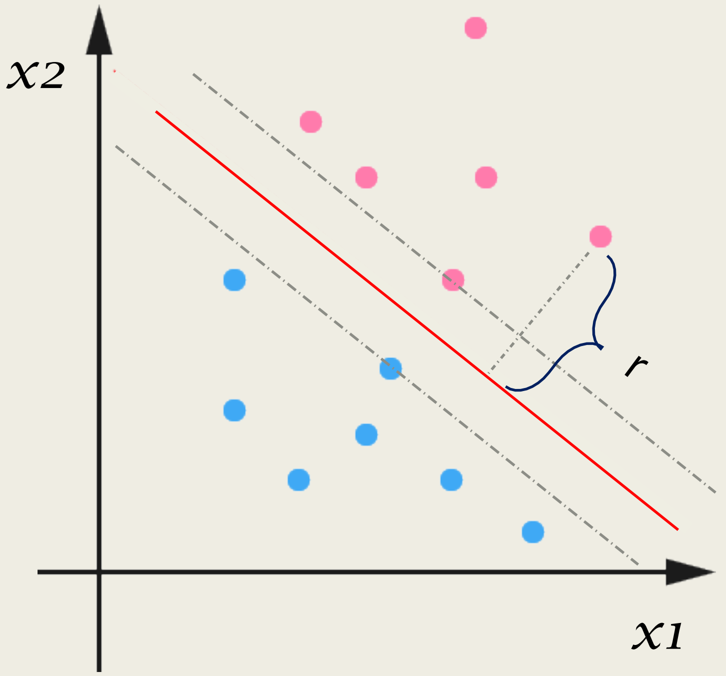


$D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, y_m)\}, y_i \in \{-1, +1\}$

Hyperplane: $\boldsymbol{\omega}^T \boldsymbol{x} + b = 0$

The distance from a sample to the hyperplane:

$$r = \frac{|\boldsymbol{\omega}^T \boldsymbol{x} + b|}{\|\omega\|}$$

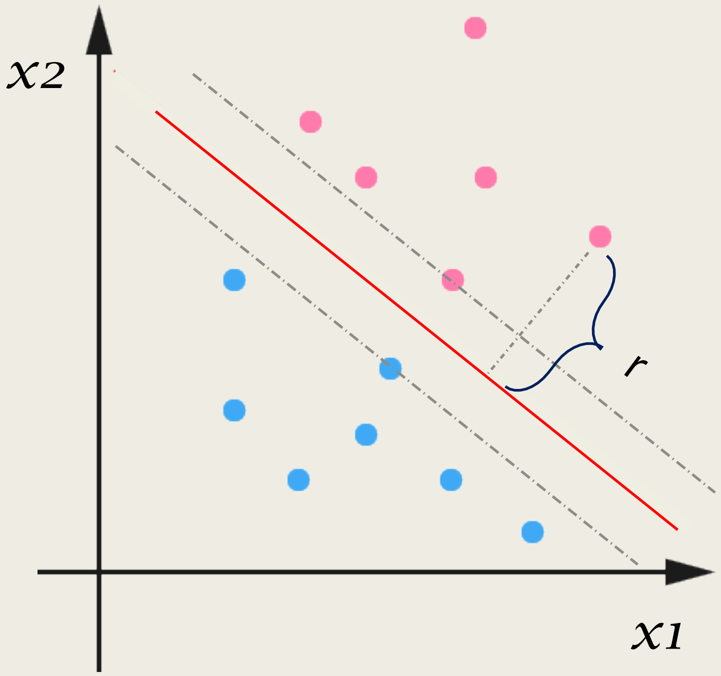


The distance from a sample to the hyperplane:

$$r = \frac{\left|\boldsymbol{\omega}^T \boldsymbol{x} + b\right|}{\|\omega\|}$$

If $y_i = +1, \boldsymbol{\omega}^T \boldsymbol{x}_i + b > 0$

$y_i = -1, \boldsymbol{\omega}^T \boldsymbol{x}_i + b < 0$



The distance from a sample to the hyperplane:

$$r = \frac{|\boldsymbol{\omega}^T \boldsymbol{x} + b|}{\|\omega\|}$$

If

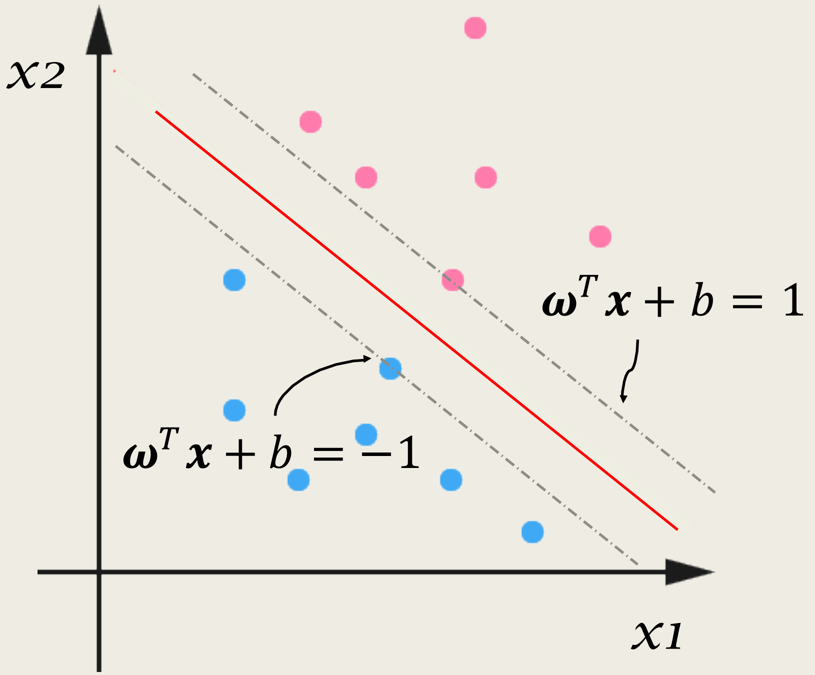
$$y_i = +1, \boldsymbol{\omega}^T \boldsymbol{x}_i + b > 0$$

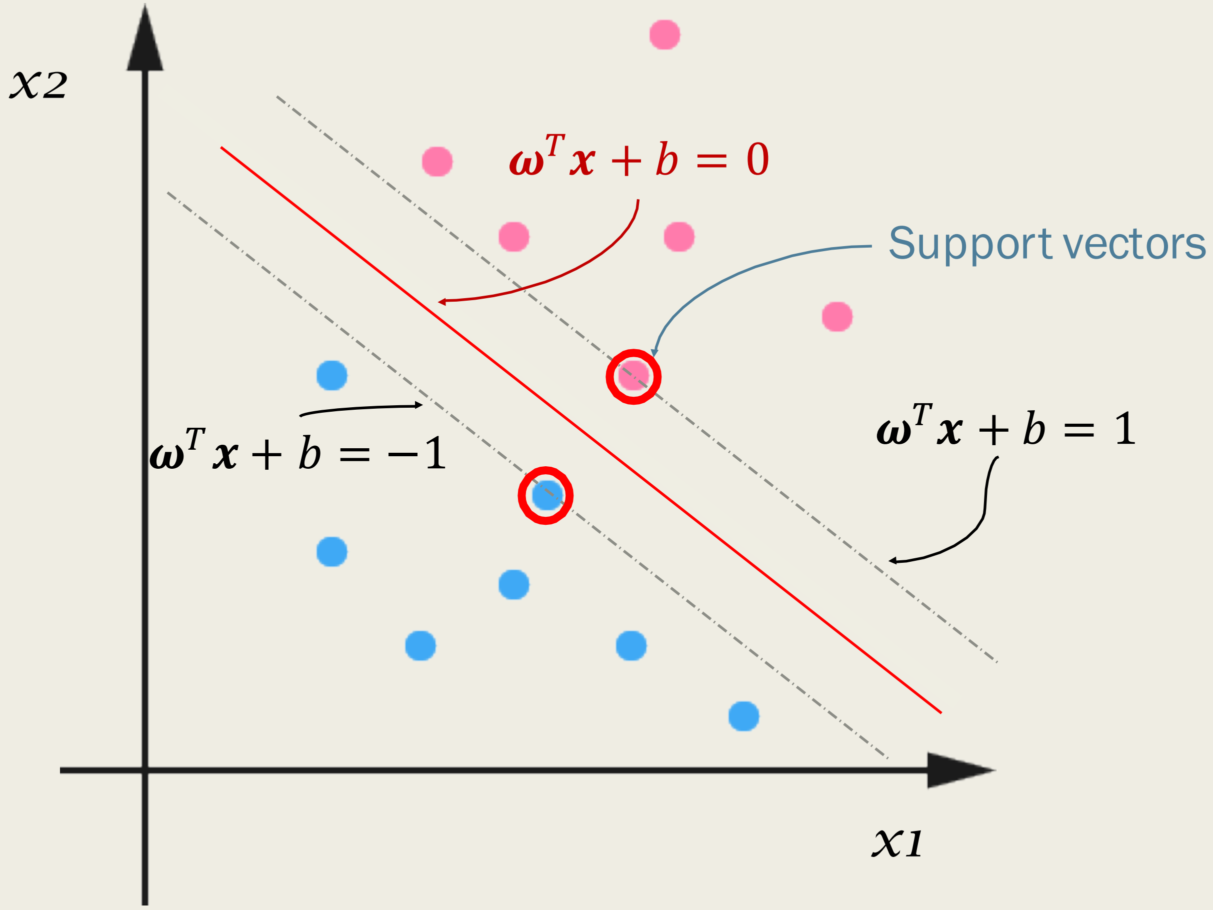
$$y_i = -1, \boldsymbol{\omega}^T \boldsymbol{x}_i + b < 0$$

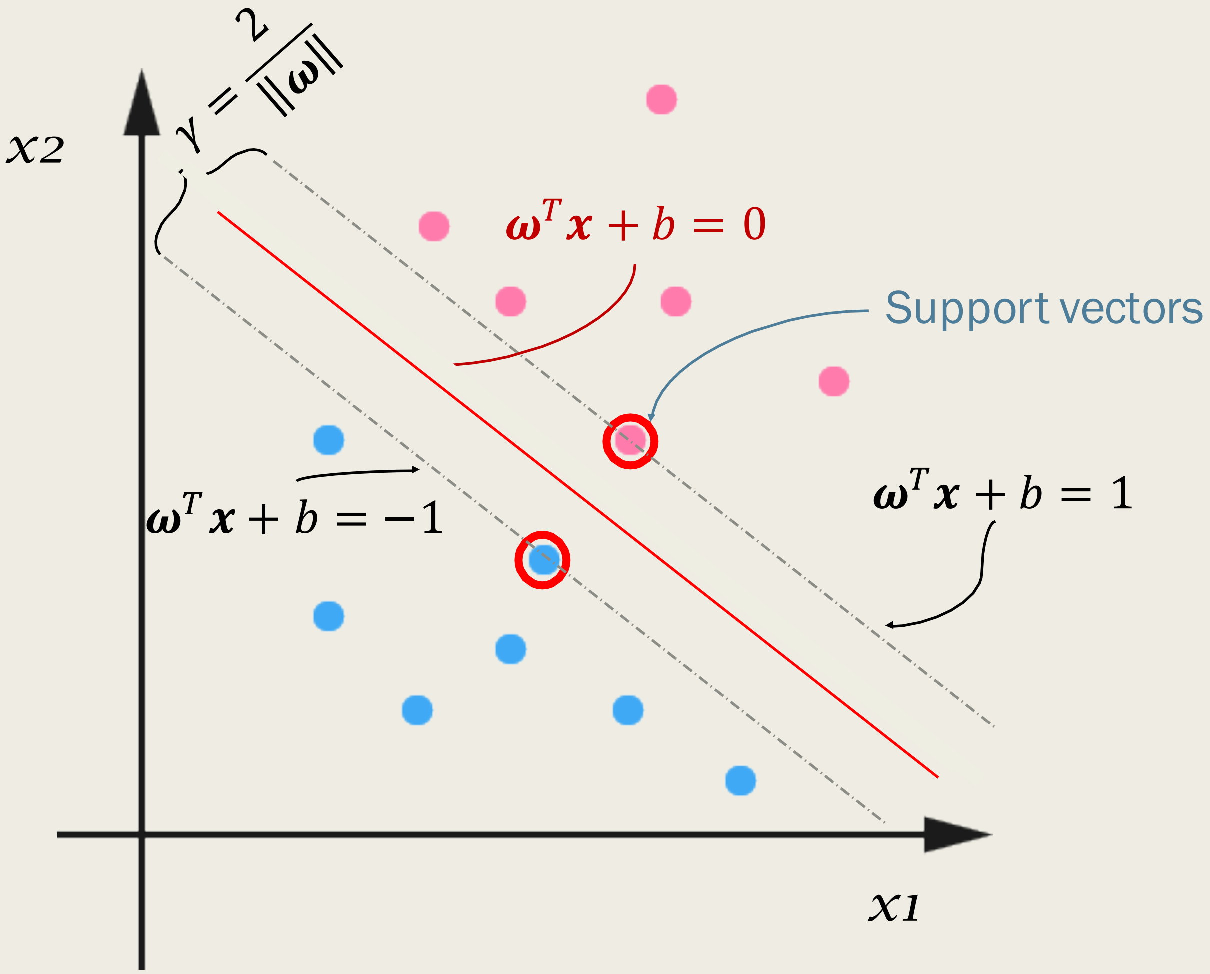
Let

$$\begin{cases} \boldsymbol{\omega}^T \boldsymbol{x}_i + b \geq +1, & y_i = +1 \\ \boldsymbol{\omega}^T \boldsymbol{x}_i + b \leq -1, & y_i = -1 \end{cases}$$

Transformations $\varsigma\boldsymbol{\omega} \longmapsto \boldsymbol{\omega}'$ and $\varsigma b \longmapsto b'$ will always exist so that the formula is established if the hyperplane could classify the samples correctly.



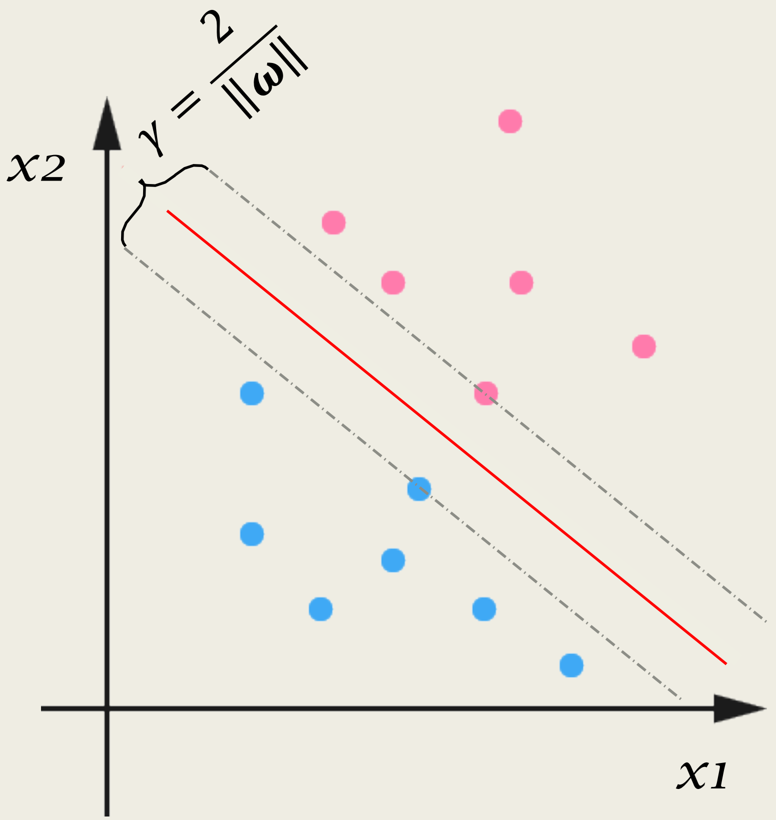

$x2$
$\omega^T x + b = 0$
Support vectors
$\omega^T x + b = 1$
$\omega^T x + b = -1$
$x1$



The distance between these two support vectors is $\gamma = \frac{2}{\|\omega\|}$, called margin

Goal:

Find the maximum-margin hyperplane

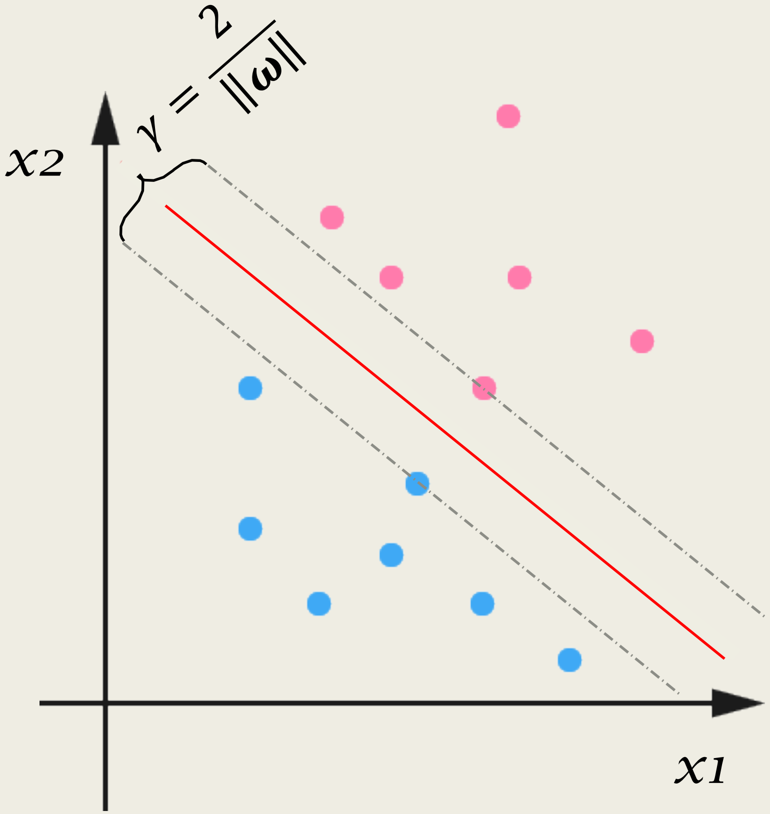


Goal:

Find the maximum-margin hyperplane

$$\max_{\boldsymbol{\omega},b} \frac{2}{\|\boldsymbol{\omega}\|}$$

s.t. $y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1, \qquad i = 1, 2, \dots, m$



Goal:

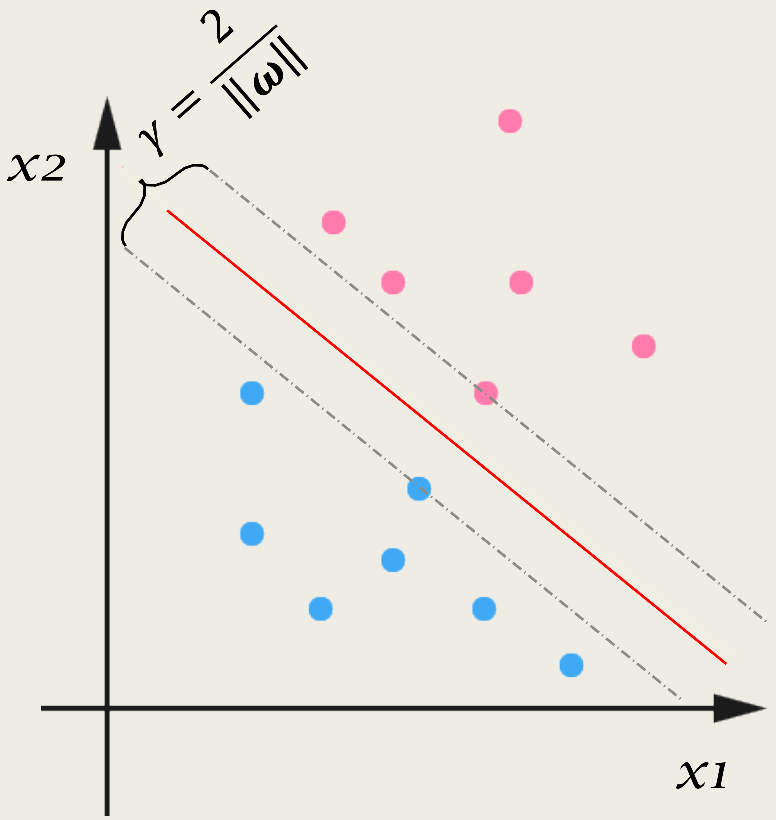
Find the maximum-margin hyperplane

$$\max_{\boldsymbol{\omega},b} \frac{2}{\|\boldsymbol{\omega}\|}$$

s.t. $y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1, \quad i = 1, 2, \dots, m$

$$\min_{\boldsymbol{\omega},b} \frac{1}{2}\|\boldsymbol{\omega}\|^2$$

s.t. $y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1, \quad i = 1, 2, \dots, m$



Goal:

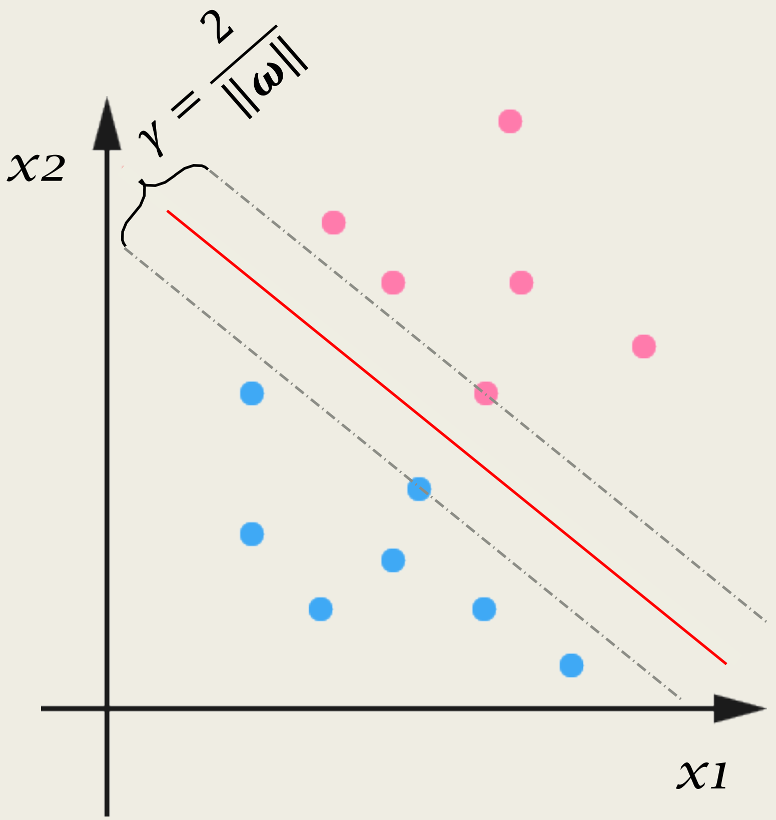
Find the maximum-margin hyperplane

$$\max_{\boldsymbol{\omega},b} \frac{2}{\|\boldsymbol{\omega}\|}$$

$$\text{s.t. } y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1, \qquad i = 1, 2, \dots, m$$

$$\min_{\boldsymbol{\omega},b} \frac{1}{2}\|\boldsymbol{\omega}\|^2 \tag{1-1}$$

$$\text{s.t. } y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1, \qquad i = 1, 2, \dots, m$$



model

$$f(\boldsymbol{x}) = \boldsymbol{\omega}^T \boldsymbol{x} + b$$

Goal:

Find the maximum-margin hyperplane

$$\max_{\boldsymbol{\omega},b} \frac{2}{\|\boldsymbol{\omega}\|}$$

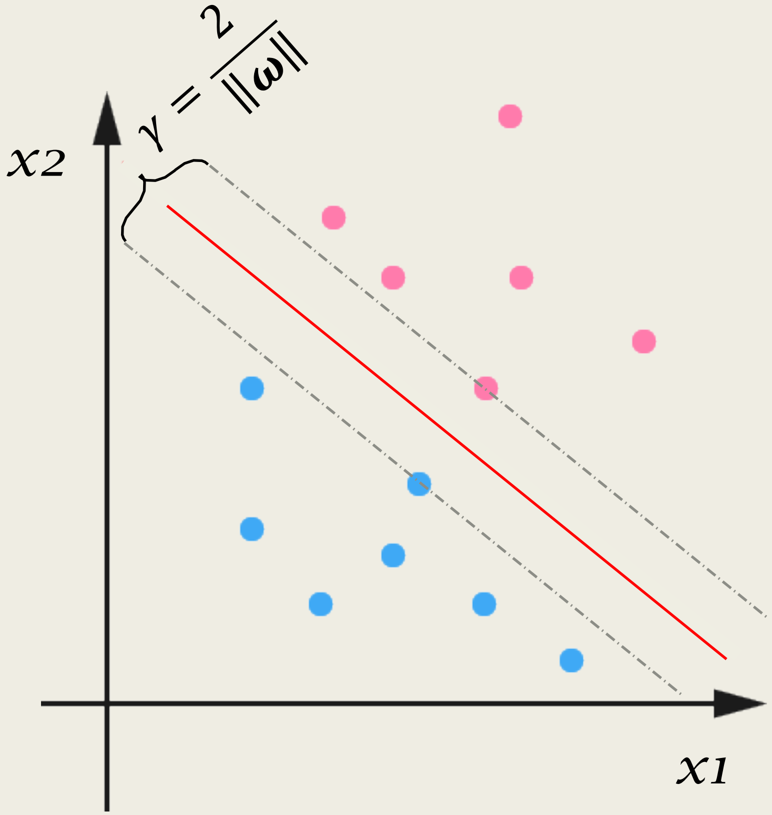
$$\text{s.t. } y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1, \qquad i = 1, 2, \ldots, m$$

$$\min_{\boldsymbol{\omega},b} \frac{1}{2}\|\boldsymbol{\omega}\|^2 \qquad (1\text{-}1)$$

$$\text{s.t. } y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1, \qquad i = 1, 2, \ldots, m$$

model

$$\boxed{f(x) = \boldsymbol{\omega}^T \boldsymbol{x} + b} \begin{cases} -1 \\ +1 \end{cases}$$



Dual problem 对偶问题

Solving for parameters ($\boldsymbol{\omega}$ and b) more effectively

Get a dual problem of equation (1-1) by using Lagrangian multiplier method

Introduce a Lagrangian multiplier $\alpha_i \geq 0$ to the limitation in equation (1-1), and it's Lagrange function is :

$$L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{\omega}\|^2 + \sum_{i=1}^{m} \alpha_i (1 - y_i (\boldsymbol{\omega}^T \boldsymbol{x}_i + b)) \qquad (1\text{-}2)$$

$$\boldsymbol{\alpha} = (\alpha_1; \alpha_2; \ldots; \alpha_m)$$

$$\min_{\boldsymbol{\omega}, b} \frac{1}{2}\|\boldsymbol{\omega}\|^2 \qquad (1\text{-}1)$$

$$\text{s.t. } y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1, \qquad i = 1, 2, \ldots, m$$

Partial derivative of $\boldsymbol{\omega}$ and b equal to 0:

$$\boldsymbol{\omega} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i$$

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

Partial derivative of $\boldsymbol{\omega}$ and b equal to 0:

$$\boxed{\boldsymbol{\omega} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i}$$

plugged it into (1-2)

$$L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{\omega}\|^2 + \sum_{i=1}^{m} \alpha_i (1 - y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b))$$

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

constraint

Dual problem of (1-1) is:

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

$$s.t. \quad \sum_{i=1}^{m} \alpha_i y_i = 0 \,,$$

$$\alpha_i \geq 0, \qquad i = 1, 2, \ldots, m.$$

Dual problem of (1-1) is:

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

$$s.t. \quad \sum_{i=1}^{m} \alpha_i y_i = 0 \,,$$

$$\alpha_i \geq 0, \qquad i = 1, 2, \ldots, m.$$

Use Sequential Minimal Optimization (**SMO**) method to solve $\boldsymbol{\alpha}$, then $\boldsymbol{\omega}$ and *b:*

$$f(x) = \boldsymbol{\omega}^T \boldsymbol{x} + b = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i^T \boldsymbol{x} + b$$

KKT condition[1]:

$$\begin{cases} \alpha_i \geq 0 \\ y_i f(\boldsymbol{x}_i) - 1 \geq 0 \\ \alpha_i (y_i f(\boldsymbol{x}_i) - 1) = 0 \end{cases}$$

because of

$$\min_{\boldsymbol{\omega}, b} \frac{1}{2} \|\boldsymbol{\omega}\|^2$$

$$\boxed{\text{s.t. } y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1, \quad i = 1, 2, \dots, m}$$

1. http://blog.csdn.net/xianlingmao/article/details/7919597

KKT condition:

$$\begin{cases} \alpha_i \geq 0 \\ y_i f(\boldsymbol{x}_i) - 1 \geq 0 \\ \alpha_i (y_i f(\boldsymbol{x}_i) - 1) = 0 \end{cases}$$

because of

$$\min_{\boldsymbol{\omega}, b} \frac{1}{2} \|\boldsymbol{\omega}\|^2$$

s.t. $y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1, \quad i = 1, 2, \dots, m$

α_i=0 or $y_i f(\boldsymbol{x}_i) = 1$. For $f(\boldsymbol{x}) = \boldsymbol{\omega}^T \boldsymbol{x} + b$,

if α_i=0 , $f(\boldsymbol{x}) = b$, the sample will not have any impact on $f(\boldsymbol{x})$

if $\alpha_i > 0$, $y_i f(\boldsymbol{x}_i) = 1$, corresponding samples are on the max-margin, these $\boldsymbol{x}_i$ are called support vectors

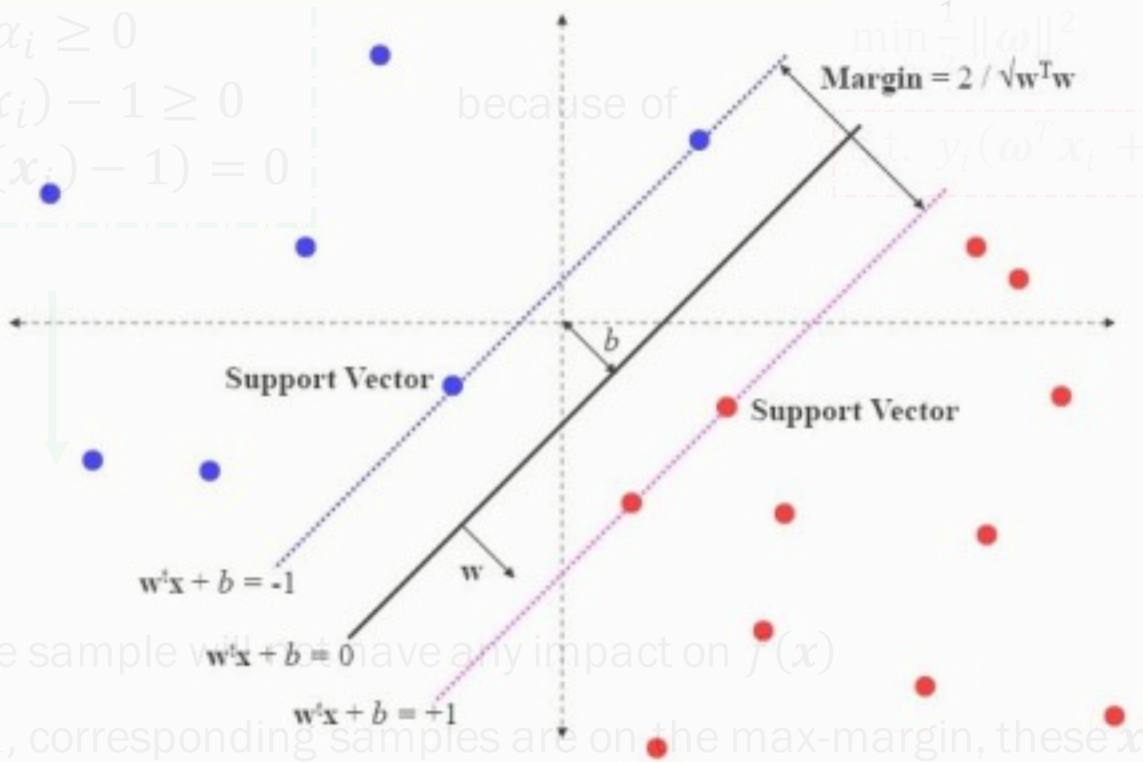
KKT condition:

$$\begin{cases} \alpha_i \geq 0 \\ y_i f(x_i) - 1 \geq 0 \\ \alpha_i (y_i f(x_i) - 1) = 0 \end{cases}$$

because of

$$\min \frac{1}{2} \|\omega\|^2$$

$$\text{s.t. } y_i(\omega^T x_i + b) \geq 1, \quad i = 1, 2, \ldots, m$$



α_i=0 or $y_i f(x_i) = 1$

If α_i=0 , $f(x) = b$, the sample will not have any impact on $f(x)$

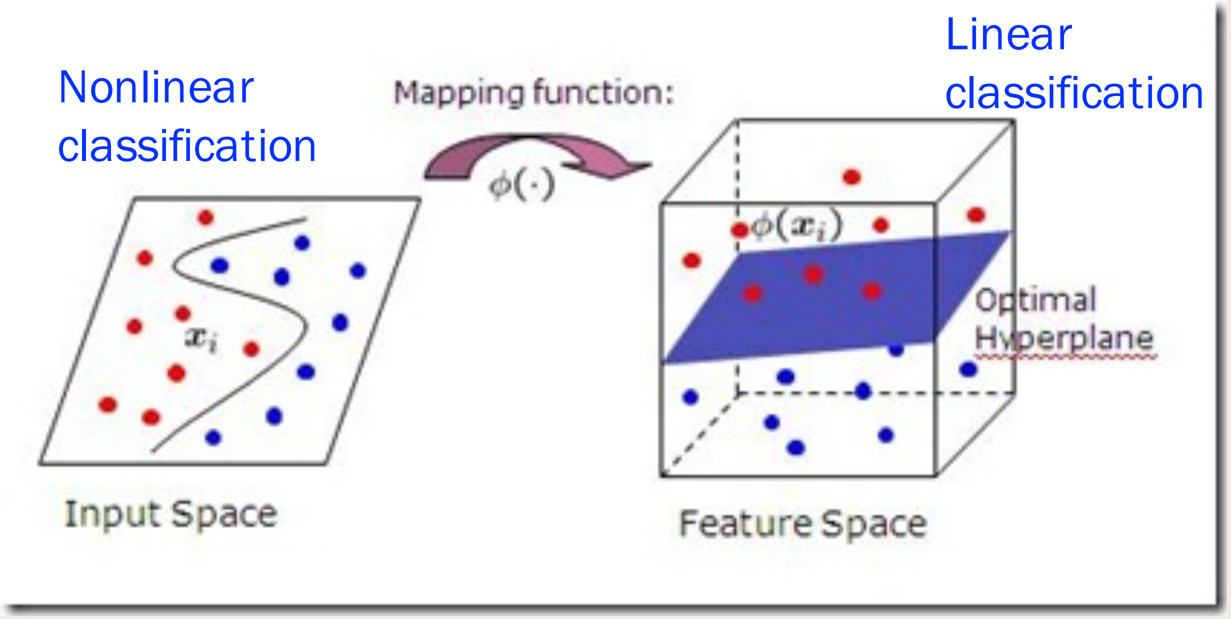
If $\alpha_i > 0$, $y_i f(x_i) = 1$, corresponding samples are on the max-margin, these x_i are called support vectors

The final model is only relevant to support vectors

Kernel function 核函数

$$f(x) = \boldsymbol{\omega}^T \boxed{\boldsymbol{x}} + b \longrightarrow f(x) = \boldsymbol{\omega}^T \boxed{\phi(\boldsymbol{x})} + b$$

feature vector
after mapping



Dual problem :

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j)$$

$$s.t. \quad \sum_{i=1}^{m} \alpha_i y_i = 0 \,,$$

$$\alpha_i \geq 0, \qquad\qquad i = 1, 2, \dots, m.$$

Dual problem :

Difficult to compute

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boxed{\phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j)}$$

$$s.t. \quad \sum_{i=1}^{m} \alpha_i y_i = 0 \,,$$

$$\alpha_i \geq 0, \qquad\qquad i = 1, 2, \dots, m.$$

Dual problem :

Difficult to compute

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boxed{\phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j)}$$

$$s.t. \quad \sum_{i=1}^{m} \alpha_i y_i = 0\,,$$

$$\alpha_i \geq 0, \qquad\qquad i = 1, 2, \ldots, m.$$

$$\kappa\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j)$$

Dual problem :

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boxed{\phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j)}$$

$$s.t. \ \sum_{i=1}^{m} \alpha_i y_i = 0 \, ,$$

$$\alpha_i \geq 0, \qquad i = 1, 2, \ldots, m.$$

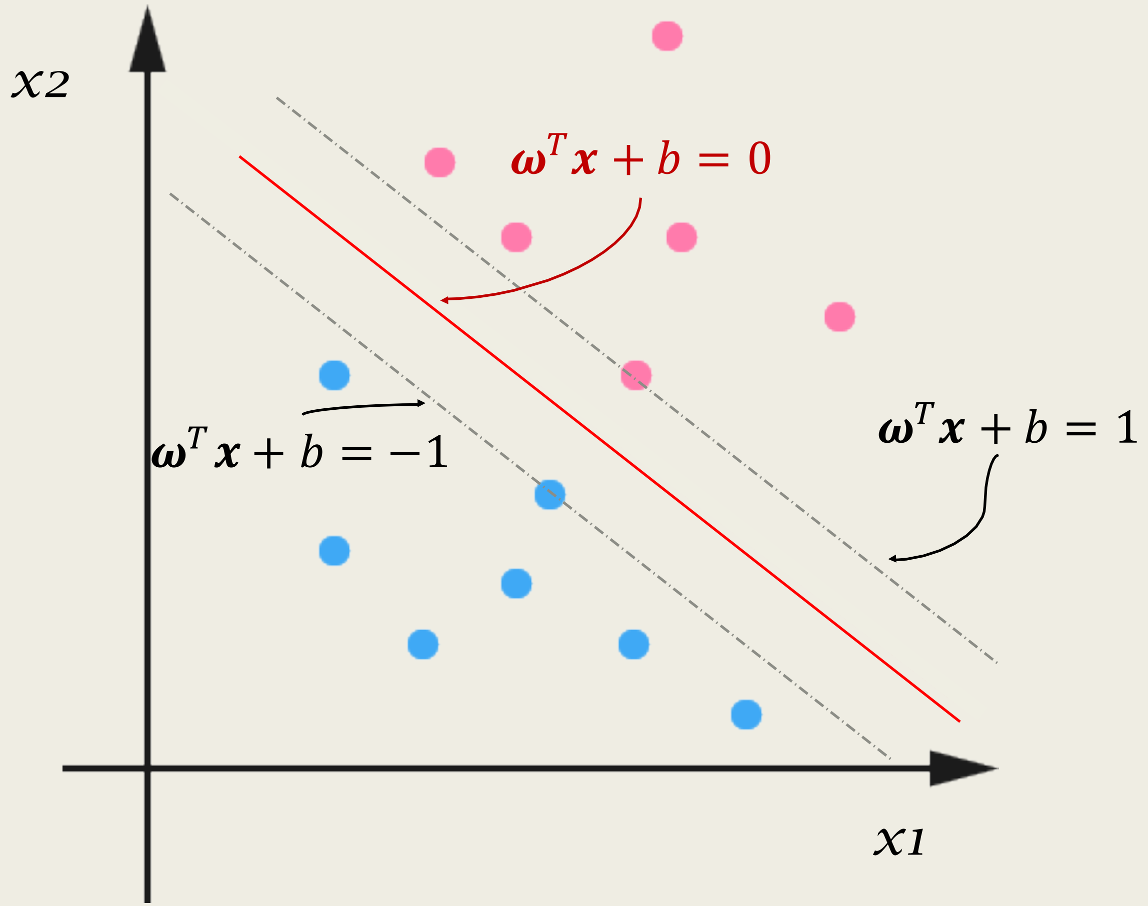
$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j)$$

kernel function

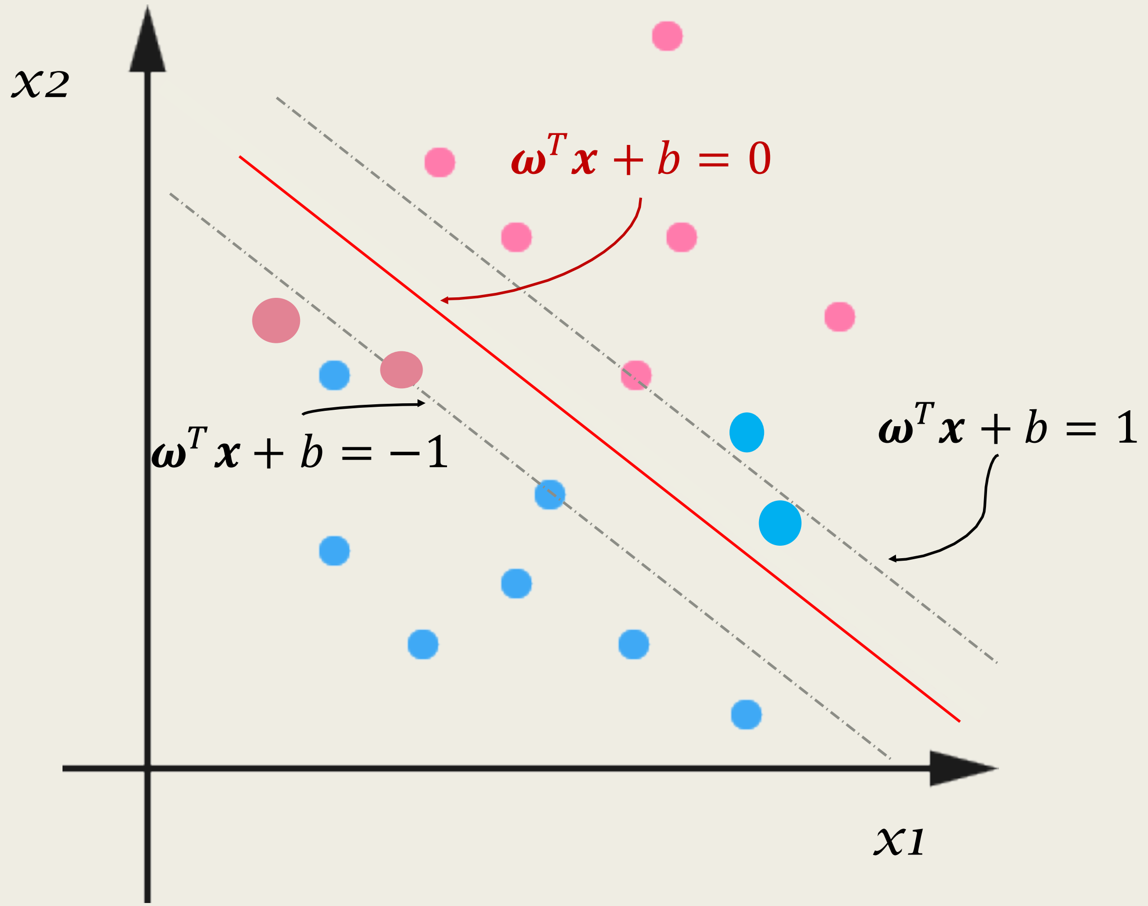
Some common kernel functions:

Name	Expression	Parameters
Linear	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^T \boldsymbol{x}_j$	
Polynomial	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^T \boldsymbol{x}_j)^d$	$d \geq 1$, degree of a polynomial
Gaussian(RBF)	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\lVert \boldsymbol{x}_i - \boldsymbol{x}_j \rVert^2}{2\sigma^2}\right)$	$\sigma > 0$, width
Laplacian	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\lVert \boldsymbol{x}_i - \boldsymbol{x}_j \rVert}{\sigma}\right)$	$\sigma > 0$
	……	

Soft margin and regularization



Soft margin and regularization



Some samples are allowed not to satisfy constraints

$$y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1$$

Some samples are allowed not to satisfy constraints

$$y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1$$

Optimization target:

$$\min_{\boldsymbol{\omega},b} \frac{1}{2}\|\boldsymbol{\omega}\|^2 \longrightarrow \min_{\boldsymbol{\omega},b} \frac{1}{2} \|\boldsymbol{\omega}\|^2 + \underbrace{\boxed{C}}_{\text{cost（大于0）}} \sum_{i=1}^{m} \boxed{\ell_{0/1}} (y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) - 1)$$

0/1损失函数 $\ell_{0/1} = \begin{cases} 1, & if\ z < 0 \\ 0, & otherwise \end{cases}$

Some samples are allowed not to satisfy constraints

引入容错性，给1这个硬性的阈值加一个"松弛变量"（slack variables） $\xi_i \geq 0$

$$y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1 - \xi_i$$

Optimization target :

$$\min_{\boldsymbol{\omega},b} \frac{1}{2}\|\boldsymbol{\omega}\|^2 \quad \Longrightarrow \quad \min_{\boldsymbol{\omega},b} \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{i=1}^{m} \ell_{0/1}\left(y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) - 1\right)$$

Some samples are allowed not to satisfy constraints

引入容错性，给1这个硬性的阈值加一个"松弛变量"（slack variables） $\xi_i \geq 0$

$$y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1 - \xi_i$$

Optimization target:

$$\min_{\boldsymbol{\omega},b} \frac{1}{2}\|\boldsymbol{\omega}\|^2 \quad \longrightarrow \quad \min_{\boldsymbol{\omega},b} \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{i=1}^{m} \ell_{0/1}\left(y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) - 1\right)$$

引入松弛变量后可重写成

$$\min_{\boldsymbol{\omega},b} \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{i=1}^{m} \xi_i$$

Some samples are allowed not to satisfy constraints

$$y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1$$

Optimization target:

用来描述划分超平面的“间隔”大小

$$\sum_{i=1}^{m} \xi_i$$

$$\min_{\boldsymbol{\omega},b} \frac{1}{2}\|\boldsymbol{\omega}\|^2 \longrightarrow \min_{\boldsymbol{\omega},b} \boxed{\frac{1}{2}\|\boldsymbol{\omega}\|^2} + C \boxed{\sum_{i=1}^{m} \ell_{0/1}(y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) - 1)}$$

用来表述训练集上的误差

General model

$$\min_{f} \Omega(f) + C \sum_{i=1}^{m} \ell(f(\boldsymbol{x}_i), y_i)$$

Some samples are allowed not to satisfy constraints

$$y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i + b) \geq 1$$

Optimization target :

$$\min_{\boldsymbol{\omega},b} \frac{1}{2}\|\boldsymbol{\omega}\|^2 \Longrightarrow \min_{\boldsymbol{\omega},b} \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{i=1}^{m} \ell_{0/1}(y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i + b) - 1)$$

General model

$$\min_f \Omega(f) + C\sum_{i=1}^{m} \ell(f(\boldsymbol{x}_i), y_i)$$

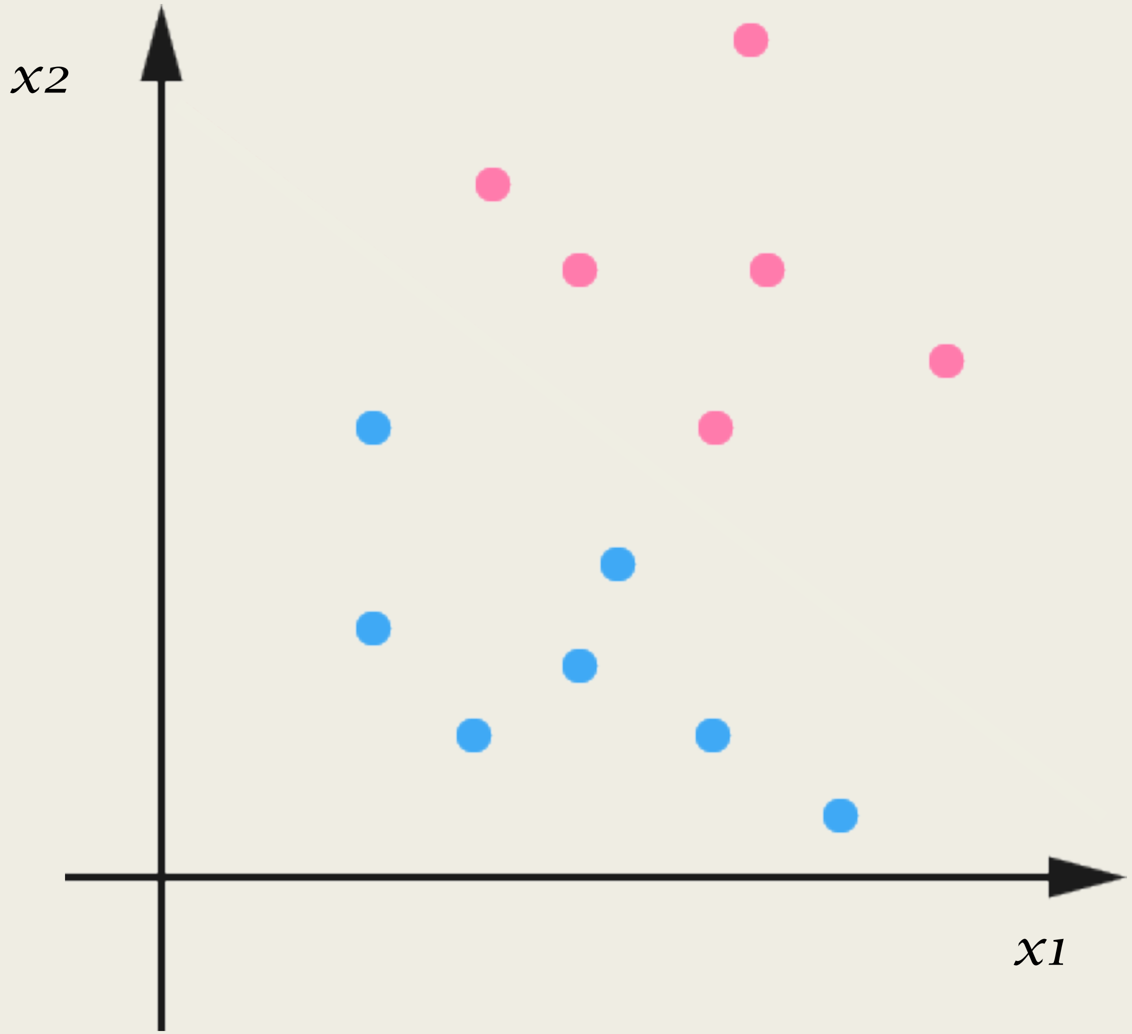
regularization

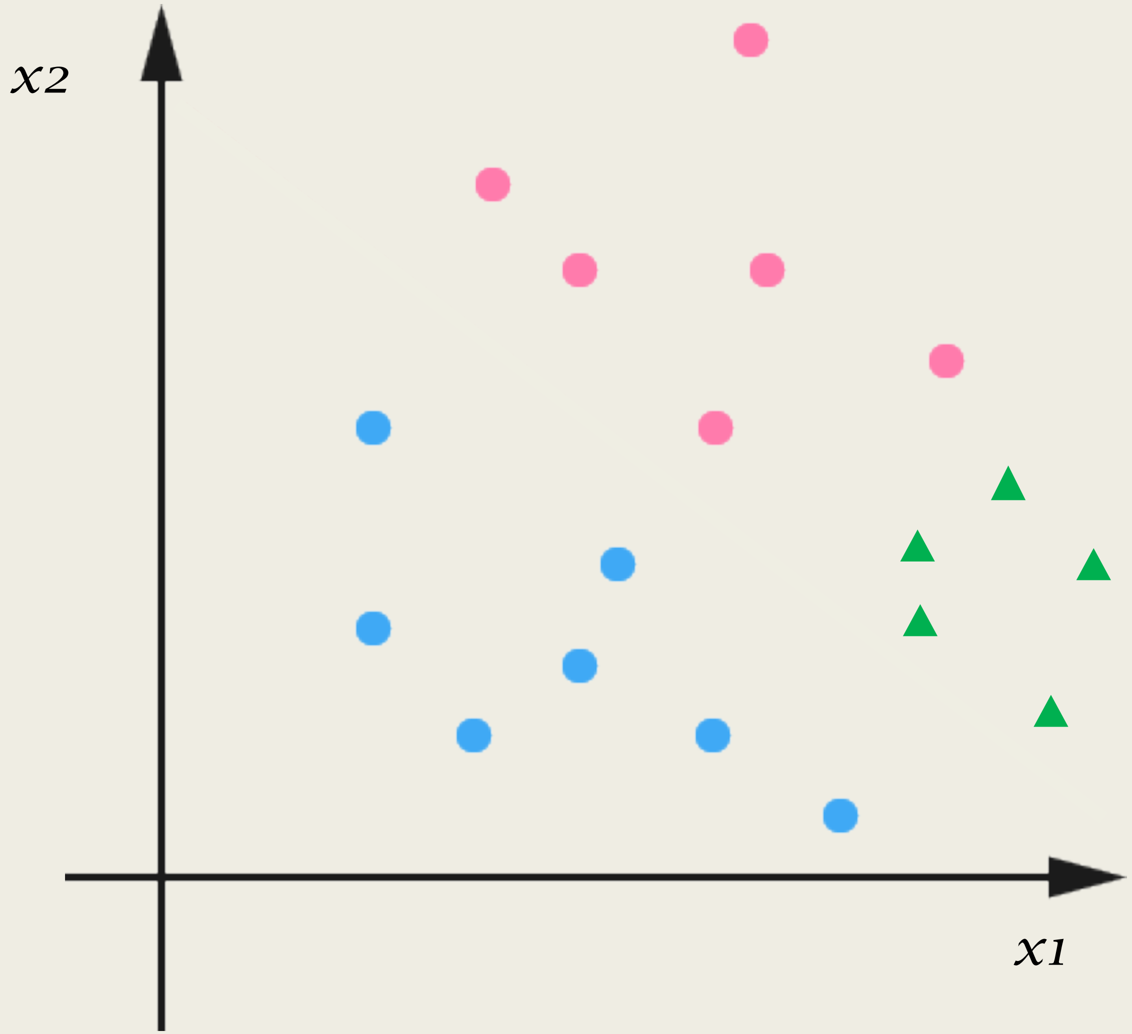
MULTI-CLASS

1-v-r SVMs

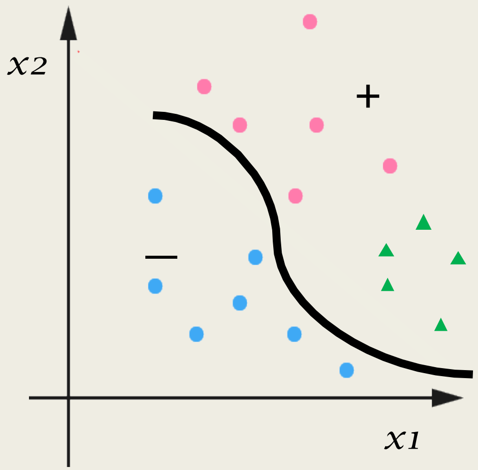
OVR SVMs

OVO SVMs


x2
x1

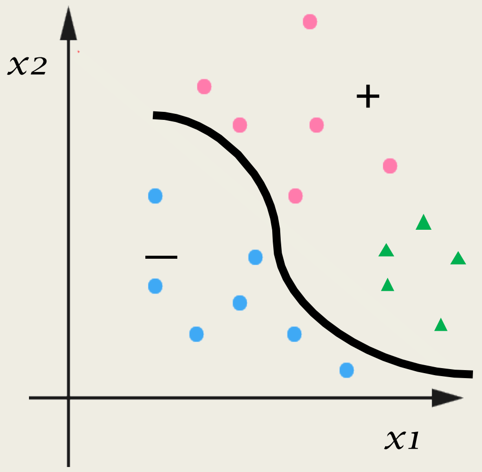

x2
x1

One-versus-rest (1-v-r SVMs) 一对多法

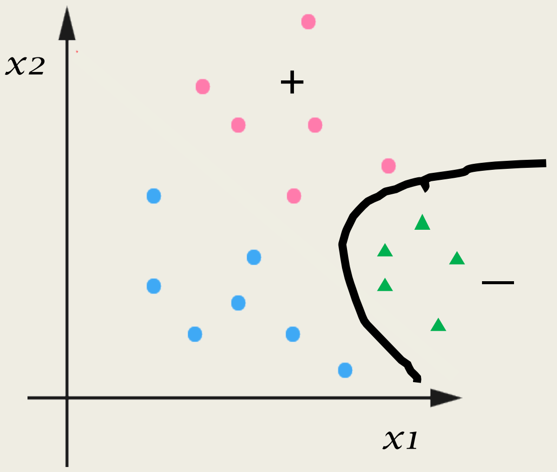


$f(\boldsymbol{x}_1)$

One-versus-rest (1-v-r SVMs) 一对多法

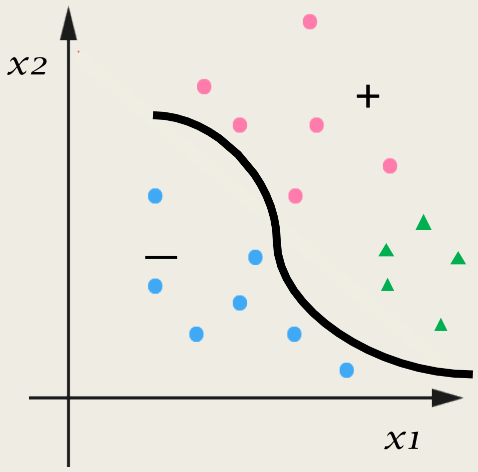


$f(\boldsymbol{x}_1)$

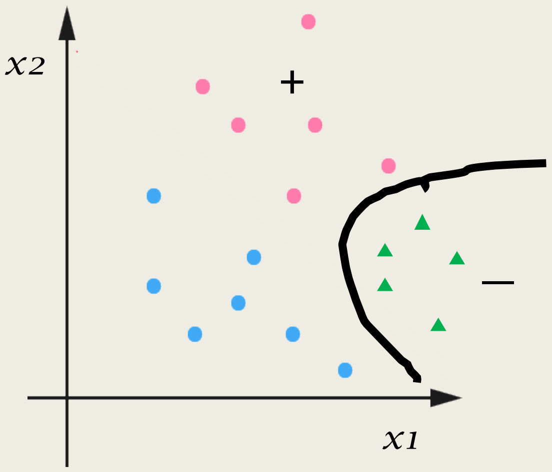


$f(\boldsymbol{x}_2)$

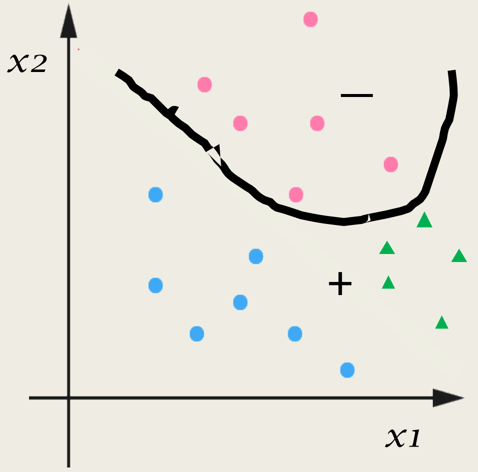
One-versus-rest (1-v-r SVMs) 一对多法



$f(\boldsymbol{x}_1)$



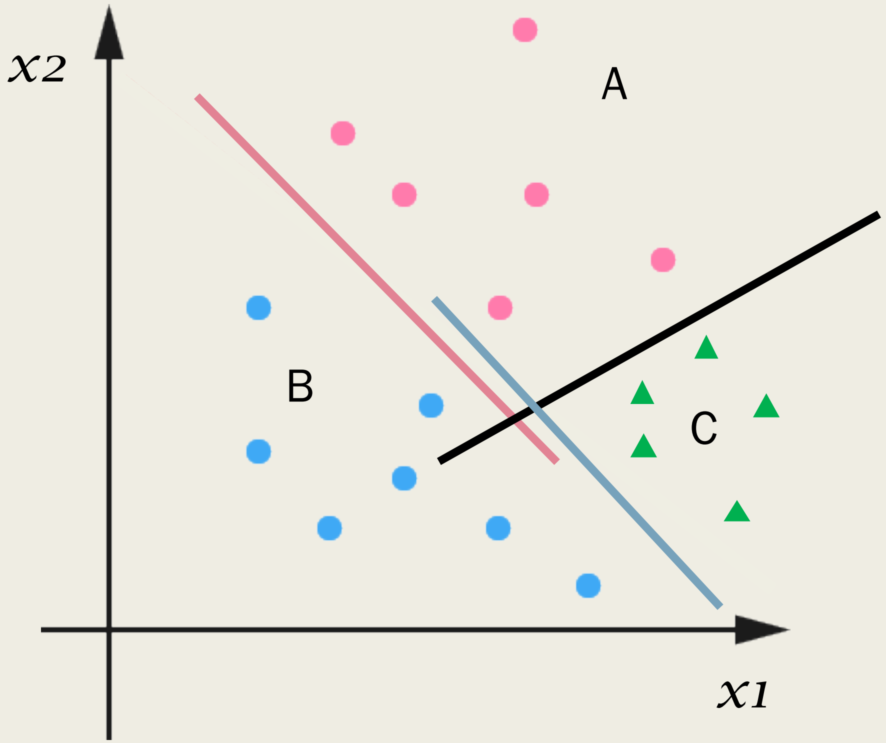
$f(\boldsymbol{x}_2)$



$f(\boldsymbol{x}_3)$

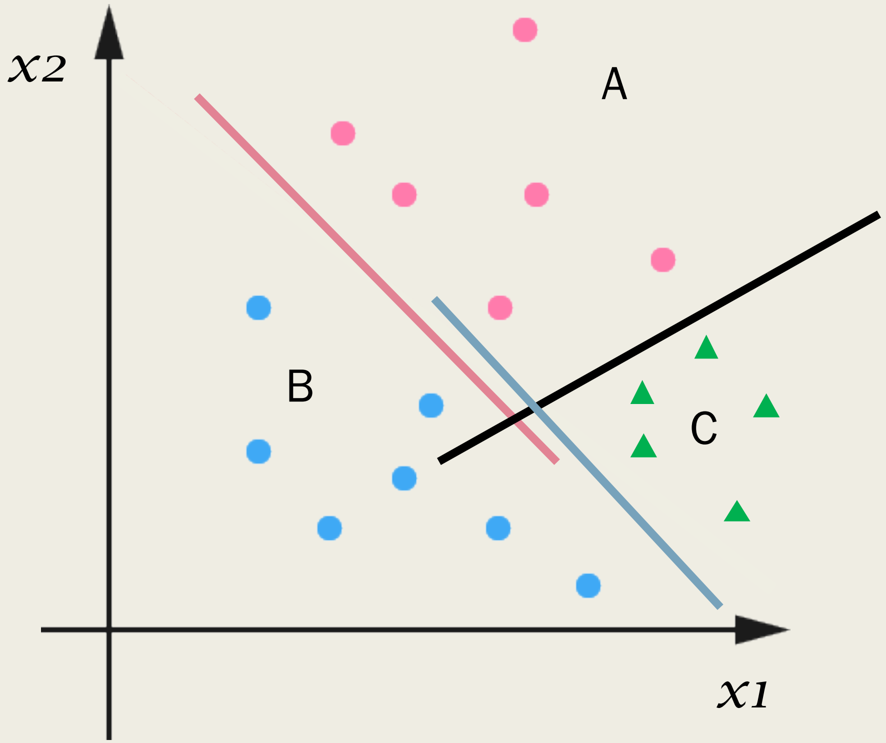
$$f(\boldsymbol{x}) = \max\{f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), f(\boldsymbol{x}_3)\}$$

One-versus-one (OVO SVMs) 一对一法



$$\frac{n\times(n-1)}{2} \longrightarrow \frac{3\times\,(3-1)}{2} = 3$$ classifiers

One-versus-one (OVO SVMs) 一对一法



(A,B) – classifier, if A win, voteA=voteA+1

(A,C) – classifier, if C win, voteC=voteC+1

(B,C) – classifier, if B win, voteA=voteB+1

$$f(\boldsymbol{x}) = \max\{\text{voteA}, \text{voteB}, \text{voteC}\}$$

$\frac{3\times(3-1)}{2} = 3$ classifiers

END

Discussion