《机器学习》课程——

支持向量机

主讲人： 徐义田教授

学　校：中国农业大学

目　录

- 研究背景
- 支持向量分类机
- 支持向量回归机
- 非平衡类支持向量机
- 总结

➢研究背景

一、研究背景

支持向量机是解决监督学习问题中一种强大的算法

发展历史

- 1963年：Vapnik 提出支持向量机的概念
- 1968年：Vapnik 和 Chervonenkis 提出VC维
- 1974年：提出结构风险最小化原则

……　　苏联解体前一年（1990），Vapnik 来到美国

- 1995年：Support vector network 文章发表

"Support-vector networks". Machine Learning, 1995, 20(3):273-297. (被引数：3.0万)

- 1995年：《The Nature of Statistical Learning》出版
- 1998年：SVM在文本分类中取得巨大成功
- 1998年：《Statistical Learning Theory》出版

……

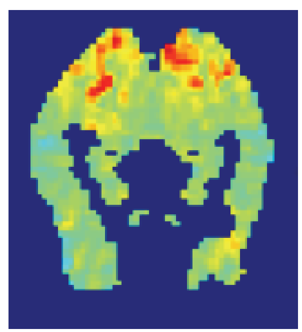
一、研究背景

应用：

① **分类问题：** 特征空间：X ⟶ 标签：Y

文本

→ {运动、新闻、政治、…}

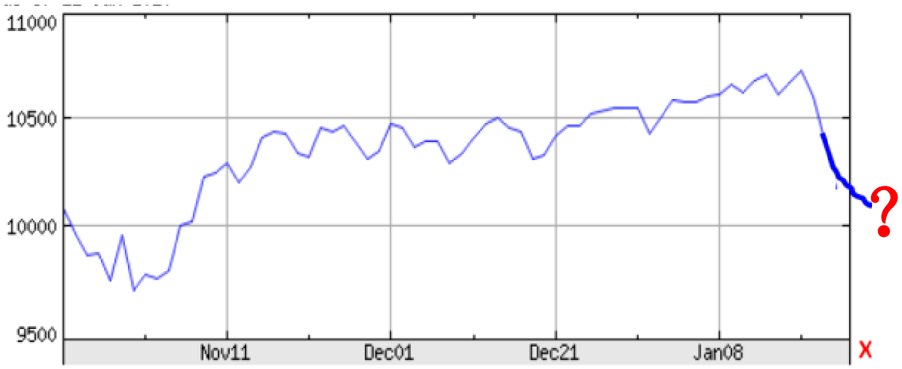


刺激反应

→ {有反应、无反应}

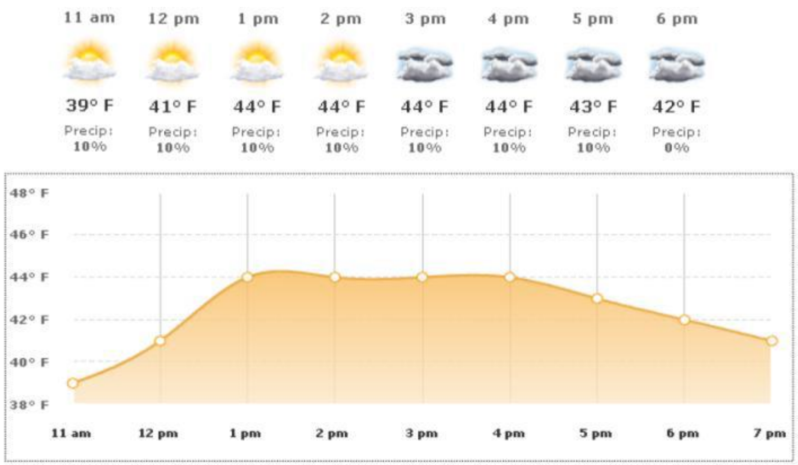
一、研究背景

② **回归问题：** 特征空间：X ⟶ 标签：Y



价格预测

价格：\$20.50



气温预测

气温：42^oF

中国农业大学 China Agricultural University

➢ 支持向量分类机

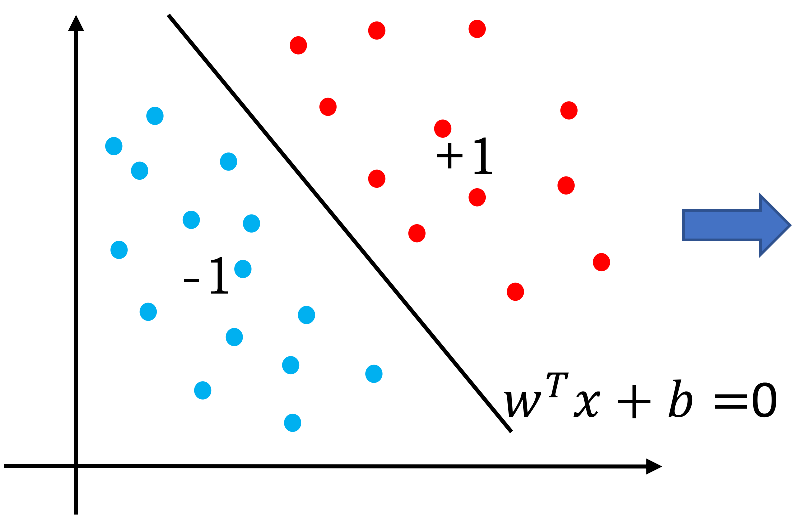
二、支持向量分类机（SVC）

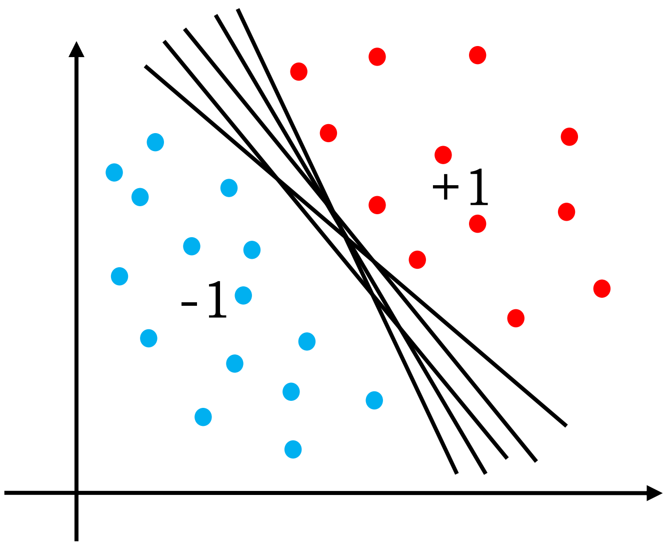
1. 线性情况

- **完全可分（硬间隔SVC）**

给定训练集$T=\{(x_1,y_1),(x_2,y_2),\dots,(x_l,y_l)\}$，其中$x_i\in R^n$，$y_i\in\{-1,+1\}$

完全可分： 可以找到一条线将两类点完全分开





如何找到最优的决策超平面$f(x)=\mathrm{sgn}(w^Tx+b)$？

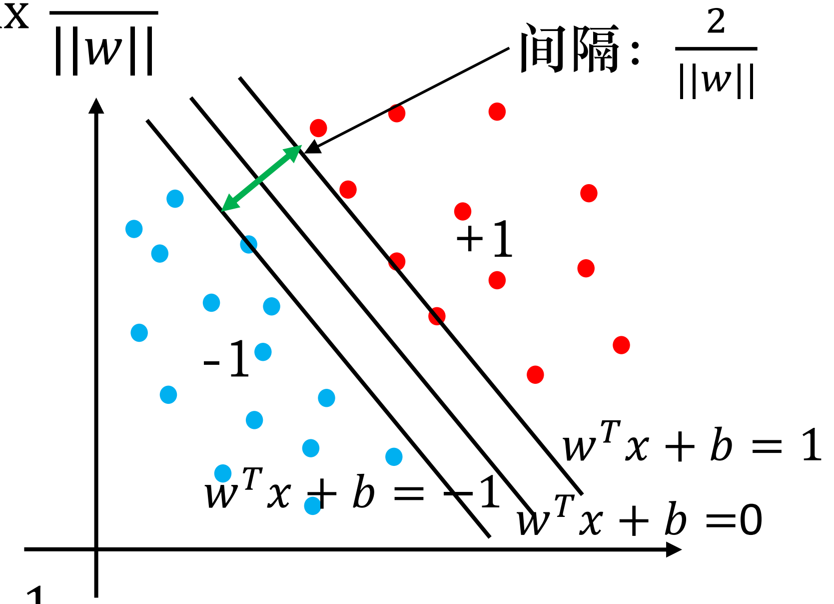
二、支持向量分类机（SVC）

1. 线性情况

- **完全可分（硬间隔SVC）**

基本思想： 基于最大化间隔的思想，分离两类点

最大间隔： $\max\limits_{w} \dfrac{2}{||w||}$



模型： $\min\limits_{w,b} \quad \dfrac{1}{2}||w||^2$

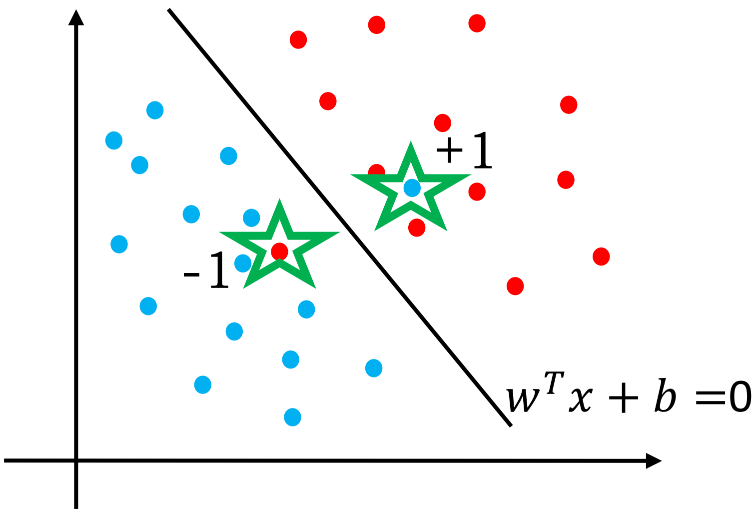
s.t. $\quad y_i\big((w \cdot x_i) + b\big) \geq 1, i = 1, \cdots, l$

二、支持向量分类机（SVC）

1. 线性情况

- **不完全可分（软间隔SVC）**

不完全可分：找不到一条线将两类点完全分开



（注：完全可分是不完全可分的特殊情况）

基本思想：基于最大化间隔的思想，尽可能的分离两类点

中国农业大学 China Agricultural University

1. 线性情况

- **不完全可分（软间隔SVC）**

原问题：

$$\min_{w,b,\xi} \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i$$

$$s.t.\ \ y_i(w^T x_i + b) \geq 1 - \xi_i\,, \quad i = 1, \dots, l$$

$$\xi_i \geq 0, \qquad i = 1, \dots, l$$

对偶问题：

$$\min_{\alpha} \ \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j (x_i, x_j)\alpha_i\alpha_j - \sum_{j=1}^{l}\alpha_j$$

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

$$0 \leq \alpha_i \leq C\,, \qquad i = 1, \dots, l.$$

二、支持向量分类机（SVC）

1. 线性情况

对偶问题推导： 引入拉格朗日函数：

$$L=\frac{1}{2}||w||^2+C\sum_{i=1}^{l}\xi_i-\sum_{i=1}^{l}\alpha_i(y_i(w\cdot x_i+b)-1+\xi_i)-\sum_{i=1}^{l}\beta_i\xi_i$$

将$L(w,b,\alpha)$分别关于w,b,ξ_i求导，最优性条件可得：

$$\begin{cases}
① \ \dfrac{\partial L}{\partial w}=w-\sum_{i=1}^{l}\alpha_i y_i x_i=0 \Rightarrow w=\sum_{i=1}^{l}\alpha_i y_i x_i \\
② \ \dfrac{\partial L}{\partial b}=\sum_{i=1}^{l}\alpha_i y_i=0 \\
③ \ \dfrac{\partial L}{\partial \xi_i}=C-\alpha_i-\beta_i=0 \Rightarrow 0\le\alpha_i\le C \\
\left.\begin{array}{l} ④ \ \alpha_i(y_i(w\cdot x_i+b)-1+\xi_i)=0 \\ ⑤ \ \beta_i\xi_i=0 \end{array}\right\} \text{松弛互补条件} \\
⑥ \ \alpha_i\ge 0,\beta_i\ge 0 \\
⑦ \ y_i(w\cdot x_i+b)\ge 1-\xi_i
\end{cases}$$

二、支持向量分类机（SVC）

1. 线性情况

- **不完全可分（软间隔SVC）**

Step 1：解得 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_l^*)^T$

Step 2：计算 b^*：选取位于开区间$(0, C)$中的α^*的分量α_j^*，据此计算

$$b^* = y_i - \sum_{i=1}^{l} y_i \alpha_i^* (x_i, x_j)$$

Step 3：构造分划超平面$(w^* \cdot x) + b^* = 0$，由此求得决策函数

$$f(x) = sgn(g(x))$$

其中

$$g(x) = \sum_{i=1}^{l} \alpha_i y_i (x_i, x) + b$$

1. 线性情况

- **不完全可分（软间隔SVC）**

对偶问题的解与样本位置的关系：

当$\alpha_i = 0$时 $\xrightarrow{③} \beta_i = C \xrightarrow{⑤} \xi_i = 0 \xrightarrow{⑦} y_i(w \cdot x_i + b) \geq 1 \longrightarrow$ 线上或线外

当$\alpha_i = \mathrm{C}$时 $\xrightarrow{③} \beta_i = 0 \xrightarrow{⑤} \xi_i \neq 0 \xrightarrow{⑦} y_i(w \cdot x_i + b) \leq 1 \longrightarrow$ 线上或线内

当$0 < \alpha_i < C$时 $\xrightarrow{③} \beta_i > 0 \xrightarrow{⑤} \xi_i = 0 \xrightarrow{⑦} y_i(w \cdot x_i + b) = 1 \longrightarrow$ 线上

反之，

线上的点：$\alpha_i = 0$，$\alpha_i = C$，$0 < \alpha_i < C$

线两侧的点：$\alpha_i = 0$

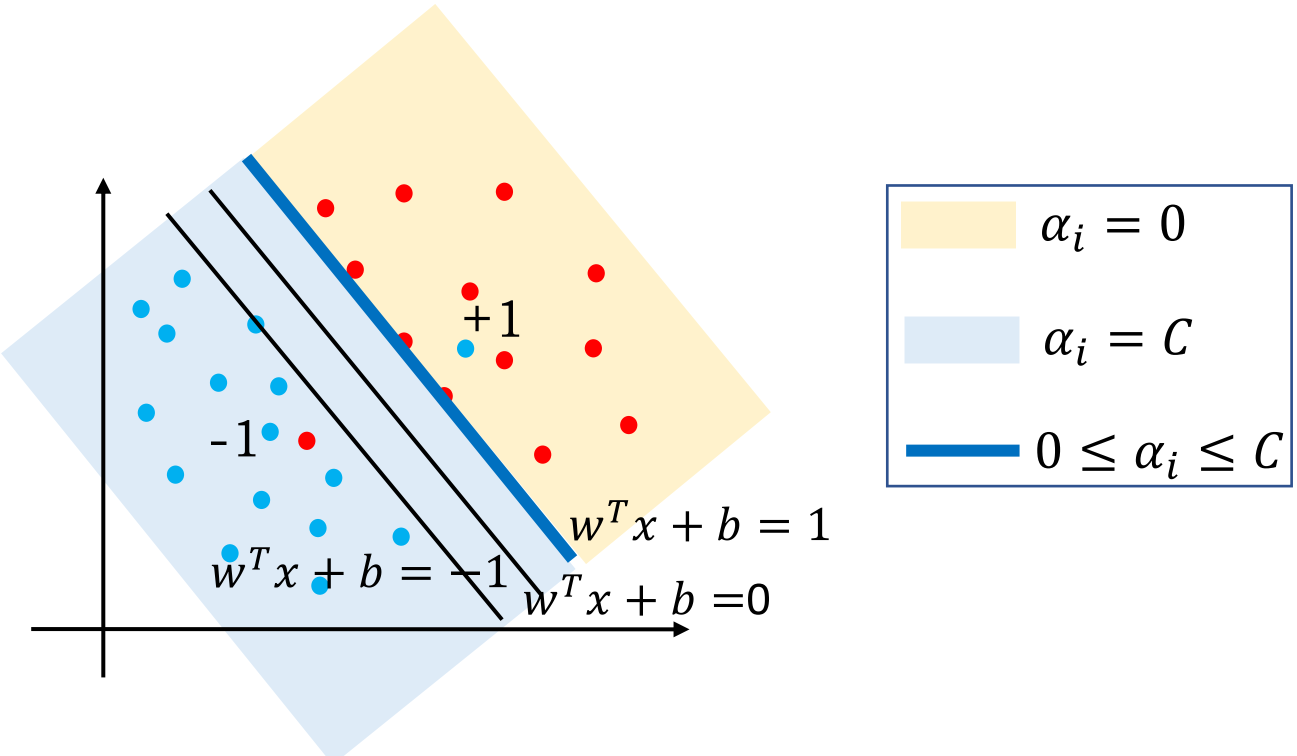
线中间的点：$\alpha_i = C$

二、支持向量分类机（SVC）

1. 线性情况

- 不完全可分（软间隔SVC）

几何解释：



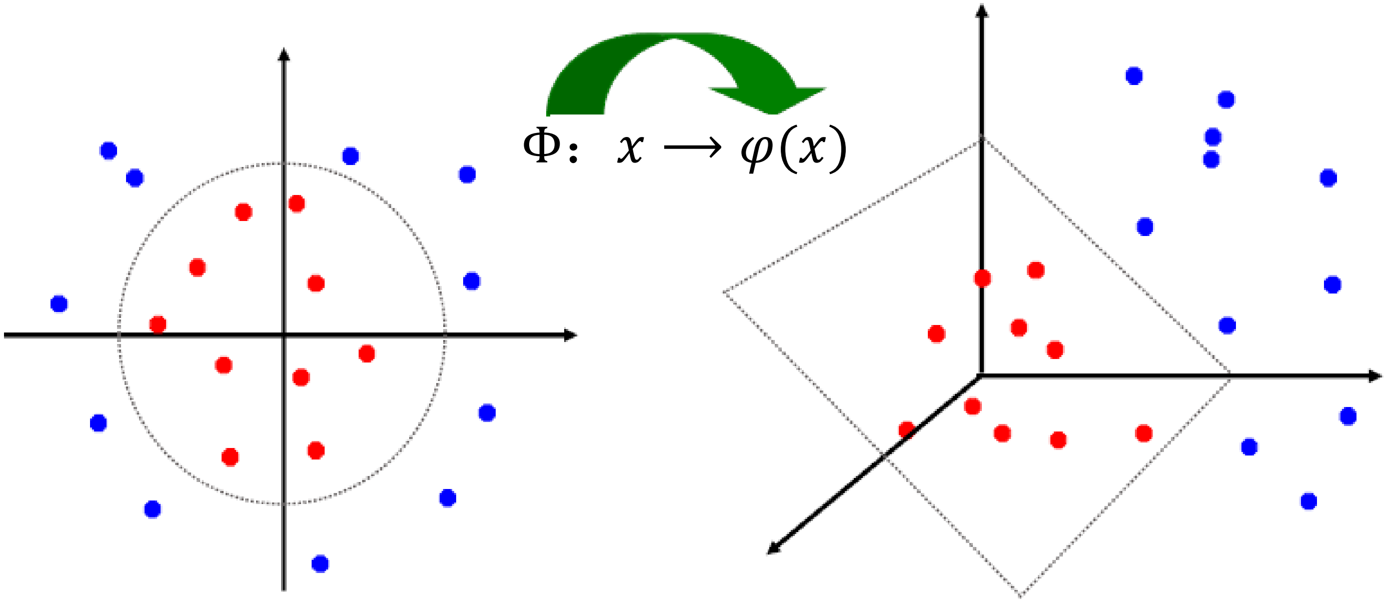
支持向量： 只要是$\alpha_i \neq 0$对应的样本点都称为支持向量

（注：支持向量有可能在线上，可能不在线上，并且线上的点不一定都是支持向量）

二、支持向量分类机（SVC）

2. 非线性情况

非线性可分： 如果在低维空间中无法线性可分的样本，在高维空间中可以找到一个超平面可以将其分开



线性不可分　　　　线性可分

原因：非线性问题往往不好求解，所以希望能用解线性分类问题的方法解决

2. 非线性情况

主要思想： 在原始空间中无法线性可分时，

- 将其映射到高维空间，使其在高维空间中线性可分
- 通过求解变换后的线性问题来求解原来的非线性问题

核函数： 设X是输入空间，H为特征空间，如果存在一个从X到H的映射：

$$\Phi\text{：} X \longrightarrow H$$

使得对所有$x, z \in X$，函数$K(x,z)$满足条件：

$$\boxed{K(x,z)=\varphi(x)\cdot\varphi(z)}$$

则称$K(x,z)$为核函数，$\varphi(x)$为映射函数.

二、支持向量分类机（SVC）

2. 非线性情况

常见的核函数：

基本经验：文本数据常用线性核，
情况不明时可先尝试高斯核

- 线性核：$K(x_i,x_j)=x_i^T x_j$
- 多项式核：$K(x_i,x_j)=(x_i^T x_j)^d$ （$d\geq 1$为多项式的次数）
- 高斯核：

$$K(x_i,x_j)=\exp\left(-\frac{||x_i-x_j||^2}{2\sigma^2}\right)$$ （ σ > 0为高斯核的带宽）

- 拉普拉斯核：

$$K(x_i,x_j)=\exp\left(-\frac{||x_i-x_j||}{\sigma}\right)$$ （ σ > 0）

- Sigmoid核：

$$K(x_i,x_j)=\tanh(\beta x_i^T x_j+\theta)$$

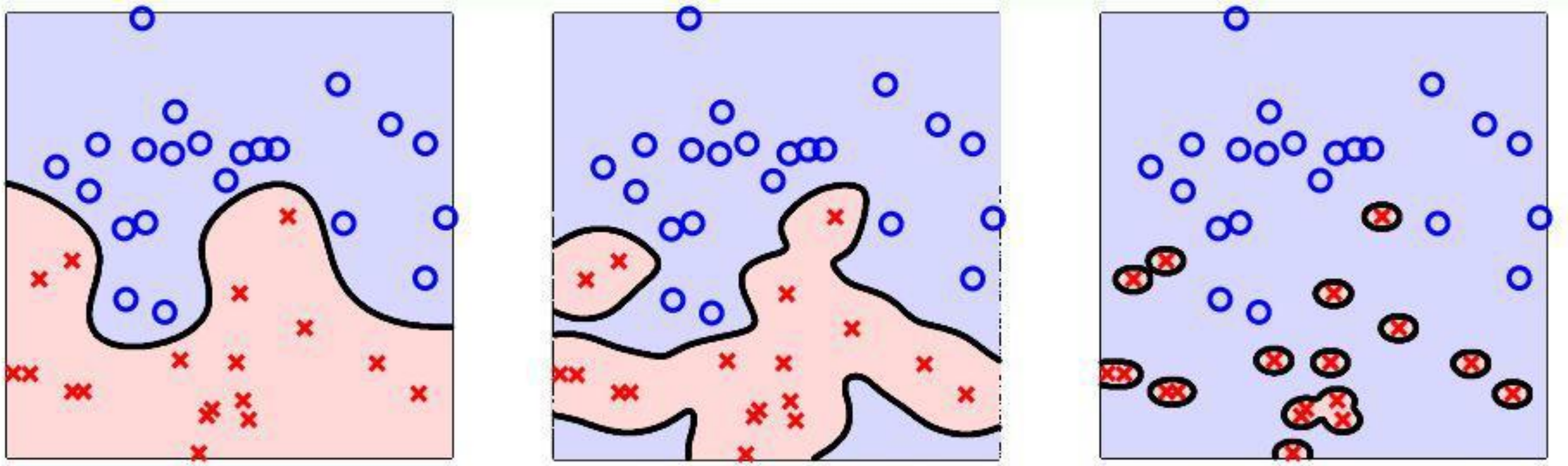
（tanh为双曲正切函数，$\beta>0,\theta<0$）

二、支持向量分类机（SVC）

2. 非线性情况

核函数参数的取值不同分类效果也不同

举例： 高斯核函数



$\exp(-1||x_i - x_j||^2)$　　　$\exp(-10||x_i - x_j||^2)$　　　$\exp(-100||x_i - x_j||^2)$

过拟合

一般通过交叉验证选取合适的核参数

二、支持向量分类机（SVC）

2. 非线性情况

核函数在支持向量机中的应用

设样本x映射后的向量为$\varphi(x)$，划分超平面为$f(x) = w^T\varphi(x) + b$

原始问题：

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i$$

$$s.t.\ \ y_i(w^T\varphi(x_i) + b) \geq 1 - \xi_i\,, \quad i = 1, \dots, l$$

$$\xi_i \geq 0, \qquad i = 1, \dots, l$$

对偶问题：

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j K(x_i, x_j)\alpha_i\alpha_j - \sum_{j=1}^{l}\alpha_j$$

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

$$0 \leq \alpha_i \leq C\,, \qquad i = 1, \dots, l.$$

预测：

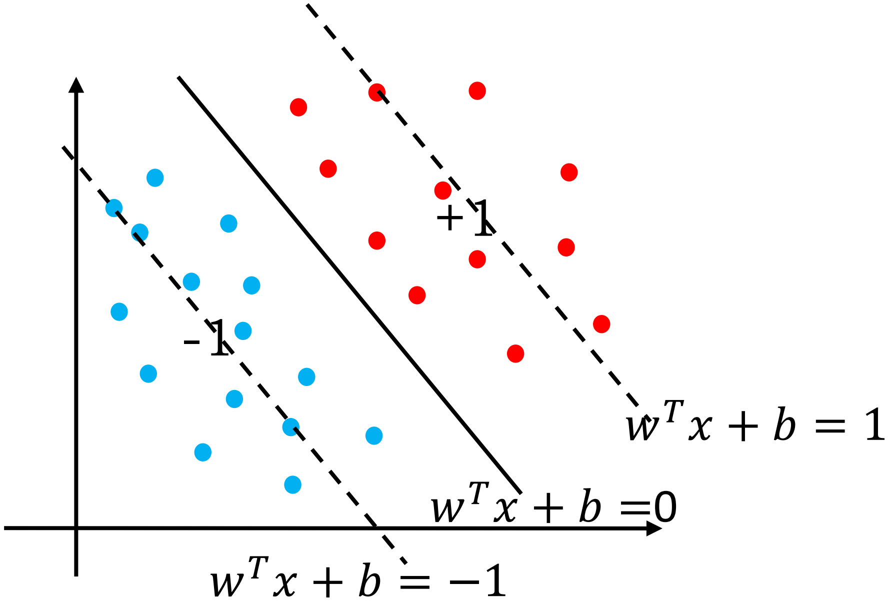
$$f(x) = w^T\varphi(x) + b = \sum_{i=1}^{l}\alpha_i y_i K(x_i, x_j) + b$$

3. 拓展模型

① 最小二乘支持向量机（LS-SVM）

主要思想： 将SVM优化问题的非等式约束用等式约束替换

几何解释：



① 间隔最小化

② 正负类样本点分别尽可能在两条虚线周围

3. 拓展模型

① 最小二乘支持向量机（LS-SVM）

原问题：

$$\min_{w,\eta,b} \quad \frac{1}{2}||w||^2 + \frac{C}{2}\sum_{i=1}^{l}\eta_i^2$$

$$s.t. \quad y_i\left(\left(w \cdot \phi(x_i)\right) + b\right) = 1 - \eta_i, i = 1, \cdots, l$$

对偶模型：

$$\min_{\alpha} \quad -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \left(K(x_i, x_j) + \frac{\delta_{ij}}{C}\right)\alpha_i\alpha_j + \sum_{i=1}^{l}\alpha_i$$

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

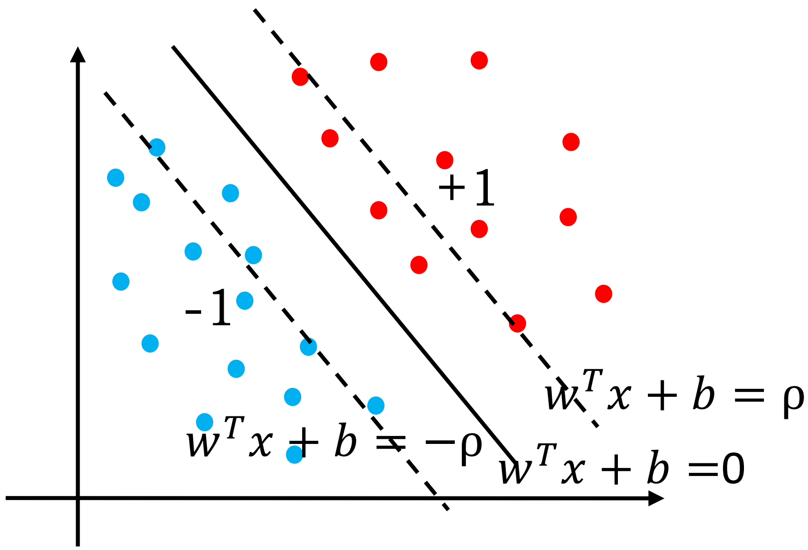
其中 $\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$

3. 拓展模型

② ν −支持向量机（ν −SVM）

主要思想： 用参数ν代替原始SVM中的参数C，使参数具有一定的含义

几何解释：



注：当$\rho = 1$时，即为原始的SVM

3. 拓展模型

② ν –支持向量机（ν –SVM）

原问题：

$$\min_{w,b,\xi,\rho} \frac{1}{2}\|w\|^2 - \nu\rho + \frac{1}{l}\sum_{i=1}^{l}\xi_i$$

$$s.t.\ \ y_i(w\cdot\phi(x_i)+b) \geq \rho - \xi_i\,,\quad i=1,\dots,l$$

$$\xi_i \geq 0, i=1,\dots,l;\ \rho \geq 0$$

对偶问题：

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{\mathrm{l}}\sum_{j=1}^{l} y_i y_j K(x_i,x_j)\alpha_i\alpha_j$$

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

$$\sum_{i=1}^{l}\alpha_i \geq \nu$$

$$0 \leq \alpha_i \leq \frac{1}{l}\,,\qquad i=1,\dots,l.$$

中國農業大學 China Agricultural University

3. 拓展模型

② ν −支持向量机（ν −SVM）

参数ν的意义：

间隔错误训练点：没有被“充分”正确划分的训练点

- 若记间隔错误训练点的个数为p, 则$\nu \geq \frac{p}{l}$

  (ν是建个错误训练点的个数所占总训练点数的上界)

- 若记支持向量的个数为q, 则$\nu \leq \frac{q}{l}$

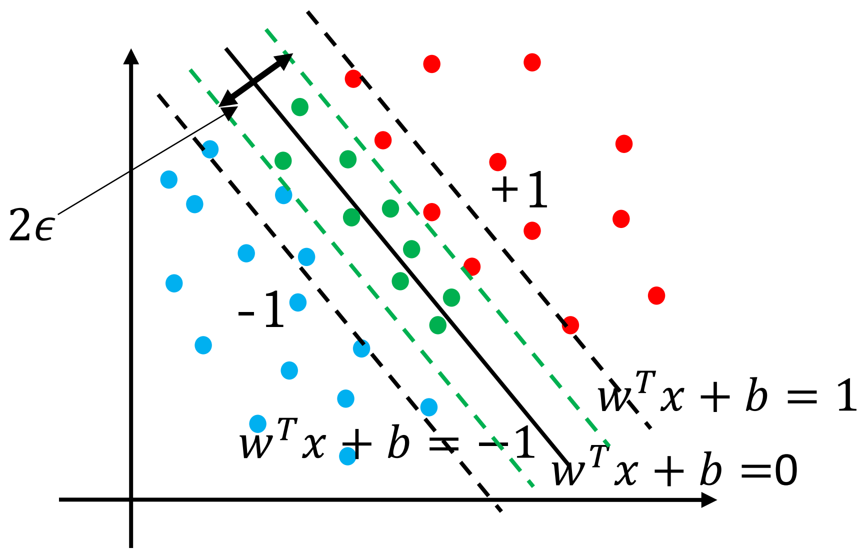
  (ν是支持向量的个数所占总训练点数的下界)

3. 拓展模型

③ U –支持向量机（U –SVM）

给定训练集$T=\{(x_1,y_1),(x_2,y_2),\dots,(x_l,y_l)\}$，其中$x_i\in R^n$，$y_i\in\{-1,+1\}$, 和无类别标记的样本集合$U=\{x_1^*,\dots,x_u^*\}$.

几何解释:



① 正负类样本点尽可能在两条黑色虚线以外

② U-类点尽可能在两条绿色虚线中间

二、支持向量分类机（SVC）

3. 拓展模型

③ U –支持向量机（U –SVM）

原问题：

$$\min_{w,b,\xi,\psi} \frac{1}{2}\|w\|^2 + C_t\sum_{i=1}^{l}\xi_i + C_u\sum_{s=1}^{u}(\psi_s+\psi_s^*)$$

$$s.t.\ \ y_i(w\cdot x_i+b)\ge 1-\xi_i\,,\xi_i\ge 0, i=1,\dots,l$$

$$\boxed{-\varepsilon-\psi_s^*\le (w\cdot x_s^*)+b\le \varepsilon+\psi_s, s=1,\dots,u}$$

$$\psi_s,\psi_s^*\ge 0, s=1,\dots,u$$

对偶问题：

$$\max_{\alpha,\mu,\nu}\ -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(x_i,x_j)-\frac{1}{2}\sum_{s,t=1}^{u}(\mu_s-\nu_s)(\mu_t-\nu_t)K(x_s^*,x_t^*)$$

$$-\sum_{i=1}^{l}\sum_{s=1}^{u}\alpha_i y_i(\mu_s-\nu_s)K(x_i,x_s^*)+\sum_{i=1}^{l}\alpha_i-\varepsilon\sum_{s=1}^{u}(\mu_s+\nu_s)$$

$$s.t.\quad \sum_{i=1}^{l}y_i\alpha_i+\sum_{s=1}^{u}(\mu_s-\nu_s)=0$$

$$C_t-\alpha_i-\eta_i=0, i=1,\dots,l$$

$$C_u-\nu_s-\gamma_s=0, s=1,\dots,u$$

$$C_u-\mu_s-\gamma_s^*=0, s=1,\dots,u$$

$$\nu_s,\gamma_s\ge 0, s=1,\dots,u$$

$$\mu_s,\gamma_s^*\ge 0, s=1,\dots,u$$

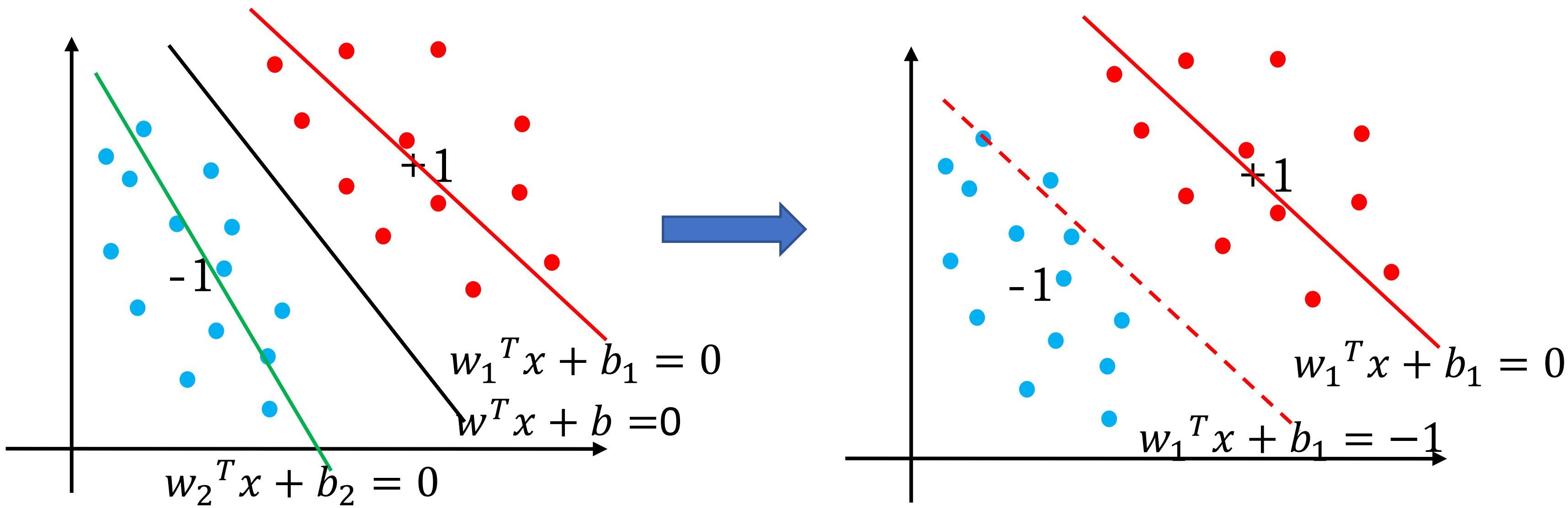
注：充分利用了未标记的样本，但对偶问题的规模变大

3. 拓展模型

④ 双子支持向量机（TSVM）

主要思想：通过构造两条非平行的超平面，分离正负类样本点

几何解释：



$\begin{cases} \text{① 正类样本点尽可能在红色实线周围} \\ \text{② 负类样本点尽可能在红色虚线下方} \end{cases}$

China Agricultural University 中国农业大学

3. 拓展模型

④ 双子支持向量机（TSVM）

原问题：

$$\min_{w^{(1)},b^{(1)},q} \frac{1}{2}\left(Aw^{(1)}+e_1 b^{(1)}\right)^T\left(Aw^{(1)}+e_1 b^{(1)}\right)+c_1 e_2^T q$$
$$s.t.\ -\left(Bw^{(1)}+e_2 b^{(1)}\right)+q \geq e_2, q \geq 0$$

$$\min_{w^{(2)},b^{(2)},q} \frac{1}{2}\left(Bw^{(2)}+e_2 b^{(2)}\right)^T\left(Bw^{(2)}+e_2 b^{(2)}\right)+c_2 e_1^T q$$
$$s.t.\ \left(Aw^{(2)}+e_1 b^{(2)}\right)+q \geq e_1, q \geq 0$$

对偶问题：

$$\max_{\alpha}\ e_2^T\alpha-\frac{1}{2}\alpha^T G(H^T H)^{-1}G^T\alpha$$
$$s.t.\ 0 \leq \alpha \leq c_1$$

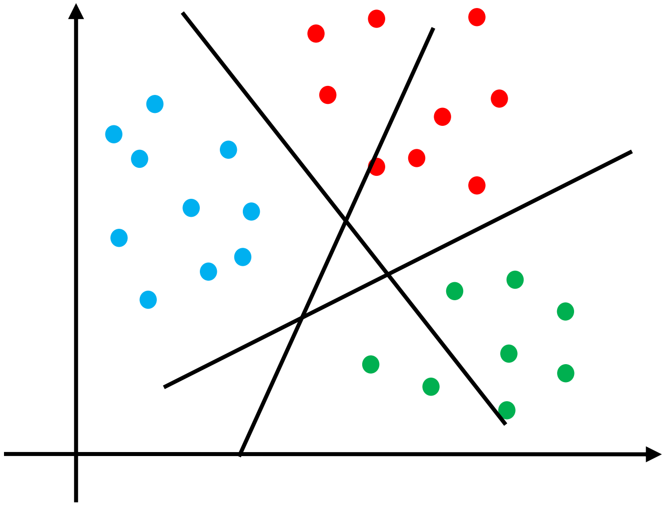
$$\max_{\gamma}\ e_1^T\gamma-\frac{1}{2}\gamma^T P(Q^T Q)^{-1}P^T\gamma$$
$$s.t.\ 0 \leq \gamma \leq c_2$$

3. 拓展模型

⑤ 多分类支持向量机（1 vs 1）

主要思想： 将多分类问题转化为多个二分类问题处理

几何解释：



二、支持向量分类机（SVC）

4. 求解算法：

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j (x_i, x_j)\alpha_i \alpha_j - \sum_{j=1}^{l}\alpha_j$$

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

$$0 \le \alpha_i \le C, \qquad i = 1, \dots, l.$$

序列最小化算法：（SMO）

基本思想：

- 如果α^*的所有分量都满足KKT条件，那么α^*即为最优解
- 否则，选择α^*两个分量，固定其他分量，
- 针对这两个变量构建一个二次规划问题，然后求解它，
- 直到所有分量都满足KKT条件

4. 求解算法：

常用算法总结

Algorithm	Citation	SVM type	Optimization type	Style	Runtime
SMO	[Platt, 1999]	Kernel	Dual QP	Batch	$\Omega(n^2 d)$
$\text{SVM}^{\text{light}}$	[Joachims, 1999]	Kernel	Dual QP	Batch	$\Omega(n^2 d)$
Core Vector Machine	[Tsang et al., 2005, 2007]	SL Kernel	Dual geometry	Batch	$O(s/\rho^4)$
SVM^{perf}	[Joachims, 2006]	Linear	Dual QP	Batch	$O(ns/\lambda\rho^2)$
NORMA	[Kivinen et al., 2004]	Kernel	Primal SGD	Online(-style)	$\tilde{O}(s/\rho^2)$
SVM-SGD	[Bottou, 2007]	Linear	Primal SGD	Online-style	Unknown
Pegasos	[Shalev-Shwartz et al., 2007]	Kernel	Primal SGD/SGP	Online-style	$\tilde{O}(s/\lambda\rho)$
LibLinear	[Hsieh et al., 2008]	Linear	Dual coordinate descent	Batch	$O(nd \cdot \log(1/\rho))$
SGD-QN	[Bordes and Bottou, 2008]	Linear	Primal 2SGD	Online-style	Unknown
FOLOS	[Duchi and Singer, 2008]	Linear	Primal SGP	Online-style	$\tilde{O}(s/\lambda\rho)$
BMRM	[Smola et al., 2007]	Linear	Dual QP	Batch	$O(d/\lambda\rho)$
OCAS	[Franc and Sonnenburg, 2008]	Linear	Primal QP	Batch	$O(nd)$

➢支持向量回归机

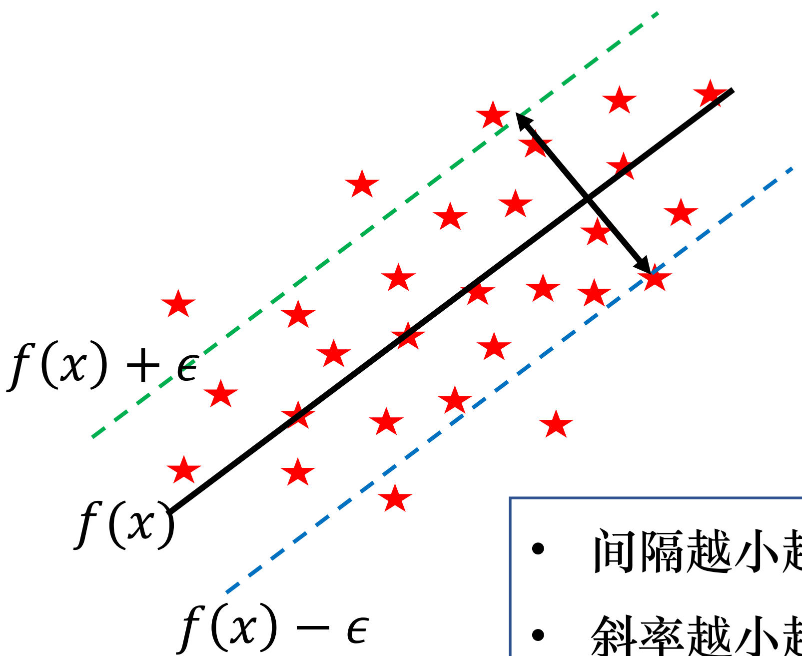
三、支持向量回归机（SVR）

1. 线性情况

给定训练集$T=\{(x_1,y_1),(x_2,y_2),\dots,(x_l,y_l)\}$，其中$x_i\in R^n$，$y_i\in R$

主要思想： 构建尽可能窄的ϵ带子，使的最终拟合线落在中间

几何解释：



- 间隔越小越好
- 斜率越小越好
- 使的所有样本点尽可能在两条虚线中间

三、支持向量回归机（SVR）

1. 线性情况

原问题：

$$\min_{w,b,\xi_i,\xi_i^*} \quad \frac{1}{2}||w||^2 + c\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

$$s.t.\ \ y_i - ({x_i}^T w + b) \le \epsilon + \xi_i,$$

$$({x_i}^T w + b) - y_i \le \epsilon + \xi_i^*,$$

$$\xi_i \ge 0, \xi_i^* \ge 0, i = 1,2,\dots,n.$$

对偶模型：

$$\min_{\alpha} \frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(x_i \cdot x_j) + \varepsilon\sum_{i=1}^{l}(\alpha_i^* + \alpha_i) - \sum_{i=1}^{l} y_i(\alpha_i^* - \alpha_i)$$

$$s.t. \quad \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \qquad 0 \le \alpha_i^{(*)} \le C\ , i = 1,\dots,l.$$

预测： $\mathrm{y} = \mathrm{g}(\mathrm{x}) = \sum_{i=1}^{l}(\bar{\alpha}_i^* - \bar{\alpha}_i)(x_i \cdot x) + \bar{b}$

1. 线性情况

引入拉格朗日函数：

$$L\left(w,b,\xi^{(*)},\alpha^{(*)},\eta^{(*)}\right)=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}(\xi_i+\xi_i^*)-\sum_{i=1}^{l}(\beta_i\xi_i+\beta_i^*\xi_i^*)$$

$$-\sum_{i=1}^{l}\alpha_i(\varepsilon+\xi_i+y_i-(w\cdot x_i)-b)-\sum_{i=1}^{l}\alpha_i^*(\varepsilon+\xi_i^*-y_i+(w\cdot x_i)+b)$$

最优性条件：

① $\frac{\partial L}{\partial w}=w+\sum_{i=1}^{l}\alpha_i x_i-\sum_{i=1}^{l}\alpha_i^* x_i=0$

② $\frac{\partial L}{\partial b}=\sum_{i=1}^{l}\alpha_i-\sum_{i=1}^{l}\alpha_i^*=0$

③ $\frac{\partial L}{\partial \xi_i}=C-\alpha_i-\beta_i=0$

④ $\frac{\partial L}{\partial \xi_i^*}=C-\alpha_i^*-\beta_i^*=0$

⑤ $\alpha_i(\varepsilon+\xi_i+y_i-(w\cdot x_i)-b)=0$

⑥ $\alpha_i^*(\varepsilon+\xi_i^*-y_i+(w\cdot x_i)+b)=0$

⑦ $\beta_i\xi_i=0$

⑧ $\beta_i^*\xi_i^*=0$

⑨ $(w\cdot x_i)+b-y_i\le\varepsilon+\xi_i$

⑩ $y_i-(w\cdot x_i)-b\le\varepsilon+\xi_i^*$

中国农业大学 China Agricultural University

1. 线性情况

对偶问题的解与样本位置的关系：

$$\alpha_i = 0 \xrightarrow{③} \beta_i = C \xrightarrow{⑦} \xi_i = 0 \xrightarrow{⑨} (w \cdot x_i) + b - y_i \le \varepsilon$$

$$\alpha_i^* = 0 \xrightarrow{④} \beta_i^* = C \xrightarrow{⑧} \xi_i^* = 0 \xrightarrow{⑩} y_i - (w \cdot x_i) - b \le \varepsilon$$

$$\alpha_i = C \xrightarrow{③} \beta_i = 0 \xrightarrow{⑦} \xi_i \ge 0 \xrightarrow{⑤} (w \cdot x_i) + b - y_i \ge \varepsilon$$

$$\alpha_i^* = C \xrightarrow{④} \beta_i^* = 0 \xrightarrow{⑧} \xi_i^* \ge 0 \xrightarrow{⑥} y_i - (w \cdot x_i) - b \ge \varepsilon$$

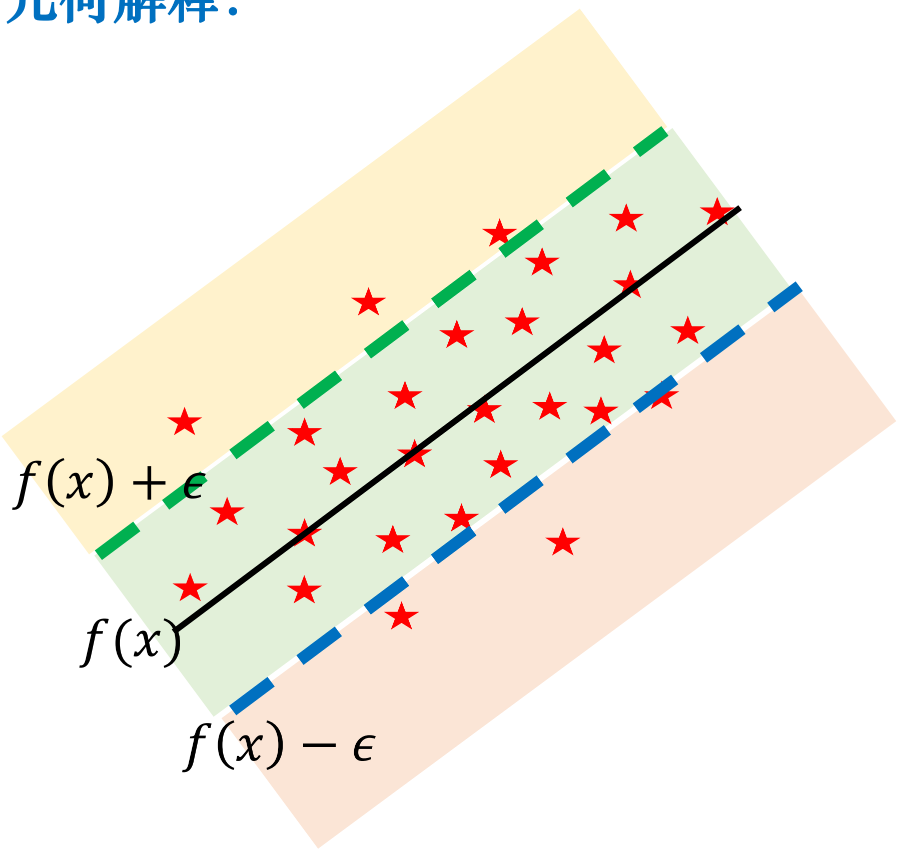
$$0 < \alpha_i < C \xrightarrow{③} \beta_i \ne 0 \xrightarrow{⑦} \xi_i = 0 \xrightarrow{⑤} (w \cdot x_i) + b - y_i = \varepsilon$$

$$0 < \alpha_i^* < C \xrightarrow{④} \beta_i^* \ne 0 \xrightarrow{⑧} \xi_i^* = 0 \xrightarrow{⑥} y_i - (w \cdot x_i) - b = \varepsilon$$

三、支持向量回归机（SVR）

1. 线性情况

几何解释：



$\alpha_i = 0,\ \alpha_i^* = C$

$\alpha_i = 0,\ 0 \le \alpha_i^* \le C$

$\alpha_i = 0,\ \alpha_i^* = 0$

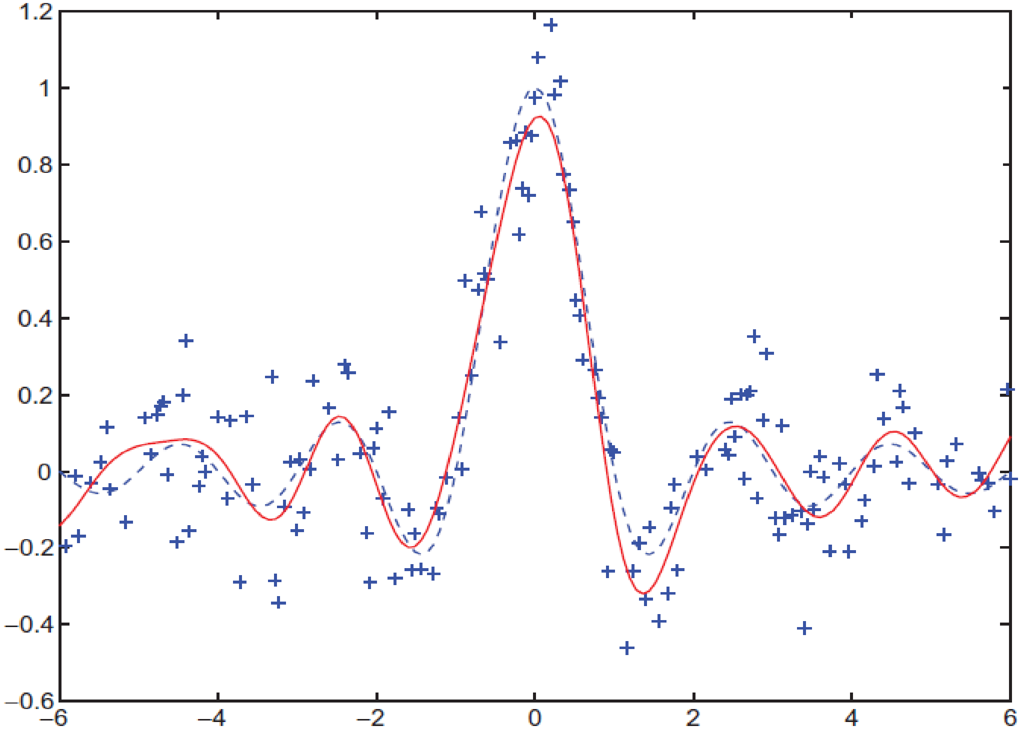
$0 \le \alpha_i \le C,\ \alpha_i^* = 0$

$\alpha_i = C,\ \alpha_i^* = 0$

三、支持向量回归机（SVR）

2. 非线性情况

线性不可分：



2. 非线性情况

原问题：

$$\min_{w,b,\xi_i,\xi_i^*} \quad \frac{1}{2}||w||^2 + c\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

$$s.t.\ \ y_i - (\boxed{\varphi(x_i)}^T w + b) \le \epsilon + \xi_i,$$

$$(\boxed{\varphi(x_i)}^T w + b) - y_i \le \epsilon + \xi_i^*,$$

$$\xi_i \ge 0, \xi_i^* \ge 0, i = 1,2,\dots,n.$$

对偶模型：

$$\min_{\alpha} \ \frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)\boxed{K(x_i \cdot x_j)} + \varepsilon\sum_{i=1}^{l}(\alpha_i^* + \alpha_i) - \sum_{i=1}^{l} y_i(\alpha_i^* - \alpha_i)$$

$$s.t. \quad \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \qquad 0 \le \alpha_i^{(*)} \le C, i = 1,\dots,l.$$

预测： $\mathrm{y} = \mathrm{g(x)} = \sum_{i=1}^{l}(\bar{\alpha}_i^* - \bar{\alpha}_i)K(x_i \cdot x) + \bar{b}$

三、支持向量回归机（SVR）

3. 拓展模型

① **最小二乘支持向量回归机（LS-SVR）**

将SVR模型中的不等式约束改为等式约束

① **ν-支持向量回归机（ν-SVR）**

用参数ν代替原始SVM中的参数C，使参数具有一定的含义

① **双子支持向量回归机（TSVR）**

通过构造两条非平行的ϵ超平面，得到最终的拟合超平面

➢非平衡类支持向量机

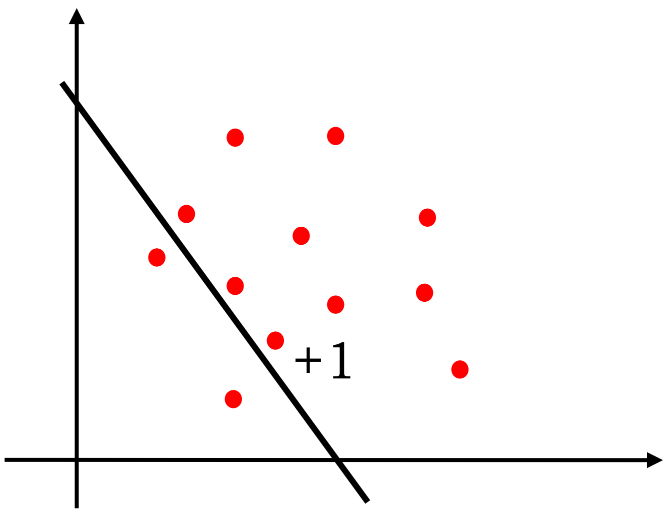
四、非平衡类支持向量机

非平衡类支持向量机是处理非平衡数据的一种有效的手段

1. One-class SVM

主要思想： 只有一类训练点时，我们只利用这一类点，构造最优超平面

几何解释：



① 所有样本点尽可能远离坐标原点

② 超平面尽可能远离坐标原点

中国农业大学 China Agricultural University

1. One-class SVM

原问题:

$$\min_{w,\xi,\rho} \frac{1}{2}\|w\|^2 - \rho + \frac{1}{\nu l}\sum_{i=1}^{l}\xi_i$$

$$s.t.\quad y_i\big(w\cdot\phi(x_i)\big) \geq \rho - \xi_i\,,\quad i = 1,\dots,l$$

$$\xi_i \geq 0, i = 1,\dots,l;$$

对偶问题:

$$\min_{\alpha}\ \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(x_i,x_j)\alpha_i\alpha_j$$

$$s.t.\quad \sum_{i=1}^{l}\alpha_i = 1$$

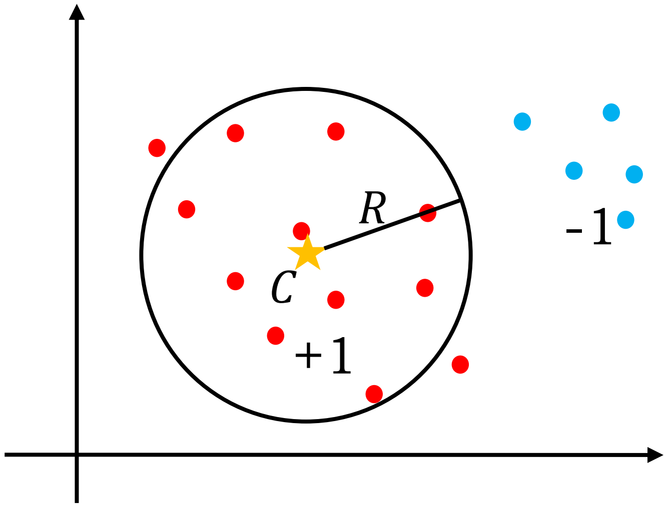
$$0 \leq \alpha_i \leq \frac{1}{\nu l},\qquad i = 1,\dots,l.$$

2. 球类支持向量机

① SVDD

主要思想： 只利用数量较多的那一类，构造最优超平面

几何解释：



① 正类样本点尽可能在圆内

② 负类样本点尽可能在圆外

四、非平衡类支持向量机

① SVDD

原问题:

$$\min_{R^2,C,\xi} \quad R^2+\frac{1}{v_1 l^+}\sum_{i=1}^{l^+}\xi_i+\frac{1}{v_2 l^-}\sum_{j=1}^{l^-}\xi_j$$

$$s.t. \quad ||\phi(x_i)-C||^2 \le R^2+\xi_i$$

$$||\phi(x_j)-C||^2 \ge R^2-\xi_j$$

$$\xi_k \ge 0$$

对偶问题:

$$\min_{\alpha} \quad \alpha^T\begin{bmatrix} K(A,A) & -K(A,B) \\ -K(B,A) & K(B,B) \end{bmatrix}\alpha-\alpha^T\begin{bmatrix} K_{AA} \\ -K_{BB} \end{bmatrix}$$

$$s.t. \quad \sum_{i=1}^{l^+}\alpha_i-\sum_{j=1}^{l^-}\beta_j=1$$

$$0\le\alpha_i\le\frac{1}{v_1 l^+}, 0\le\beta_j\le\frac{1}{v_2 l^-}$$

决策函数:

$$f(x)=\begin{cases} 1, & if||\phi(x)-C||^2\le R^2 \\ -1, & otherwise \end{cases}$$

四、非平衡类支持向量机

① SVDD

球心C和半径R的求解:

构造拉格朗日函数:

$$L = R^2 + \frac{1}{v_1 l^+}\sum_{i=1}^{l^+}\xi_i + \frac{1}{v_2 l^-}\sum_{j=1}^{l^-}\xi_j + \sum_{i=1}^{l^+}\alpha_i(||\phi(x_i) - C||^2 - R^2 - \xi_i)$$
$$-\sum_{j=1}^{l^-}\beta_j(||\phi(x_j) - C||^2 - R^2 + \xi_j) - \sum_{k=1}^{n}\lambda_k\xi_k$$

关于参数R^2, C, ξ_i求导数:

$$\frac{\partial L}{\partial R^2} = 1 - \sum_{i=1}^{l^+}\alpha_i + \sum_{j=1}^{l^-}\beta_j \Rightarrow \sum_{i=1}^{l^+}\alpha_i + \sum_{j=1}^{l^-}\beta_j = 1$$

$$\frac{\partial L}{\partial C} = -2\sum_{i=1}^{l^+}\alpha_i(\phi(x_i) - C) + 2\sum_{j=1}^{l^-}\beta_j\left(\phi(x_j) - C\right) = 0$$

中国农业大学 China Agricultural University

① SVDD

球心C和半径R的求解：

$$\frac{\partial L}{\partial \xi_i}=\frac{1}{v_1 l^+}-\alpha_i-\lambda_i=0 \qquad \alpha_i(||\phi(x_i)-C||^2-R^2-\xi_i)=0$$

$$\frac{\partial L}{\partial \xi_j}=\frac{1}{v_2 l^-}-\beta_j-\lambda_j=0 \qquad \beta_j\left(||\phi(x_j)-C||^2-R^2+\xi_j\right)=0$$

可知：当$0<\alpha_i<\frac{1}{v_1 l^+}$，有$R^2=||\phi(x_i-C)||^2$

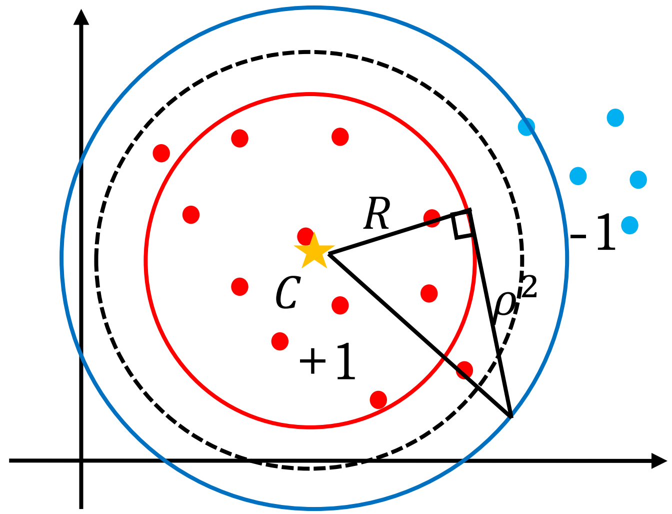
当$0<\beta_j<\frac{1}{v_2 l^-}$，有$R^2=||\phi(x_j-C)||^2$

中国农业大学 China Agricultural University

② 最小球最大间隔法（SSLM）

主要思想： 同时利用正负类样本点，构造两个同心球超平面

几何解释：



① 正类样本点尽可能在红色圆内

② 负类样本点尽可能在蓝色圆外

③ 两球间隔尽量大

中国农业大学 China Agricultural University

② 最小球最大间隔法（SSLM）

原问题：

$$\min_{R^2,C,\rho^2,\xi} \quad R^2 - \nu\rho^2 + \frac{1}{v_1 l^+}\sum_{i=1}^{l^+}\xi_i + \frac{1}{v_2 l^-}\sum_{j=1}^{l^-}\xi_j$$

$$s.t. \quad ||\phi(x_i) - C||^2 \le R^2 + \xi_i, i \in I^+$$

$$||\phi(x_j) - C||^2 \ge R^2 + \rho^2 - \xi_j, j \in I^-$$

$$\xi_k \ge 0, \mathrm{k} = 1,2,\cdots,l$$

对偶问题：

$$\min_{\alpha} \quad \alpha^T \begin{bmatrix} K(A,A) & -K(A,B) \\ -K(B,A) & K(B,B) \end{bmatrix} \alpha - \alpha^T \begin{bmatrix} K_{AA} \\ -K_{BB} \end{bmatrix}$$

$$s.t. \begin{cases} \alpha^T y = 1 \\ \alpha^T e = 2\nu + 1 \end{cases} \Leftrightarrow \alpha^T [y \quad e] = [1 \quad 2\nu + 1]$$

$$0 \le \alpha \le \begin{bmatrix} \frac{1}{v_1 l^+} \\ \frac{1}{v_2 l^1} \end{bmatrix}$$

决策函数：

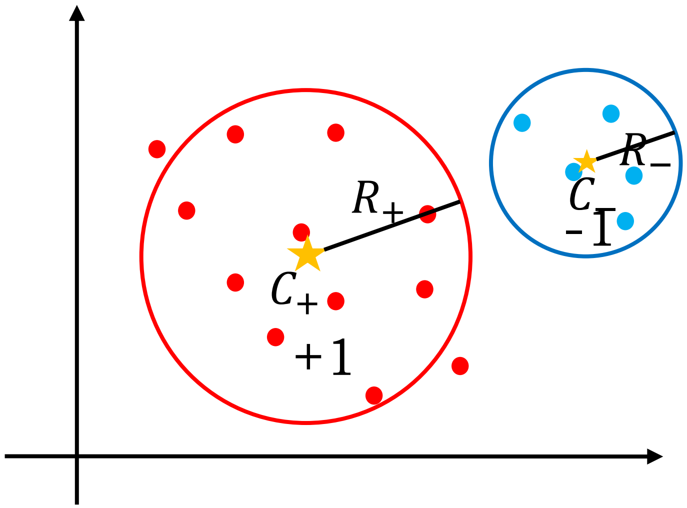
$$f(x_t) = \begin{cases} 1, & if \ ||\phi(x_i) - C|| < \frac{R + \sqrt{R^2 + \rho^2}}{2} \\ -1, & else \end{cases}$$

四、非平衡类支持向量机

③ 双子球支持向量机

主要思想： 同时利用正负类样本点，构造两个球超平面

几何解释：



① 正类样本点尽可能在红色圆内

② 负类样本点尽可能在蓝色圆内

四、非平衡类支持向量机

③ 双子球支持向量机

原问题：

$$\min_{C_+,R_+^2,\xi_i} \quad R_+^2 - \frac{\nu_1}{l^-}\sum_{j\in I^-}||\phi(x_j)-C_+||^2 + \frac{c_1}{l^+}\sum_{i\in I^+}\xi_i$$

$$s.t. \quad ||\phi(x_i)-C_+||^2 \le R_+^2 + \xi_i$$

$$R_+^2 \ge 0, \xi_i \ge 0, i \in I^+$$

$$\min_{C_-,R_-^2,\xi_j} \quad R_-^2 - \frac{\nu_2}{l^+}\sum_{i\in I^+}||\phi(x_i)-C_-||^2 + \frac{c_2}{l^-}\sum_{j\in I^-}\xi_j$$

$$s.t. \quad ||\phi(x_j)-C_-||^2 \le R_-^2 + \xi_j$$

$$R_-^2 \ge 0, \xi_j \ge 0, j \in I^-$$

决策函数：

$$f(\boldsymbol{x}) = arg\min_{+,-}\left\{\frac{||\phi(\boldsymbol{x})-C_+||^2}{R_+^2}, \frac{||\phi(\boldsymbol{x})-C_-||^2}{R_-^2}\right\}$$

（距离哪个球近就属于哪一类）

➢ 总结

五、总结

1. **支持向量分类机**

   线性SVM、非线性SVM，拓展模型，求解算法

2. **支持向量回归机**

   线性SVR、非线性SVR，拓展模型

3. **球类支持向量机**

   OCSVM，SVDD、SSLM、THSVM

五、总结

优点：

① 小样本：并不是说样本的绝对数量少，而是说与问题的复杂度比起来，SVM算法要求的样本数是相对比较小的

② 非线性：SVM擅长应付样本数据线性不可分的情况，主要通过松弛变量（也有人叫惩罚变量）和核函数技术来实现

③ 高维模式识别：指样本维数很高，SVM可以处理高维向量，因为SVM产生的分类器很简洁，用到的样本信息很少（仅仅用到那些称之为“支持向量”的样本）

缺点：

① 如果特征维度远远大于样本数，则SVM表现一般

② SVM在样本量非常大，计算量过大

③ 难以选择一个合适的核函数

④ SVM对缺失数据敏感

谢　谢！