河北师范大学软件学院
Software College of Hebei Normal University

分类模型

第一部分 线性分类模型

第二部分 非线性分类模型

SVM

张朝晖

2018-2019学年 20180918-0926

1.分类模块

产生式分类模型 A.贝叶斯分类模型

判别式分类模型

- **线性分类模型**
  - B. Fisher判别分类
  - C. 感知器分类模型
  - D. 大间隔分类模型(线性SVM)
- **非线性分类模型**
  - E. 核SVM (非线性SVM)
  - F. 核Fisher判别分类
  - G. 神经网络

其它分类模型

- H.KNN分类模型
- I.决策树分类模型
- J.$Logistic$回归
- K.$Softmax$回归

2.聚类模块

- L.K-均值聚类
- M.高斯混合聚类
- N.DBSCAN聚类
- O.层次聚类

3.回归模块

- P.KNN回归
- Q.回归树
- R.最小二乘线性回归
- S.岭回归
- T.LASSO回归

4.集成学习

- U.$Bagging$
- V.随机森林
- W.$Boosting$

5.特征工程

- X.主成分分析(PCA)
- ...

6.评价模块 混淆矩阵(及其相关指标)、ROC曲线、交叉验证

主要内容

A.引言

B. Fisher线性判别分析

C.感知器

<u>D. 线性支持向量机(线性SVM)</u>

Support Vector Machine: SVM

学习要求：

- 理解SVM分类模型构建的基本思想
- 掌握SVM 目标函数的构造形式、意义
- 掌握SVM分类模型的学习步骤、使用流程
- 掌握SVM进行超参数寻优的典型方法
- 能够熟练应用SVM进行分类(两类别、多类别)

关键词：

- 分类间隔、分类边界、判别函数
- 训练样本集的错分程度
- 分类超平面、超平面法向量、超参数
- 支持向量、非支持向量
- 判别函数、分类边界

- SVM(Support Vector Machine)是支持向量机的简称。“机”,对应机器学习中的“机器”，是“算法”的意思。

  所谓支持向量机就是一种与“支持向量”有关的机器学习算法。

- SVM 的求解最后转化成凸二次规划问题的求解，因此SVM 的解是全局唯一的最优解.

- SVM在解决小样本(样本数<特征维数)、非线性、高维模式分类问题中表现出许多特有的优势，并能够推广应用到函数拟合(回归)等其它机器学习问题中.

- SVM 用于分类，就是*SVC (Support Vector Classification)*；用于回归就是*SVR (Support Vector Regression)*。

第1部分

两类别分类情况下的线性SVC

主要内容

1 两类别分类问题：线性可分情况

2 两类别分类问题：近似线性可分情况

3 多类别的线性SVM

4 应用举例

SVM始于两类问题的解决需求

两类别分类问题的描述：

给定训练集$\boldsymbol{T}=\{(\boldsymbol{x}_1,y_1),\cdots,(\boldsymbol{x}_N,y_N)\}$

$$\text{其中}\begin{cases}\boldsymbol{x}_i\in R^d \\ y_i\in\mathcal{Y}=\{1,-1\}\end{cases}\qquad i=1,2,\ldots,N$$

要求：寻找R^d空间上的一个实值函数$g(\boldsymbol{x})$, 从而可用决策函数 $f(\boldsymbol{x})=\mathrm{sgn}(g(\boldsymbol{x})): R^d\to\mathcal{Y}$;

推断任意输入$\boldsymbol{x}$对应的输出值y.

线性可分问题-定义：

给定**训练集**$\boldsymbol{T}=\left\{\left(\boldsymbol{x}_1, y_1\right),\cdots,\left(\boldsymbol{x}_N, y_N\right)\right\}$

其中 $\boldsymbol{x}_i \in R^d, i=1,2,\ldots,N$

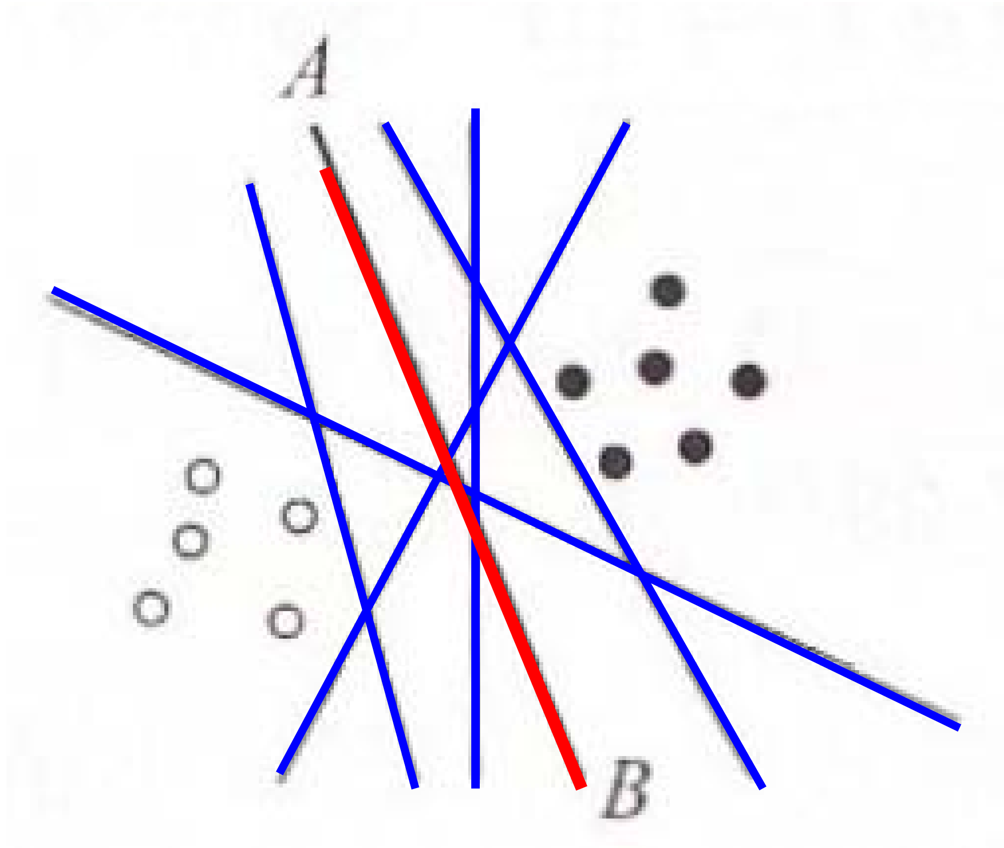
$y_i \in \mathcal{Y}=\{1,-1\}$

若存在$\omega \in R^d$，$b \in R$,以及正数ε，使得

$$\begin{cases} 对于所有y_i=+1的样本\boldsymbol{x}_i，有g\left(\boldsymbol{x}_i\right)=\omega \bullet \boldsymbol{x}_i+b \geq \varepsilon \\ 对于所有y_i=-1的样本\boldsymbol{x}_i，有g\left(\boldsymbol{x}_i\right)=\omega \bullet \boldsymbol{x}_i+b \leq -\varepsilon. \end{cases}$$

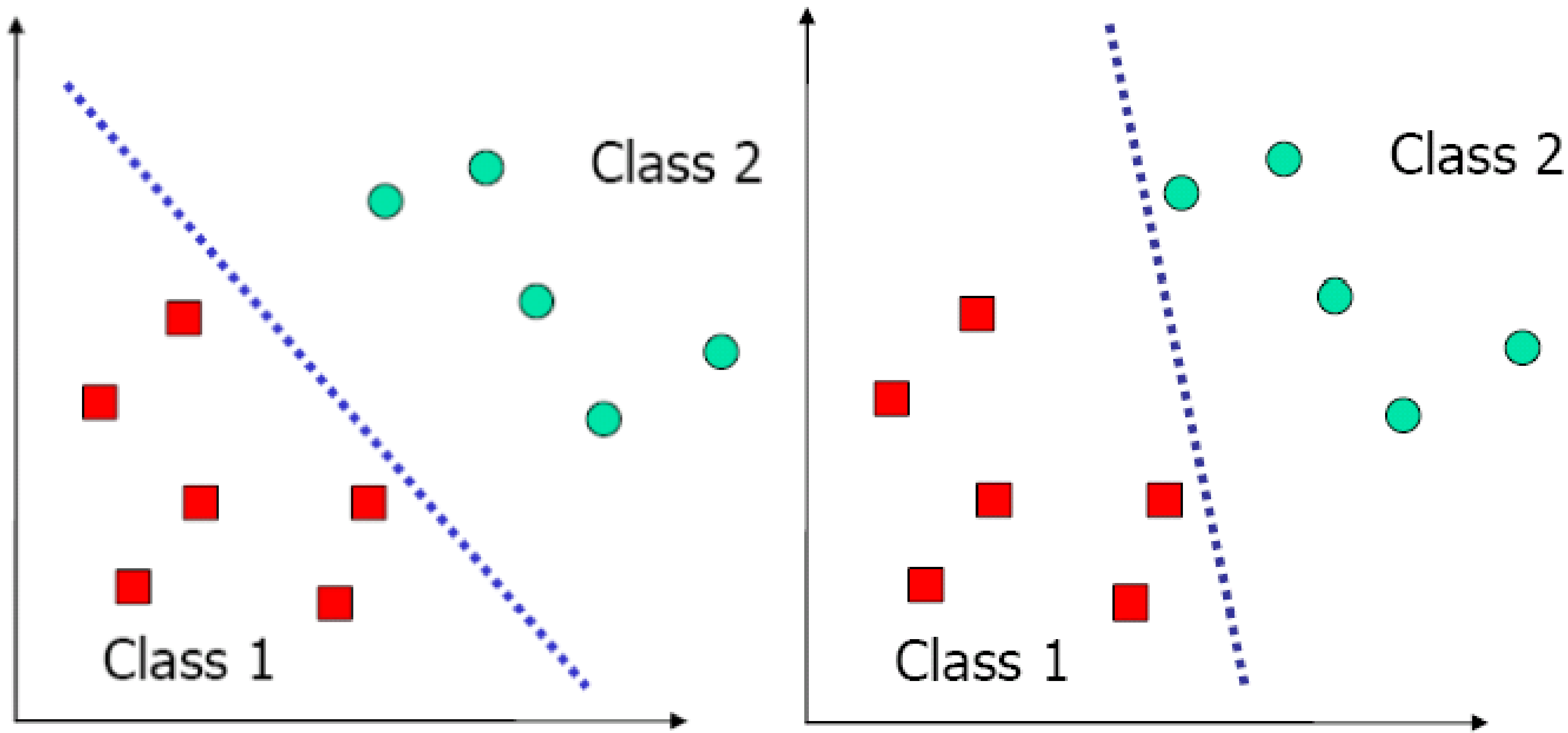
则称**训练集$\boldsymbol{T}$**线性可分。

1. 例：二维、两类、线性可分情况



问题：可能存在许多分类边界（决策边界），若将两类完全分开，应当选择哪一个?

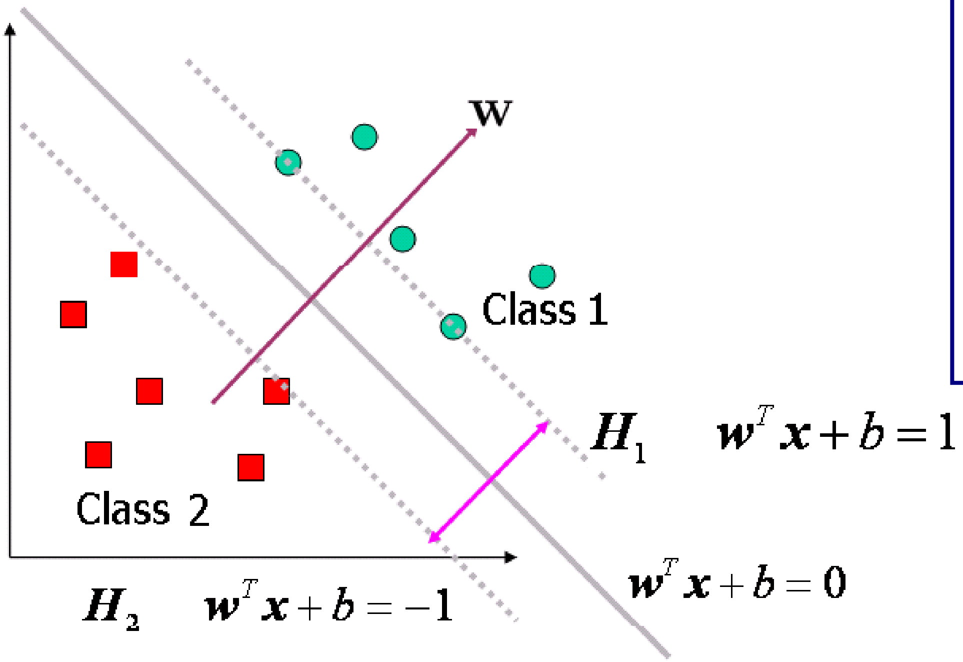
➢ 较差的分类边界(如图)



➢ 较好的分类边界：

分类间隔（margin）尽可能大；

分类边界尽可能远离两类样本数据



- **最优分类边界：**

能将两类无错误分开；

分类间隔(margin)最大

- H -----最优分类线；
- H_1, H_2----平行于H、且经过各类训练样本集中离H最近的训练样本
- 分类间隔(margin)---- H_1, H_2之间的距离
- 推广到高维空间：最优边界就变为最优超平面(*optimal hyperplane*)

如何找到最优分类边界?

分类边界 $\boldsymbol{w} \cdot \boldsymbol{x} + b = 0$

判别函数 $g(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x} + b$

2. 线性SVM的分类间隔

线性可分训练样本集 $\{\boldsymbol{x}_i, y_i\} \quad i=1,\ldots,N \qquad \boldsymbol{x}_i \in R^d$

类别标号 $y_i \in \{-1,1\}$

d维空间的线性判别函数 $g(\boldsymbol{x}) = \boldsymbol{w}\cdot\boldsymbol{x} + b$

判别函数归一化 $\begin{cases} \text{两类所有样本满足} & |g(\boldsymbol{x}_i)| \geq 1 \\ \text{离分类面最近样本} & |g(\boldsymbol{x}_i)| = 1 \end{cases}$

训练样本线性可分 $\begin{cases} g(\boldsymbol{x}_i) \geq 1 & \text{对于} y_i = 1 \\ g(\boldsymbol{x}_i) \leq -1 & \text{对于} y_i = -1 \end{cases}$

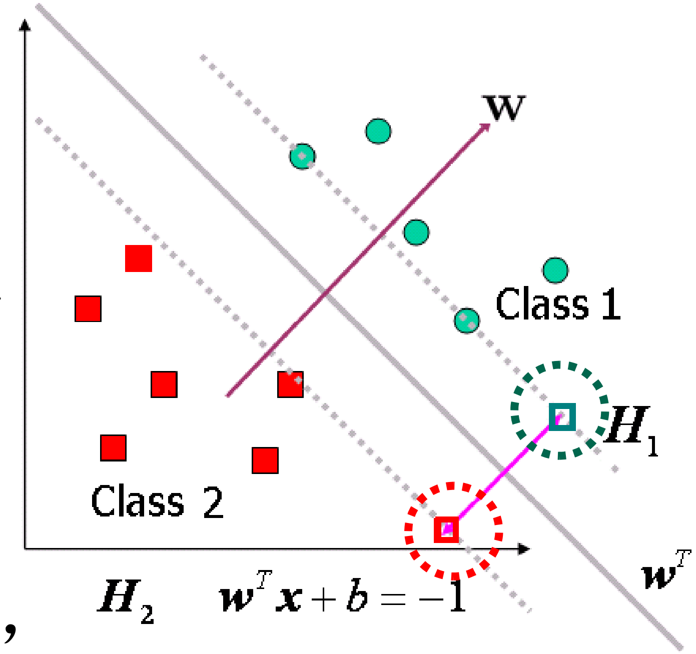
$$\Rightarrow \quad y_i \cdot g(\boldsymbol{x}_i) \geq 1$$

$$\Rightarrow y_i \cdot [\boldsymbol{w}\cdot\boldsymbol{x}_i + b] - 1 \geq 0$$

最优分类面H：$\boldsymbol{w}\cdot\boldsymbol{x}+b=0$

正平面H_1:　　$\boldsymbol{w}\cdot\boldsymbol{x}+b=1$

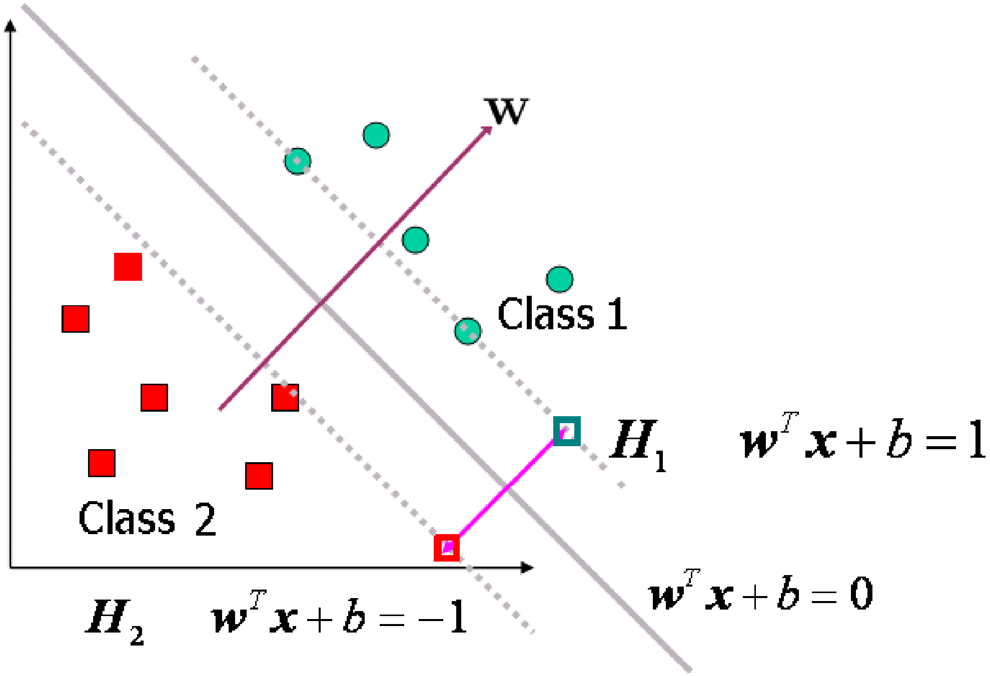
负平面H_2:　　$\boldsymbol{w}\cdot\boldsymbol{x}+b=-1$



设$\begin{cases}\boldsymbol{x}^+\text{为正平面}H_1\text{上任意点，}\\ \boldsymbol{x}^-\text{为负平面}H_2\text{上最接近}\boldsymbol{x}^+\text{的点，}\boldsymbol{x}^+=\boldsymbol{x}^-+\lambda\boldsymbol{w}\end{cases}$

则$\begin{cases}\text{正平面上}\quad \boldsymbol{w}\cdot\boldsymbol{x}^++b=1\\ \text{负平面上}\quad \boldsymbol{w}\cdot\boldsymbol{x}^-+b=-1\\ \text{两平面间隔}\quad margin=\left\|\boldsymbol{x}^+-\boldsymbol{x}^-\right\|=\left\|\lambda\boldsymbol{w}\right\|=\lambda\left\|\boldsymbol{w}\right\|\end{cases}$

$$\begin{cases} \text{正平面上：} \ \mathbf{w}\cdot\mathbf{x}^{+}+b=1 \\ \text{负平面上：} \ \mathbf{w}\cdot\mathbf{x}^{-}+b=-1 \\ \mathbf{x}^{+}=\mathbf{x}^{-}+\lambda\mathbf{w} \end{cases}$$



$$\rightarrow \mathbf{w}\cdot\left(\mathbf{x}^{-}+\lambda\mathbf{w}\right)+b=1$$

$$\rightarrow \mathbf{w}\cdot\mathbf{x}^{-}+b+\lambda\mathbf{w}\cdot\mathbf{w}=1$$

$$\rightarrow -1+\lambda\mathbf{w}\cdot\mathbf{w}=1$$

$$\rightarrow \lambda=\frac{2}{\mathbf{w}\cdot\mathbf{w}}=\frac{2}{\|\mathbf{w}\|^{2}}$$

$$\boxed{\text{分类间隔} \ \ margin=\left\|\mathbf{x}^{+}-\mathbf{x}^{-}\right\|=\|\lambda\mathbf{w}\|=\frac{2}{\|\mathbf{w}\|}}$$

3. 线性SVM—目标函数的构造

最优分类边界

$$
\begin{cases}
\text{分类间隔最大}\quad margin = \dfrac{2}{\|\mathbf{w}\|} & \Leftrightarrow \quad \text{使}\|\mathbf{w}\|\left(\text{或}\|\mathbf{w}\|^2\right)\text{最小} \\
\text{训练样本线性可分（均正确分类）} & \Leftrightarrow \quad y_i \cdot \left(\mathbf{w}\cdot\mathbf{x}_i + b\right) - 1 \geq 0 \\
 & \quad\; i = 1, 2, \ldots, N
\end{cases}
$$

$$\phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 = \frac{1}{2}(\mathbf{w}\cdot\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

原始问题的目标函数(线性可分问题的最大间隔法)

约束优化问题(凸二次规划)

如何求解?

$$
\begin{cases}
\min\limits_{\mathbf{w},b} \quad \phi(\mathbf{w}) = \dfrac{1}{2}(\mathbf{w}\cdot\mathbf{w}) \\
s.t. \quad y_i \cdot \left(\mathbf{w}\cdot\mathbf{x}_i + b\right) - 1 \geq 0 \quad i = 1, 2, \ldots, n
\end{cases}
$$

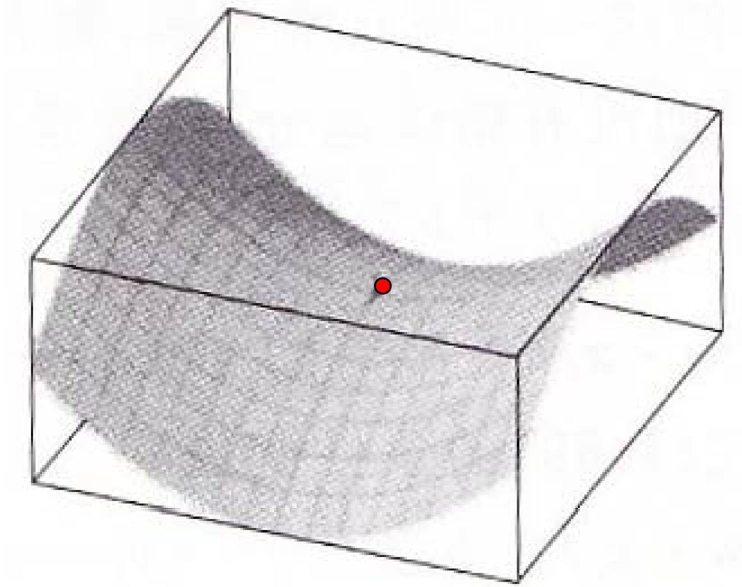
4. 线性SVM—目标函数的求解

构造***Lagrange***辅助目标函数，将**原始问题**转化为**对偶问题**.

$$L(\mathbf{w},b,\alpha)=\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})-\sum_{i=1}^{N}\alpha_i\left\{y_i\cdot\left[(\mathbf{w}\cdot\mathbf{x}_i)+b\right]-1\right\}$$

其中非负*Lagrange*乘子向量$\boldsymbol{\alpha}=\left[\alpha_1 \quad \cdots \quad \alpha_N\right]^T$

显然：$L(\mathbf{w},b,\alpha)\le\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})$



原始问题等价为：

$$\min_{\mathbf{w},b}\max_{\alpha}L(\mathbf{w},b,\alpha)=\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})-\sum_{i=1}^{N}\alpha_i\left\{y_i\cdot\left[(\mathbf{w}\cdot\mathbf{x}_i)+b\right]-1\right\}$$

$$L(\mathbf{w},b,\boldsymbol{\alpha})=\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})-\sum_{i=1}^{N}\alpha_i\left\{y_i\cdot\left[(\mathbf{w}\cdot\mathbf{x}_i)+b\right]-1\right\}$$

$$=\frac{1}{2}\left(\mathbf{w}^T\mathbf{w}\right)-\sum_{i=1}^{N}\alpha_i y_i\mathbf{w}^T\mathbf{x}_i-b\sum_{i=1}^{N}\alpha_i y_i+\sum_{i=1}^{N}\alpha_i$$

$L(\mathbf{w},b,\alpha)$在最优解处满足：

$$\begin{cases}\text{条件 1} & \dfrac{\partial L(\mathbf{w},b,\alpha)}{\partial \mathbf{w}}=0 \quad\rightarrow\quad \boxed{\mathbf{w}=\sum_{i=1}^{N}\alpha_i y_i\mathbf{x}_i} \\ \text{条件2} & \dfrac{\partial L(\mathbf{w},b,\alpha)}{\partial b}=0 \quad\rightarrow\quad \boxed{\sum_{i=1}^{N}\alpha_i y_i=0}\end{cases}$$

该条件下，$L(\boldsymbol{w},b,\alpha)$为

$$Q(\alpha)=\sum_{i=1}^{N}\alpha_i+\frac{1}{2}\left(\boldsymbol{w}^T\boldsymbol{w}\right)-\boldsymbol{w}^T\sum_{i=1}^{N}\alpha_i y_i \boldsymbol{x}_i$$

$$=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\boldsymbol{w}^T\sum_{i=1}^{N}\alpha_i y_i \boldsymbol{x}_i$$

$$=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\left(\sum_{j=1}^{N}\alpha_j y_j \boldsymbol{x}_j\right)^T\sum_{i=1}^{N}\alpha_i y_i \boldsymbol{x}_i$$

$$=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\left(\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_j{}^T\boldsymbol{x}_i\right)$$

$$=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_i\alpha_j y_i y_j\left(\boldsymbol{x}_j\cdot\boldsymbol{x}_i\right)$$

对偶问题

$$\max_{\boldsymbol{\alpha}}. \quad Q(\boldsymbol{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{N} \alpha_i \alpha_j y_i y_j \left(\boldsymbol{x}_j \cdot \boldsymbol{x}_i\right)$$

$$s.t. \quad \begin{cases} (1) \quad \sum_{i=1}^{N} \alpha_i y_i = 0 \\ (2) \quad \alpha_i \geq 0 \qquad \text{对于} i = 1, \ldots, N \end{cases}$$

等价形式

对偶问题

$$\min_{\boldsymbol{\alpha}}. \quad W(\boldsymbol{\alpha}) = \frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{N} \alpha_i \alpha_j y_i y_j \left(\boldsymbol{x}_j \cdot \boldsymbol{x}_i\right) - \sum_{i=1}^{N} \alpha_i$$

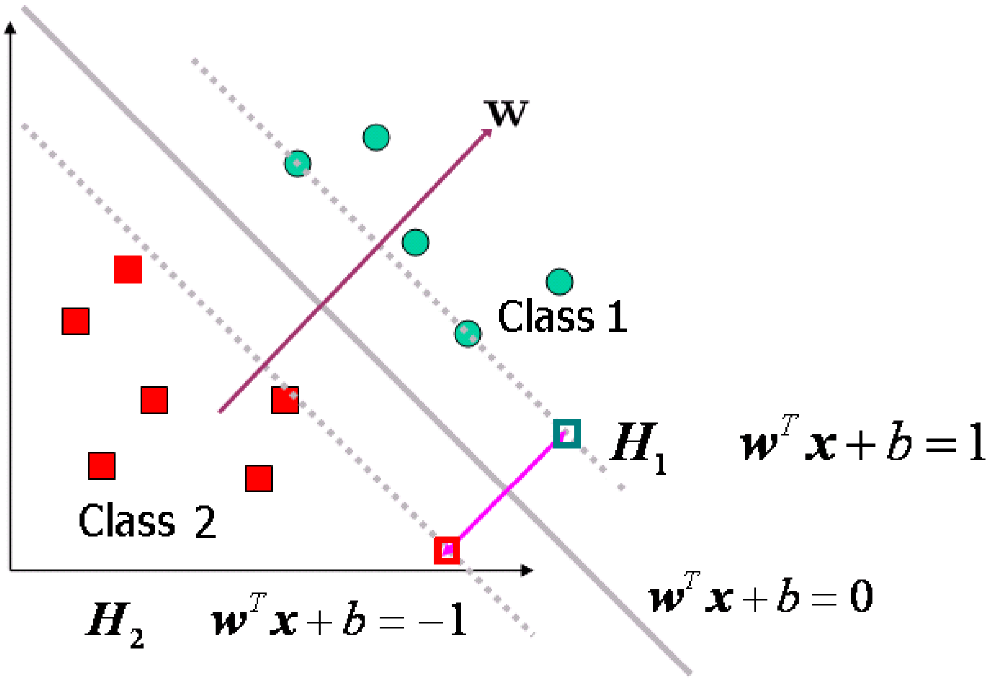
$$s.t. \quad \begin{cases} (1) \quad \sum_{i=1}^{N} \alpha_i y_i = 0 \\ (2) \quad -\alpha_i \leq 0 \qquad \text{对于} i = 1, \ldots, N \end{cases}$$

解得**非负**乘子向量的最优解：$\boldsymbol{\alpha}^* = \left[\alpha_1^* \quad \cdots \quad \alpha_N^*\right]^T$

线性SVM中，判别函数的最优参数w^*、b^*的确定：

(1) $\boldsymbol{w}^* = \sum_{i=1}^{N} \alpha_i^* y_i \boldsymbol{x}_i$

$\boldsymbol{\alpha}^* = \begin{bmatrix} \alpha_1^* & \cdots & \alpha_N^* \end{bmatrix}^T$ 中多数分量为0



对$\boldsymbol{w}^*$有贡献的只有占少数的非零α_i^*

相应地，对应非零α_i^*的训练样本就称为**支持向量**（*Suport Vector*）

作为支持向量的训练样本就位于H_1, H_2**超平面**。

线性*SVM*中，判别函数的最优参数w^*、b^*的确定：

(2) $\boldsymbol{b}^*$的确定

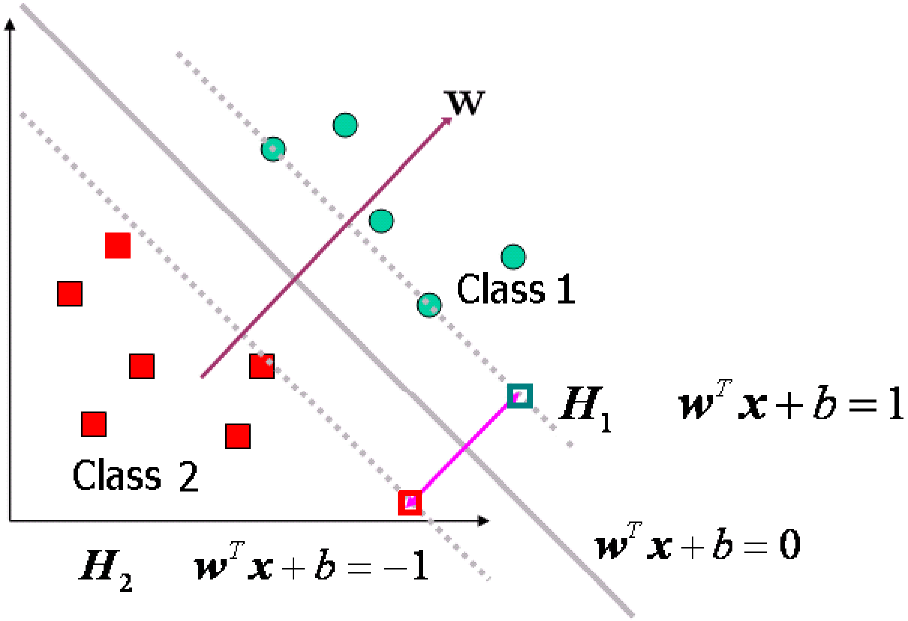
找到所有非零的α_i^*对应的支持向量，构成集合：

$$U=\left\{(\boldsymbol{x}_i,y_i)\,|\,(\boldsymbol{x}_i,y_i)\in\text{训练集}\boldsymbol{T}\text{，并且}\alpha_i^*>0\right\}$$

对于$\forall(\boldsymbol{x}_i,y_i)\in U$，必有：$y_i=\boldsymbol{w}^*\bullet\boldsymbol{x}_i+\boldsymbol{b}^*$

为了确保稳定性，取平均：

$$b^*=\frac{1}{\#U}\sum_{(\boldsymbol{x}_i,y_i)\in U}\left[y_i-\boldsymbol{w}^*\bullet\boldsymbol{x}_i\right]$$



$\#U$----集合U中元素数目(或表示为$|U|$)

原始问题的目标函数(线性可分问题的最大间隔法)

$$\begin{cases} \min\limits_{\boldsymbol{w},b} \quad \phi(\boldsymbol{w}) = \frac{1}{2}(\boldsymbol{w}\cdot\boldsymbol{w}) \\ s.t. \quad y_i \cdot (\boldsymbol{w}\cdot\boldsymbol{x}_i + b) - 1 \geq 0 \quad i = 1,2,\ldots,n \end{cases}$$

非负*Lagrange*乘子向量的最优解：$\boldsymbol{\alpha}^* = \begin{bmatrix} \alpha_1^* & \cdots & \alpha_N^* \end{bmatrix}^T$

最优分类边界 $(\boldsymbol{w}^*\cdot\boldsymbol{x}) + b^* = 0$

$$\begin{cases} \boldsymbol{w}^* = \sum_{i=1}^{N} \alpha_i^* y_i \boldsymbol{x}_i \\ b^* = \frac{1}{\#U} \sum_{(\boldsymbol{x}_i, y_i)\in U} \left[y_i - \boldsymbol{w}^* \bullet \boldsymbol{x}_i \right] \\ U = \{(\boldsymbol{x}_i, y_i) \mid (\boldsymbol{x}_i, y_i) \in \textbf{训练集}\boldsymbol{T}，\text{并且}\alpha_i^* > 0\} \end{cases}$$

判别函数 $g(\boldsymbol{x}) = (\boldsymbol{w}^*\cdot\boldsymbol{x}) + b^* = \sum_{i=1}^{N} \alpha_i^* y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}) + b$

$$f(\boldsymbol{x}) = \operatorname{sgn}\left[(\boldsymbol{w}^*\cdot\boldsymbol{x}) + b^*\right] = \operatorname{sgn}\left\{\sum_{i=1}^{N} \alpha_i^* y_i (\boldsymbol{x}_i\cdot\boldsymbol{x}) + b^*\right\}$$

5. 利用线性SVM进行分类

对于未知类别的测试样本$\boldsymbol{x}$,

若 $f(\boldsymbol{x})=\mathrm{sgn}\left[\left(\boldsymbol{w}^{*}\cdot\boldsymbol{x}\right)+b^{*}\right]=1$

则 $\boldsymbol{x}$决策为第一类；否则$\boldsymbol{x}$决策为第二类。

6.线性可分SVC算法

(1)给定训练样本集$T=\{(\boldsymbol{x}_1,y_1),\cdots(\boldsymbol{x}_N,y_N)\}$

其中$\boldsymbol{x}_i\in R^d, y_i\in\mathcal{Y}=\{1,-1\}, i=1,\ldots,N$

训练样本的预处理

(2)构造并求解凸二次规划问题

$$\begin{cases}\min\limits_{\boldsymbol{\alpha}}. & W(\boldsymbol{\alpha})=\dfrac{1}{2}\sum\limits_{j=1}^{N}\sum\limits_{i=1}^{N}\alpha_i\alpha_j y_i y_j\left(\boldsymbol{x}_j\cdot\boldsymbol{x}_i\right)-\sum\limits_{i=1}^{N}\alpha_i \\ s.t. & \begin{cases}(1) & \sum\limits_{i=1}^{N}\alpha_i y_i=0 \\ (2) & -\alpha_i\leq 0 \qquad \text{对于} i=1,\ldots,N\end{cases}\end{cases}$$

得解$\boldsymbol{\alpha}^*=\left[\alpha_1^*, \quad \cdots, \quad \alpha_N^*\right]^T$

(3) 计算 $\boldsymbol{w}^* = \sum_{j=1}^{N} \alpha_j^* y_j \boldsymbol{x}_j$

(4) 计算 $b^* = \frac{1}{\#U} \sum_{(\boldsymbol{x}_i, y_i) \in U} \left[y_i - \boldsymbol{w}^* \bullet \boldsymbol{x}_i \right]$

其中 $U = \left\{ \left(\boldsymbol{x}_i, y_i \right) \mid \left(\boldsymbol{x}_i, y_i \right) \in$ **训练集** $\boldsymbol{T}$，并且 $\alpha_i^* > 0 \right\}$

(5) 分类模型的使用：

$$g\left(\boldsymbol{x}\right) = \left(\boldsymbol{w}^* \cdot \boldsymbol{x}\right) + b^* = \sum_{i=1}^{N} \alpha_i^* y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}) + b^*$$

对于未知类别的测试样本 $\boldsymbol{x}$，

待决策样本的预处理

$$\begin{cases} 若 f(\boldsymbol{x}) = \operatorname{sgn}\left[g\left(\boldsymbol{x}\right) \right] = +1, & 则决策\boldsymbol{x}为第1类 \\ 若 f(\boldsymbol{x}) = \operatorname{sgn}\left[g\left(\boldsymbol{x}\right) \right] = -1, & 则决策\boldsymbol{x}为第2类 \end{cases}$$

主要内容

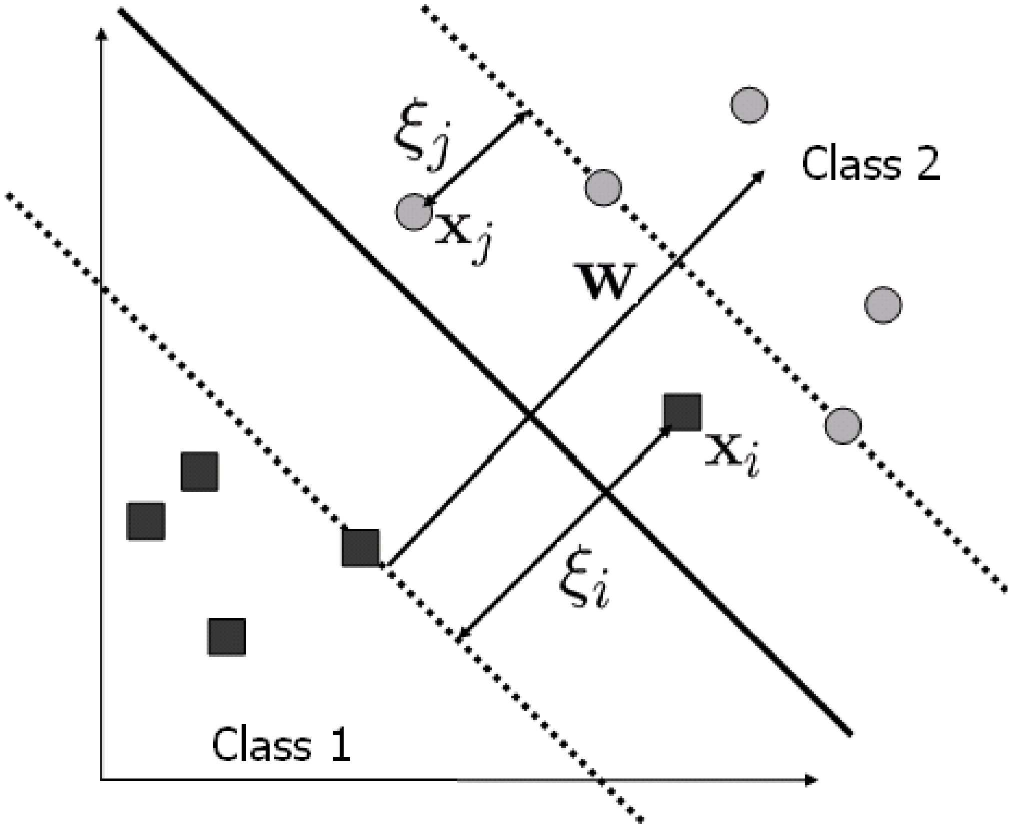
1　两类别分类问题：线性可分情况

2　两类别分类问题：近似线性可分情况

3　多类别的线性SVM

4　应用举例

1.例 二维特征空间、两类别、线性不可分情况



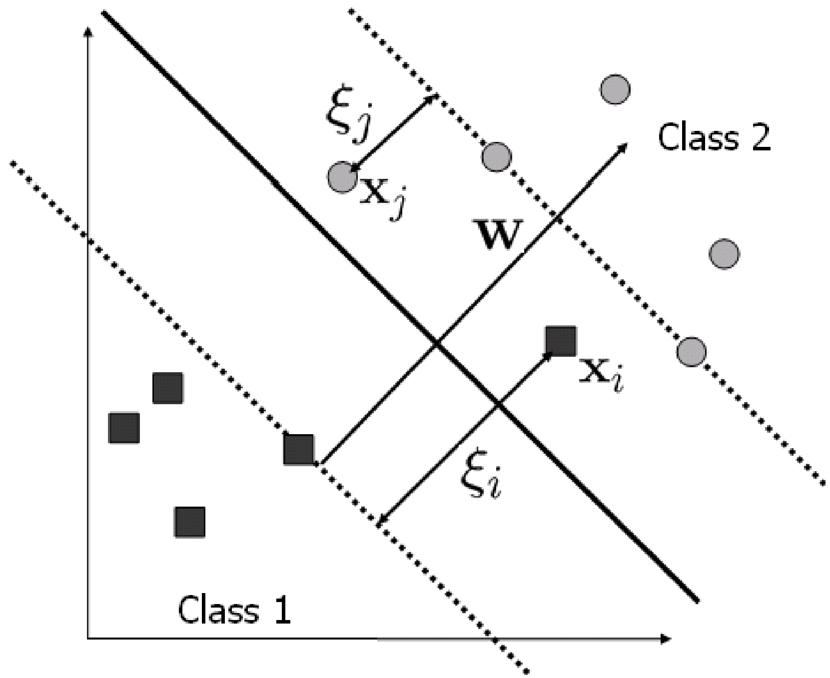
适用：含噪声的训练样本；

容许对训练样本进行线性错分

2.广义线性SVM—最优分类面的求取

(1)问题描述

$$\begin{cases} \text{已知：训练样本集}\{\boldsymbol{x}_i, y_i\}, i=1,\ldots,N \qquad \boldsymbol{x}_i \in R^d \\ \qquad\qquad \text{类别标号 } y_i \in \{-1,1\} \\ \text{估计：广义最优线性判别函数 } g(\boldsymbol{x}) = \boldsymbol{w}\cdot\boldsymbol{x}+b \end{cases}$$



(2)目标函数的构造

折中：”最大分类间隔＋ 训练样本集的最小错分程度”

归一化判别函数 $g(\boldsymbol{x})=\boldsymbol{w}\cdot\boldsymbol{x}+b$

对所有训练样本，引入松弛变量 $\xi_i\geq 0,\ i=1,\ldots,N$

$$for\quad i=1,\ldots,N\quad \begin{cases} g(\boldsymbol{x}_i)=\boldsymbol{w}\cdot\boldsymbol{x}_i+b\geq 1-\xi_i & \text{若}\, y_i=+1 \\ g(\boldsymbol{x}_i)=\boldsymbol{w}\cdot\boldsymbol{x}_i+b\leq -\left(1-\xi_i\right) & \text{若}\, y_i=-1 \end{cases}$$

$$\rightarrow y_i\cdot g(\boldsymbol{x}_i)+\xi_i\geq 1$$

$$\rightarrow y_i\cdot\left[\boldsymbol{w}\cdot\boldsymbol{x}_i+b\right]\geq 1-\xi_i$$

目标函数构造的基本思想

$$\begin{cases} \text{分类间隔} margin = \dfrac{2}{\|\mathbf{w}\|} \text{尽可能大} & \Leftrightarrow \quad \dfrac{1}{2}(\mathbf{w}\cdot\mathbf{w}) \downarrow \\ \text{训练样本集的错分程度尽可能低} & \Leftrightarrow \quad \sum_{i=1}^{N} \xi_i \downarrow \end{cases}$$

原始目标函数

松弛因子向量 $\boldsymbol{\xi} = [\xi_1 \quad \cdots \quad \xi_N]^T$

C - SVC（标准SVC）

$$\min_{\mathbf{w},\xi,b} \quad \phi(\mathbf{w},\xi,b) = \frac{1}{2}(\mathbf{w}\cdot\mathbf{w}) + C\left(\sum_{i=1}^{N}\xi_i\right)$$

$$s.t. \quad \begin{cases} -\left[y_i \cdot (\mathbf{w}\cdot\mathbf{x}_i + b) - 1 + \xi_i\right] \le 0 \\ -\xi_i \le 0 \end{cases} \quad i = 1, 2, \ldots, \mathrm{N}$$

$C > 0$——模型的超参数，控制错分样本的惩罚程度

(3)目标函数的求解

构造*Lagrange*辅助目标函数，将**原始问题**转化为**对偶问题**.

$$
\begin{cases}
\min \quad L(\mathbf{w},b,\xi;\boldsymbol{\alpha},\boldsymbol{\mu})=\dfrac{1}{2}(\mathbf{w}\cdot\mathbf{w})+C\left(\sum_{i=1}^{N}\xi_i\right) \\
\qquad -\sum_{i=1}^{N}\alpha_i\left\{y_i\cdot[\mathbf{w}\cdot\mathbf{x}_i+b]-1+\xi_i\right\}-\sum_{i=1}^{N}\mu_i\xi_i, \\
\alpha_i\ge 0,\quad i=1,2,\ldots,N \\
\mu_i\ge 0,\quad i=1,2,\ldots,N
\end{cases}
$$

其中非负*Lagrange*乘子向量 $\begin{cases}\boldsymbol{\alpha}=[\alpha_1 \quad \cdots \quad \alpha_N]^T \\ \boldsymbol{\mu}=[\mu_1 \quad \cdots \quad \mu_N]^T\end{cases}$

显然：$L(\mathbf{w},b,\xi;\boldsymbol{\alpha},\boldsymbol{\mu})\le\phi(\mathbf{w},\xi,b)=\dfrac{1}{2}(\mathbf{w}\cdot\mathbf{w})+C\left(\sum_{i=1}^{N}\xi_i\right)$

原始问题等价为： $\min\limits_{\mathbf{w},b,\xi}\max\limits_{\boldsymbol{\alpha},\boldsymbol{\mu}} L(\mathbf{w},b,\xi;\boldsymbol{\alpha},\boldsymbol{\mu})$

原始目标函数

C - SVC（标准SVC）

$$\min_{\mathbf{w},\xi,b} \quad \phi(\mathbf{w},\xi,b)=\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})+C\left(\sum_{i=1}^{N}\xi_i\right)$$

$$s.t. \quad \begin{cases} -\left[y_i\cdot(\mathbf{w}\cdot\mathbf{x}_i+b)-1+\xi_i\right]\le 0 \\ -\xi_i\le 0 \end{cases} \quad i=1,2,\dots,\mathrm{N}$$

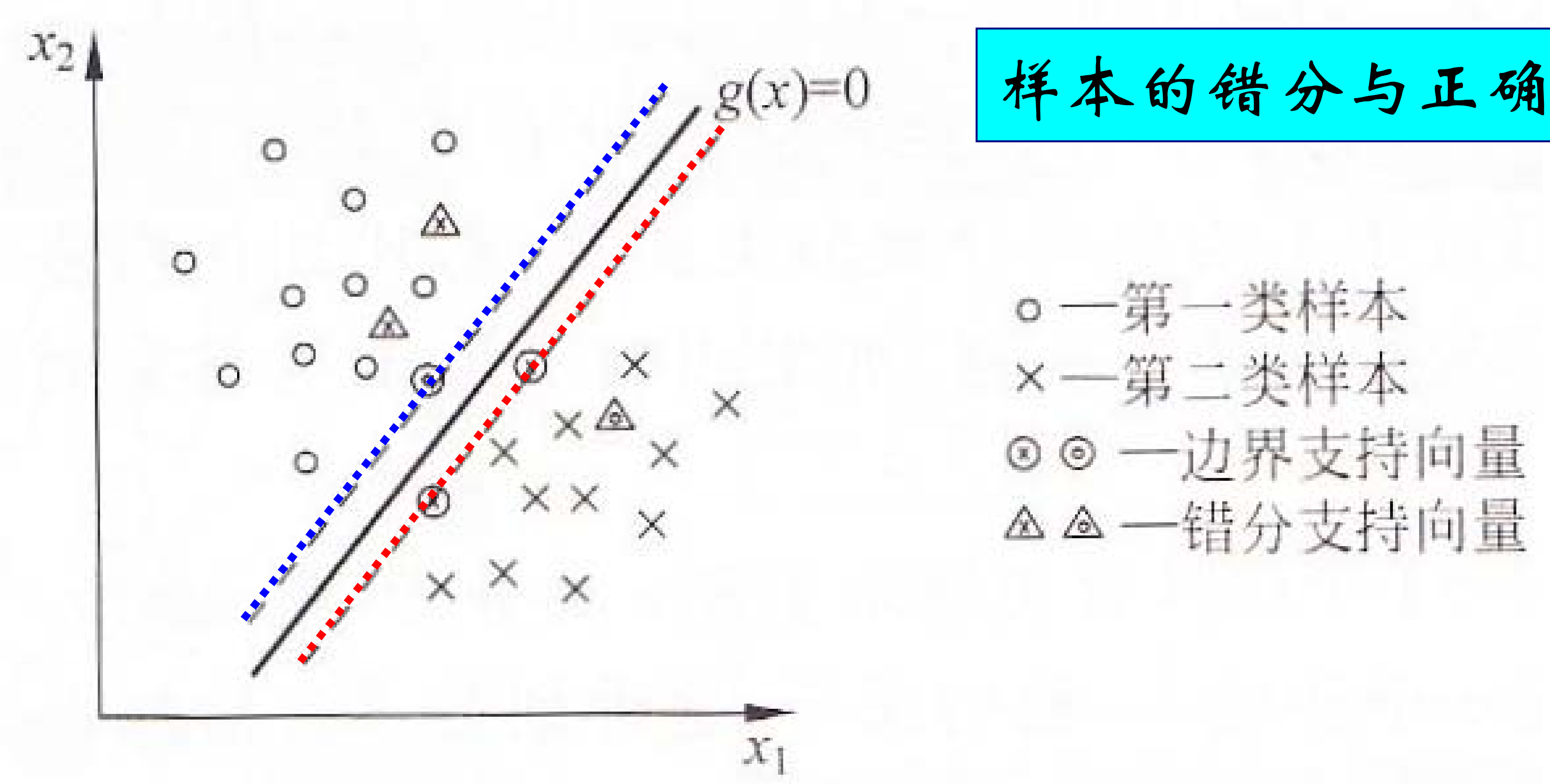
对偶目标函数

$$\max_{\boldsymbol{\alpha}}. \quad \sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_i\alpha_j y_i y_j\left(\mathbf{x}_j\cdot\mathbf{x}_i\right)$$

$$s.t.\begin{cases} (1)\quad \sum_{i=1}^{N}\alpha_i y_i=0 \\ (2)\quad 0\le\alpha_i\le C,\quad i=1,\dots,N \end{cases}$$

解得$\boldsymbol{\alpha}^*=\left[\alpha_1^*,\dots,\alpha_N^*\right]^T$

样本的错分与正确分类



训练样本集T的划分　基于$\boldsymbol{\alpha}^*=\left[\alpha_1^* \quad \cdots \quad \alpha_N^*\right]^T$中各分量取值的三种可能划分

$$\begin{cases} \alpha_i^*=0 & \text{训练样本}\boldsymbol{x}_i\text{为非支持向量} \\ 0<\alpha_i^*<C & \text{训练样本}\boldsymbol{x}_i\text{为边界支持向量（位于}H_1,H_2\text{超平面）} \\ \alpha_i^*=C & \text{训练样本}\boldsymbol{x}_i\text{为错分支持向量} \end{cases}$$

训练样本集$=\{$**非支持向量**$\}\cup\{$**边界支持向量**$\}\cup\{$**错分支持向量**$\}$

广义线性*SVM*中，判别函数的最优参数w^*、b^*的确定：

A. $$\boldsymbol{w}^* = \sum_{i=1}^{N} \alpha_i^* y_i \boldsymbol{x}_i$$

对$\boldsymbol{w}^*$有贡献的只有占少数的非零α_i^*

对应非零α_i^*的训练样本就称为**支持向量**(***Suport Vector***)

B. $\boldsymbol{b}^*$的确定

找到所有**边界支持向量**，构成集合：

$$U=\{(\boldsymbol{x}_i, y_i) \mid (\boldsymbol{x}_i, y_i) \in \text{训练集}\boldsymbol{T}，\text{并且} 0<\alpha_i^*<C\}$$

对于$\forall(\boldsymbol{x}_i, y_i) \in U$，必有：$y_i = \boldsymbol{w}^* \bullet \boldsymbol{x}_i + \boldsymbol{b}^*$

为了确保稳定性，取平均：

$$b^* = \frac{1}{\#U} \sum_{(\boldsymbol{x}_i, y_i) \in U} \left[y_i - \boldsymbol{w}^* \bullet \boldsymbol{x}_i \right]$$

$\#U$----集合U中元素数目(或表示为$|U|$)

(4)广义线性SVM的判别函数

广义最优分类决策函数：

$$f(\boldsymbol{x}) = \operatorname{sgn}\left[\left(\boldsymbol{w}^* \bullet \boldsymbol{x}\right) + b^*\right] = \operatorname{sgn}\left\{\sum_{i=1}^{N} \alpha_i^* y_i \left(\boldsymbol{x}_i \bullet \boldsymbol{x}\right) + b^*\right\}$$

广义线性SVM（C - SVC）原始问题的目标函数

$$\min_{\mathbf{w},\xi,b} \ \phi(\mathbf{w},\xi,b)=\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})+C\left(\sum_{i=1}^{N}\xi_i\right)$$

$$s.t.\quad \begin{cases}-\left[y_i\cdot(\mathbf{w}\cdot\mathbf{x}_i+b)-1+\xi_i\right]\le 0\\ -\xi_i\le 0\end{cases}\qquad i=1,2,\dots,\mathrm{N}$$

非负*Lagrange*乘子向量的最优解：$\boldsymbol{\alpha}^*=\left[\alpha_1^* \ \ \cdots \ \ \alpha_N^*\right]^T$

最优分类边界 $(\mathbf{w}^*\cdot\mathbf{x})+b^*=0$

$$\begin{cases}\mathbf{w}^*=\sum_{i=1}^{N}\alpha_i^* y_i\mathbf{x}_i\\ b^*=\frac{1}{\#U}\sum_{(\mathbf{x}_i,y_i)\in U}\left[y_i-\mathbf{w}^*\bullet\mathbf{x}_i\right]\\ U=\left\{(\mathbf{x}_i,y_i)\,|\,(\mathbf{x}_i,y_i)\in \textbf{训练集}\boldsymbol{T}\text{，并且}0<\alpha_i^*<C\right\}\end{cases}$$

判别函数 $g(\mathbf{x})=(\mathbf{w}^*\cdot\mathbf{x})+b^*=\sum_{i=1}^{N}\alpha_i^* y_i(\mathbf{x}_i\cdot\mathbf{x})+b$

$$f(\mathbf{x})=\operatorname{sgn}\left[(\mathbf{w}^*\cdot\mathbf{x})+b^*\right]=\operatorname{sgn}\left\{\sum_{i=1}^{N}\alpha_i^* y_i(\mathbf{x}_i\cdot\mathbf{x})+b^*\right\}$$

[线性C-SVC算法]

(1) 给定训练样本集 $T=\{(\boldsymbol{x}_1,y_1),\cdots(\boldsymbol{x}_N,y_N)\}$,

训练样本的预处理

其中 $\boldsymbol{x}_i \in R^d, y_i \in \mathcal{Y}=\{1,-1\}, i=1,\ldots,N$

(2) 选择适当的惩罚参数 C [问题：如何选择参数 C]

(3) 构造并求解凸二次规划问题

$$\begin{cases} \min\limits_{\boldsymbol{\alpha}}. & \dfrac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_i\alpha_j y_i y_j(\boldsymbol{x}_j\cdot\boldsymbol{x}_i)-\sum_{i=1}^{N}\alpha_i \\ s.t. & \begin{cases} (1) & \sum_{i=1}^{N}\alpha_i y_i = 0 \\ (2) & 0\le\alpha_i\le C \qquad \text{对于} i=1,\ldots,N \end{cases} \end{cases}$$

得解 $\boldsymbol{\alpha}^*=\left[\alpha_1^*, \quad \cdots, \quad \alpha_N^*\right]^T$

(4) 估计$\boldsymbol{w}^* = \sum_{i=1}^{N} \alpha_i^* y_i \boldsymbol{x}_i$

(5) 估计$\boldsymbol{b}^*$

找到所有**边界支持向量**，构成集合：

$U = \left\{ (\boldsymbol{x}_i, y_i) \mid (\boldsymbol{x}_i, y_i) \in$ **训练集**$\boldsymbol{T}$，并且$0 < \alpha_i^* < C \right\}$

对于$\forall (\boldsymbol{x}_i, y_i) \in U$，必有：$y_i = \boldsymbol{w}^* \bullet \boldsymbol{x}_i + \boldsymbol{b}^*$

$$b^* = \frac{1}{\# U} \sum_{i \in U} \left[y_i - \sum_{j=1}^{N} \alpha_j^* y_j \left(\boldsymbol{x}_j \cdot \boldsymbol{x}_i \right) \right]$$

(6) 分类模型的使用：$g(\boldsymbol{x}) = \left(\boldsymbol{w}^* \cdot \boldsymbol{x} \right) + b^* = \sum_{i=1}^{N} \alpha_i^* y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}) + b^*$

对于未知类别的测试样本$\boldsymbol{x}$，

$$\begin{cases} \text{若} f(\boldsymbol{x}) = \operatorname{sgn}\left[g(\boldsymbol{x}) \right] = +1, & \text{则决策}\boldsymbol{x}\text{为第1类} \\ \text{若} f(\boldsymbol{x}) = \operatorname{sgn}\left[g(\boldsymbol{x}) \right] = -1, & \text{则决策}\boldsymbol{x}\text{为第2类} \end{cases}$$

待决策样本的预处理

第二部分

基于两类别SVC模型的多类别分类

主要内容

1. 多类问题描述

2基于两类SVC的方法

SVM主要是解决两类问题；实际应用以多类问题为主

问题：

如何利用已有的两类SVM模型，解决多类别的分类问题？

M类分类问题描述：

给定：训练样本集$\boldsymbol{T}=\{(\boldsymbol{x}_1,y_1),\cdots,(\boldsymbol{x}_N,y_N)\}$,

其中 $\boldsymbol{x}_i \in R^d,\ i=1,2,\ldots,N$

$y_i \in \mathcal{Y}=\{1,2,\ldots,M\}$

要求：寻找R^d空间上的决策函数 $f(\boldsymbol{x}): R^d \to \mathcal{Y}$;

用以推断任意输入$\boldsymbol{x}$对应的输出y.

主要内容

1. 多类问题描述

2. 几个典型的基于两类SVC的方法

   2.1 成对分类(one verse one, one against one)

   2.2 一类对余类(one verse the rest)

   2.3 有向无环图

M类分类模型的算法描述(成对分类法):

(1) **给定标准训练样本集** $\boldsymbol{T}=\{(\boldsymbol{x}_1,y_1),\cdots,(\boldsymbol{x}_N,y_N)\}$,

其中 $\boldsymbol{x}_i\in R^d$, $y_i\in\mathcal{Y}=\{1,2,\ldots,M\}$ $i=1,2,\ldots,N$

(2) **训练阶段:**

基于$\boldsymbol{SVC}$,训练 $\boldsymbol{C}_M^2=\dfrac{\boldsymbol{M}(\boldsymbol{M}-\boldsymbol{1})}{\boldsymbol{2}}$个两类$\boldsymbol{SVC}$:

对于所有$(i,j)\in\{(i,j)\mid i<j, \quad i,j=1,\ldots,M\}$,进行如下运算:

从训练集中分别抽取对应$y=i$和$y=j$的训练样本,分别作为正样本点和负样本点,构成训练集$\boldsymbol{T}_{i对j}$;

由$\boldsymbol{T}_{i-j}$求解两类分类问题的$\boldsymbol{SVC}_{i对j}$,得到实值函数$\boldsymbol{g}_{i对j}(\boldsymbol{x})$,以及用于判定任意输入$\boldsymbol{x}$属于第$i$或$j$类的决策函数$\boldsymbol{f}_{i对j}(\boldsymbol{x})$

$$\boldsymbol{f}_{i对j}(\boldsymbol{x})=\begin{cases} i & 若\boldsymbol{g}_{i对j}(\boldsymbol{x})>0 \\ j & 若\boldsymbol{g}_{i对j}(\boldsymbol{x})<0 \end{cases}$$

(3)分类阶段：

构建针对上述$\frac{M(M-1)}{2}$个两类*SVC*的**投票表决器**；

对于未知输入***x***，考虑上述所有的*SVC*的判决结果，某*SVC*判定***x***属于第i个类别，相应的第i类的**表决器票数增1**。

票数最多的**表决器**对应的类别，即为***x***的最终判决类别。

*LIBSVM*基于“成对分类”实现多类问题的分类；得票数最多类别不止一个时，将***x***判决为对应**最高票数的最小类别**。

实际使用时，应结合具体问题做适当处理。

主要内容

1. 多类问题描述

2 基于两类SVC的方法

- 2.1 成对分类(one verse one, one against one)
- 2.2 一类对余类(one verse the rest)
- 2.3 有向无环图

算法描述

(1)给定M类分类的标准训练样本集

$$\boldsymbol{T}=\{(\boldsymbol{x}_1,y_1),\cdots,(\boldsymbol{x}_N,y_N)\}$$

其中 $\boldsymbol{x}_i \in R^d$， $y_i \in \mathcal{Y}=\{1,2,\ldots,M\}$ $i=1,2,\ldots,N$

(2)训练：对于$j=1,\ldots,M$,进行如下运算：

从训练集$\boldsymbol{T}$中分别抽取对应$\boldsymbol{j}$和其余$M-1$类的训练点，分别作为**正类**和**负类**样本点，用于训练第j个两类分类问题的$\boldsymbol{SVC}_{j对其它}$，得到$\boldsymbol{g}_{j对其它}(\boldsymbol{x})$

(3)决策：判定输入$\boldsymbol{x}$属于第J类.

其中，$J=\arg\max\limits_{j} \boldsymbol{g}_{j对其它}(\boldsymbol{x})$

主要内容

0. 多类问题描述

1. 基于两类SVC的方法

    1.1 成对分类(one verse one, one against one)

    1.2 一类对余类(one verse the rest)

    1.3 有向无环图

    Directed Acyclic Graph Support Vector Machine (DAGSVM)

有向无环图-算法(1)

(1)给定M类分类的训练样本集 $\boldsymbol{T}=\{(\boldsymbol{x}_1,y_1),\cdots,(\boldsymbol{x}_N,y_N)\}$

其中 $\boldsymbol{x}_i\in R^d$, $y_i\in\mathcal{Y}=\{1,2,\ldots,M\}$ $i=1,2,\ldots,N$

(2)训练$\dfrac{M(M-1)}{2}$个两类***SVC***.

对于所有$(i,j)\in\{(i,j)\mid i<j,\quad i,j=1,\ldots,M\}$,进行如下运算：

从训练集中分别抽取对应$y=i$和$y=j$的训练点，分别作为正样本点和负样本点，构成训练集$\boldsymbol{T}_{i对j}$；由$\boldsymbol{T}_{i对j}$求解$\boldsymbol{SVC}_{i对j}$,得到实值函数$\boldsymbol{g}_{i对j}(\boldsymbol{x})$,以及用于判定任意输入$\boldsymbol{x}$属于第$i$或$j$类的决策函数$\boldsymbol{f}_{i对j}(\boldsymbol{x})$

$$\boldsymbol{f}_{i对j}(\boldsymbol{x})=\begin{cases} i & \boldsymbol{g}_{i对j}(\boldsymbol{x})>0 \\ j & \boldsymbol{g}_{i对j}(\boldsymbol{x})<0 \end{cases}$$

有向无环图-算法(2)

(3)构建二叉树.

使之具有：$\frac{M(M-1)}{2}$个内部节点，M个叶子：

每个内部节点对应一个两类SVC;

叶子对应最终的类别。

(4)分类. 判定输入x的所属类别,从图的根节点开始输入x,

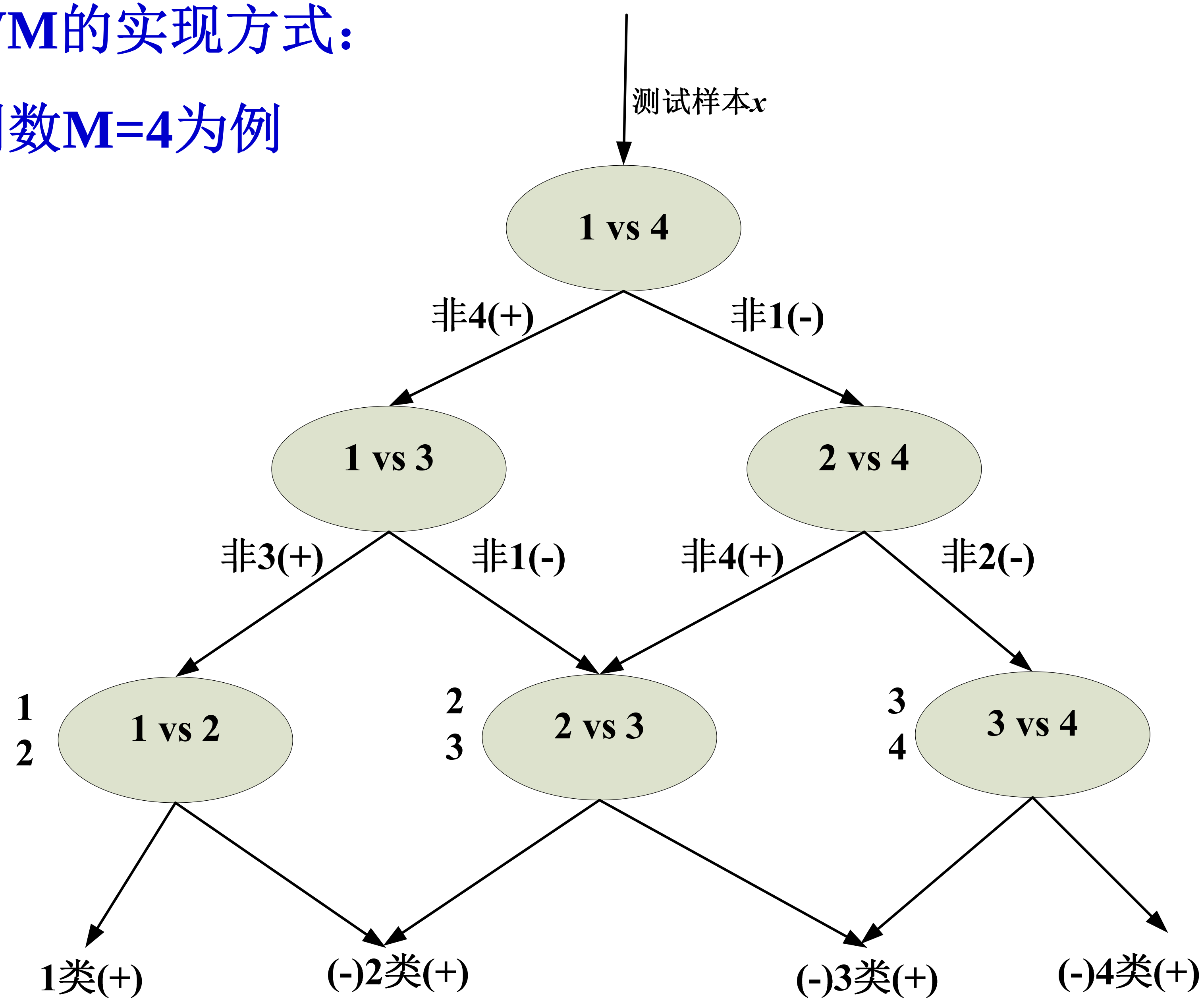
根据每个节点的输出决定左侧或者右侧路径，

一直到达叶子，即可确定所属类别。

每个x只需访问$M-1$个节点。

LIBSVM的实现方式:

以类别数M=4为例



第三部分

两类别的核SVC（非线性SVC）

主要内容

1 核SVM

线性SVM回顾

非SVM基本思想

变换空间的最优分类面

核函数作为内积的条件

几种典型核函数

核函数形式的选取

核函数的构造

C-SVC算法描述

2 应用举例

(1)线性SVC的回顾

线性SVM：训练样本线性可分

原始目标函数：

$$\min_{\mathbf{w},b} \quad \frac{1}{2}\|\mathbf{w}\|^2=\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})$$

$$s.t. \quad y_i\cdot[\mathbf{w}\cdot\mathbf{x}_i+b]-1\geq 0$$

$$i=1,2,\ldots,N$$

广义线性SVM：训练样本线性不可分

原始目标函数：

$$\min_{\mathbf{w},b,\xi} \quad \frac{1}{2}(\mathbf{w}\cdot\mathbf{w})+C\left(\sum_{i=1}^{N}\xi_i\right)$$

$$s.t. \quad y_i\cdot[\mathbf{w}\cdot\mathbf{x}_i+b]-1+\xi_i\geq 0$$

$$\xi_i\geq 0 \qquad i=1,2,\ldots,N$$

最优分类函数：

$$f(\mathbf{x})=\operatorname{sgn}\left[\left(\mathbf{w}^*\bullet\mathbf{x}\right)+b^*\right]=\operatorname{sgn}\left\{\sum_{i=1}^{N}\alpha_i^* y_i\left(\mathbf{x}_i\bullet\mathbf{x}\right)+b^*\right\}$$

其中　对应$\alpha_i^*=0$的$\mathbf{x}_i$为支持向量

最优分类面：　$\mathbf{w}^*\bullet\mathbf{x}+b^*=0$

分类函数只涉及内积运算

(1) 线性SVC的回顾

线性SVM存在局限性：

- 灵活性差
- 很多实际问题往往不是线性可分的

。。。

线性不可分情况的两种处理方式

方式1：引入松弛项 → 广义线性SVM

方式2：是否在其它空间线性可分？→ 非线性SVM

(2)非线性SVM思想

- 通过非线性映射，将原始数据由低维特征空间变换到高维空间(可能是无限维)

特征空间		变换空间
原始数据	$\xrightarrow{\text{非线性映射}}$	高维数据
$\boldsymbol{x}=\left[x_1,x_2,\cdots,x_d\right]^T$		$\Phi(\boldsymbol{x})=\left[\Phi_1(\boldsymbol{x}),\Phi_2(\boldsymbol{x}),\cdots,\Phi_j(\boldsymbol{x}),\ldots\right]^T$

- 在高维变换空间中设计线性SVM，寻求最优分类面

- 非线性变换是通过定义适当的内积函数隐式实现的

(3)高维变换空间中的最优分类面

- 对于非线性SVM,将原始数据映射到高维变换空间后,变换空间的线性SVM:

原始问题:

$$\begin{aligned} &\min_{\mathbf{w},b,\xi} \quad \frac{1}{2}(\mathbf{w}\cdot\mathbf{w})+C\left(\sum_{i=1}^{N}\xi_i\right) \\ &s.t. \quad y_i\cdot\left[\mathbf{w}\cdot\Phi\left(\mathbf{x}_i\right)+b\right]-1+\xi_i\geq 0 \\ &\qquad \xi_i\geq 0 \qquad\qquad i=1,2,\ldots,N \end{aligned}$$

对偶问题:

$$\begin{aligned} &\max_{\boldsymbol{\alpha}} \quad Q(\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_i\alpha_j y_i y_j\left(\Phi\left(\mathbf{x}_j\right)\cdot\Phi\left(\mathbf{x}_i\right)\right) \\ &s.t. \quad \begin{cases} (1) \quad \sum_{i=1}^{N}\alpha_i y_i=0 \\ (2) \quad 0\leq\alpha_i\leq C \qquad \text{对于所有}\, i \end{cases} \end{aligned}$$

最优分类函数

$$f(\boldsymbol{x}) = \operatorname{sgn}\left[\left(\boldsymbol{w}^* \cdot \Phi(\boldsymbol{x})\right) + b^*\right]$$

$$= \operatorname{sgn}\left\{\sum_{i=1}^{N} \alpha_i^* y_i \left(\Phi(\boldsymbol{x}_i) \bullet \Phi(\boldsymbol{x})\right) + b^*\right\}$$

最优分类面

$$\left(\boldsymbol{w}^* \cdot \Phi(\boldsymbol{x})\right) + b^* = 0$$

其中：$\boldsymbol{w}^* = \sum_{i=1}^{n} \alpha_i^* y_i \Phi(\boldsymbol{x}_i)$

目标函数与最优分类函数涉及内积运算，内积形式以核函数表示：

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}_j) = \sum_l \Phi_l(\boldsymbol{x}_i) \Phi_l(\boldsymbol{x}_j)$$

kernel trick

核函数形式的非线性*SVM*：

目标函数：

$\Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}_j)$

$$\max_{\boldsymbol{\alpha}} \quad Q(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j)$$

$$s.t. \quad \begin{cases} (1) \quad \sum_{i=1}^{n} \alpha_i y_i = 0 \\ (2) \quad 0 \le \alpha_i \le C \qquad \text{对于所有} i \end{cases}$$

最优分类函数：$f(\boldsymbol{x}) = \operatorname{sgn}\left\{ \sum_{i=1}^{n} \alpha_i^* y_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b^* \right\}$

对于**函数** $k(\boldsymbol{x},\boldsymbol{y})=(1+x_1y_1+x_2y_2)^2$

考虑如下变换：

$$\phi(\boldsymbol{x})=\phi\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right)=\left(1,\sqrt{2}x_1,\sqrt{2}x_2,x_1^{\ 2},x_2^{\ 2},\sqrt{2}x_1x_2\right)^T$$

$$\phi(\boldsymbol{x})=\phi\left(\begin{bmatrix}y_1\\y_2\end{bmatrix}\right)=\left(1,\sqrt{2}y_1,\sqrt{2}y_2,y_1^{\ 2},y_2^{\ 2},\sqrt{2}y_1y_2\right)^T$$

则变换后的内积计算： $\phi(x)\bullet\phi(y)$

若直接用函数计算，则：$k(\boldsymbol{x},\boldsymbol{y})=(1+x_1y_1+x_2y_2)^2=\phi(x)\bullet\phi(y)$

无需知道变换函数 $\phi(\bullet)$ 的显式形式，即可借助**核函数**，

完成等价于变换空间内积运算，这就是**核技巧**。– ***kernel trick***

(4)函数作为内积核函数的条件——*Mercer*条件

- 统计学习理论告诉我们：只要函数满足Mercer条件，就可以作为内积核函数使用。

若$K(\mathbf{x}_1,\mathbf{x}_2)=K(\mathbf{x}_2,\mathbf{x}_1)$，则$K(\bullet,\bullet)$是对称函数

定理$[Mercer's\ \ condition]$

对于任意的对称函数$K(\mathbf{x},\mathbf{x}^{'})$，它是某个变换空间中的内积运算的充要条件是

$$\iint K(\mathbf{x},\mathbf{x}^{'})\varphi(\mathbf{x})\varphi(\mathbf{x}^{'})d\mathbf{x}d\mathbf{x}^{'}\geq 0 \quad 对于任意的\varphi(\mathbf{x})\neq 0$$

且$\int\varphi^2(\mathbf{x})d\mathbf{x}<\infty$ 均成立。

相应的$K(\mathbf{x},\mathbf{x}^{'})$为$Mercer$核。

(5) SVM常用的核函数形式

(I) 多项式核函数　$K(\mathbf{x}, \mathbf{x}') = [(\mathbf{x} \cdot \mathbf{x}') + 1]^{q}$

q阶多项式***SVM***

特例 线性内积函数　$K(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x} \cdot \mathbf{x}_i)$

(II) RBF

$$K(\mathbf{x}, \mathbf{x}') = \exp\left\{-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{\sigma^2}\right\} = \exp\left\{-\gamma \|\mathbf{x} - \mathbf{x}'\|^2\right\}$$

RBF-SVM

(III) S型内积函数　$K(\mathbf{x}, \mathbf{x}') = \tanh\left(v(\mathbf{x} \cdot \mathbf{x}') + c\right)$

SVM为3层感知器神经网络

只有部分v，c值满足***MERCER***条件

(6)核函数类型及核函数参数的选择

基本原则： 由线性至非线性，由简单至复杂.

- 若原始特征空间的维数>>样本数目,可采用线性核函数

- 非线性核函数，通常使用RBF函数，理由：

- RBF函数可以将样本非线性地映射到更高维(无限维)空间中. Sigmoid核函数取某些特定参数时,性能和RBF相同。
- RBF函数的参数只有一个。
- RBF函数的数值限制条件少。RBF函数取值限制在0和1之间，而多项式核函数的值可能会趋于不定值或零值,且幂值更高；Sigmoid核函数在取某些参数值时则可能无效。
- 应先尝试较大窗宽的RBF函数，窗宽越大，越接近线性

(6) $C-$支持向量分类机算法($\boldsymbol{C\text{-}SVC}$)

STEP1 给定训练样本集$\boldsymbol{T}=\{(\boldsymbol{x}_1,y_1),\cdots,(\boldsymbol{x}_n,y_n)\}$,

其中 $\boldsymbol{x}_i\in R^d$, 类别标号 $y_i\in\mathcal{Y}=\{-1,1\}$

$i=1,2,\ldots,n$

STEP2 选取适当核函数$\boldsymbol{K}(x,x')$及惩罚参数$C>0$

$$\begin{cases}\text{核函数形式的确定}\\ \text{核函数的参数及惩罚参数C的确定}\end{cases}$$

STEP3 构造并求解凸二次规划问题

$$\min_{\boldsymbol{\alpha}} \quad Q(\boldsymbol{\alpha}) = \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n}\alpha_i\alpha_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) - \sum_{i=1}^{n}\alpha_i$$

$$s.t.\begin{cases} (1) \quad \sum_{i=1}^{n}\alpha_i y_i = 0 \\ (2) \quad 0 \le \alpha_i \le C \qquad i = 1, 2, \ldots, n \end{cases}$$

得解$\boldsymbol{\alpha}^* = \left[\alpha_1^*, \quad \cdots, \quad \alpha_n^*\right]^T$

***STEP*4　计算$\boldsymbol{b}^*$：**

从$\boldsymbol{\alpha}^* = \left[\alpha_1^*, \quad \cdots, \quad \alpha_n^*\right]^T$中选取$\alpha_j^* \in (0, C)$，计算

$$\boldsymbol{b}^* = \boldsymbol{y}_j - \sum_{i=1}^{n} \alpha_i^* y_i k\left(x_i, \boldsymbol{x}_{\boldsymbol{j}}\right)$$

为了确保稳定性，取平均运算，其中U为所有位于$(0, C)$的α_i^*下标集合。 $b^* = \frac{1}{\# U} \sum_{i \in U} \left[y_i - \sum_{j=1}^{n} \alpha_j^* y_j k\left(\boldsymbol{x}_i, \boldsymbol{x}_{\boldsymbol{j}}\right) \right]$

***STEP*5　构造分划超平面$\left(\omega^* \cdot \phi(\boldsymbol{x})\right) + \boldsymbol{b}^* = 0$，得决策函数：**

$$f(\boldsymbol{x}) = \operatorname{sgn}\left\{g(\boldsymbol{x})\right\}$$

其中$g(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i^* y_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b^*$

参数寻优方法

以 C-SVC 为例，假定核函数为 $K(\mathbf{x}, \mathbf{x}_i) = \exp\left\{-\gamma \left\| \mathbf{x} - \mathbf{x}_i \right\|^2\right\}, \gamma > 0$

进行参数 (C, γ) 的优选，如：

方法1："*grid searching*" + "*k - fold cross - validation*"

LIB - SVM 的代码实现

方法2："遗传算法" + "*k - fold cross - validation*"

关于待求参数进行染色体编码；

以识别率作为贴近度函数

方法3："粒子群优化算法" + "*k - fold cross - validation*"

参数选择

(1)**系统性能评价**----"***k - fold cross - validation***"

A.训练集划分. 将$\boldsymbol{n}$个训练样本组成的训练集，随机分成$\boldsymbol{k}$组规模一致

且互不相交子集，即 $k-$折：$\mathrm{S}_1,\ldots,\mathrm{S}_k$

$(k=n$时，称做留一法，***leave - one - out, LOO***$)$

B.训练与测试. $\text{for}: i=1,\ldots,k$

$$\begin{cases}\text{取}\mathrm{S}_i\text{为}\textbf{测试子集}\text{，其它k}-1\text{个子集共同组成}\textbf{训练子集}\text{；}\\ \text{由}\textbf{训练子集}\text{训练,得到分类器判别决策函数；}\\ \text{对}\textbf{测试子集}\mathrm{S}_i\text{测试，统计错分样本数目}l_i\end{cases}$$

C.误差估计.

基于$k-$折交叉验证的误差估计 $\dfrac{1}{k}\sum_{i=1}^{k}\dfrac{l_i}{\#\mathrm{S}_i}$

基于留一法的误差估计 $\dfrac{\sum_{i=1}^{n} l_i}{n}$

参数选择

(2)参数的搜索----"*grid-search*"

首先，选定一组(C,γ)的范围, 比如$\begin{cases}\log_2 C: & -5\sim 15 \\ \log_2 \gamma: & -15\sim 3\end{cases}$

为避免时间消耗，可采用**coarse-to-fine的搜索策略**。

first, loose grid-search----目的：识别较好的***grid***

如：搜索步长 $\left(step_{\log_2 C},\ step_{\log_2 \gamma}\right)=(2,\ 2)$

找到具有最小"$k-$折交叉验证的误差"的***grid***对应的(C_0,γ_0)

then, fine grid-search----在(C_0,γ_0)邻域，获取更好的(C,γ)

减小搜索步长，在更小的区域内精细搜索。

如：搜索步长 $\left(step_{\log_2 C},\ step_{\log_2 \gamma}\right)=(0.25,\ 0.25)$

每一组可能的(C,γ)，都对应一次性能评价(误差，或者准确率)

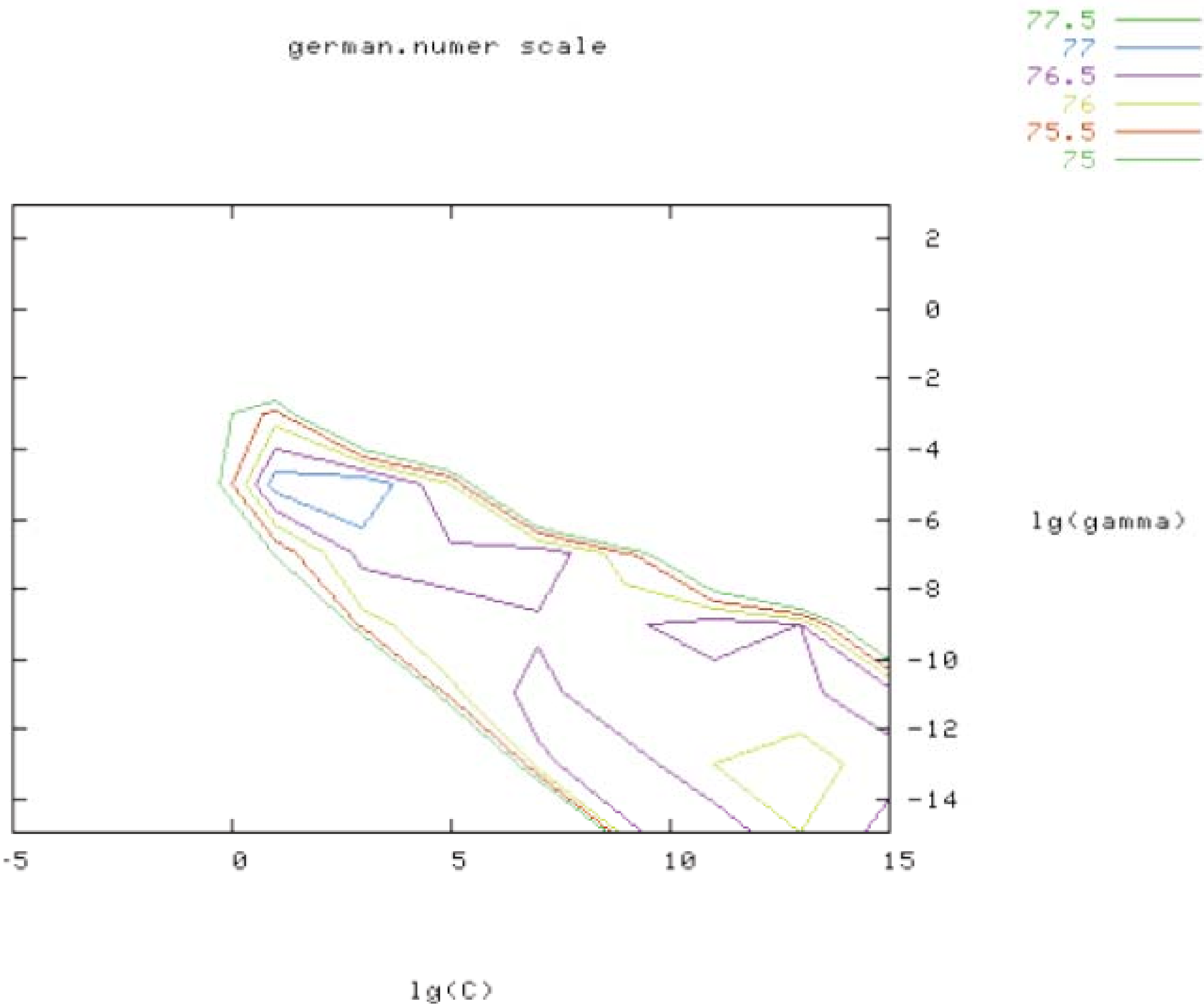
参数选择

网格搜索法的优点：

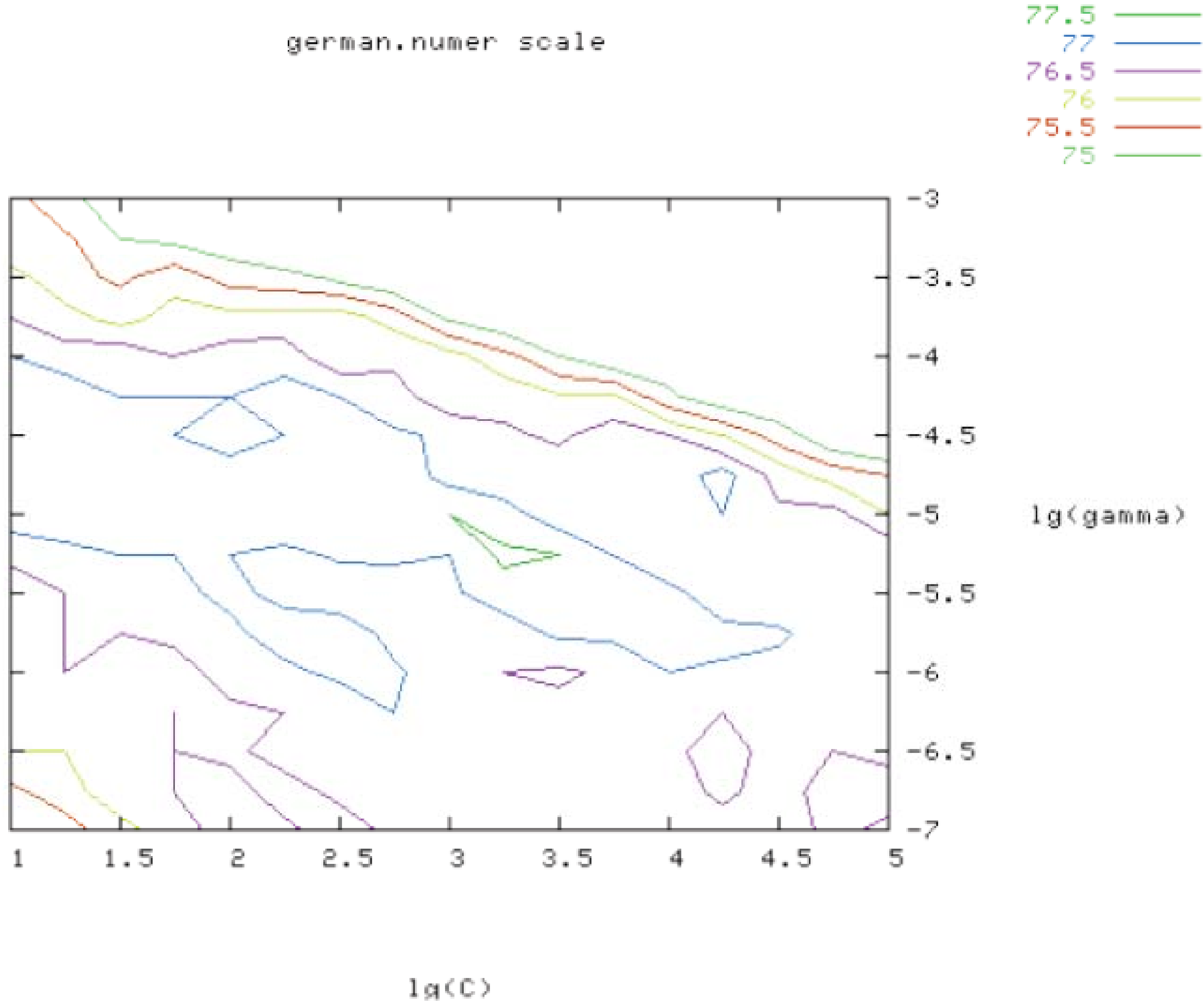
该方法比较简单，易操作；可以同时搜索多个参数。最终获取的是使分类准确率达到最优的参数组合.

该方法的缺点：

需反复进行多次训练，计算量较大.



Loose grid search on $C = 2^{-5}, 2^{-3}, \ldots, 2^{15}$ and $\gamma = 2^{-15}, 2^{-13}, \ldots, 2^{3}$.



Fine grid-search on $C = 2^1, 2^{1.25}, \ldots, 2^5$ and $\gamma = 2^{-7}, 2^{-6.75}, \ldots, 2^{-3}$.

主要内容

1 核SVM

- 线性SVM回顾
- 非SVM基本思想
- 变换空间的最优分类面
- 核函数作为内积的条件
- 几种典型核函数
- 核函数形式的选取
- 核函数的构造
- C-SVC算法描述

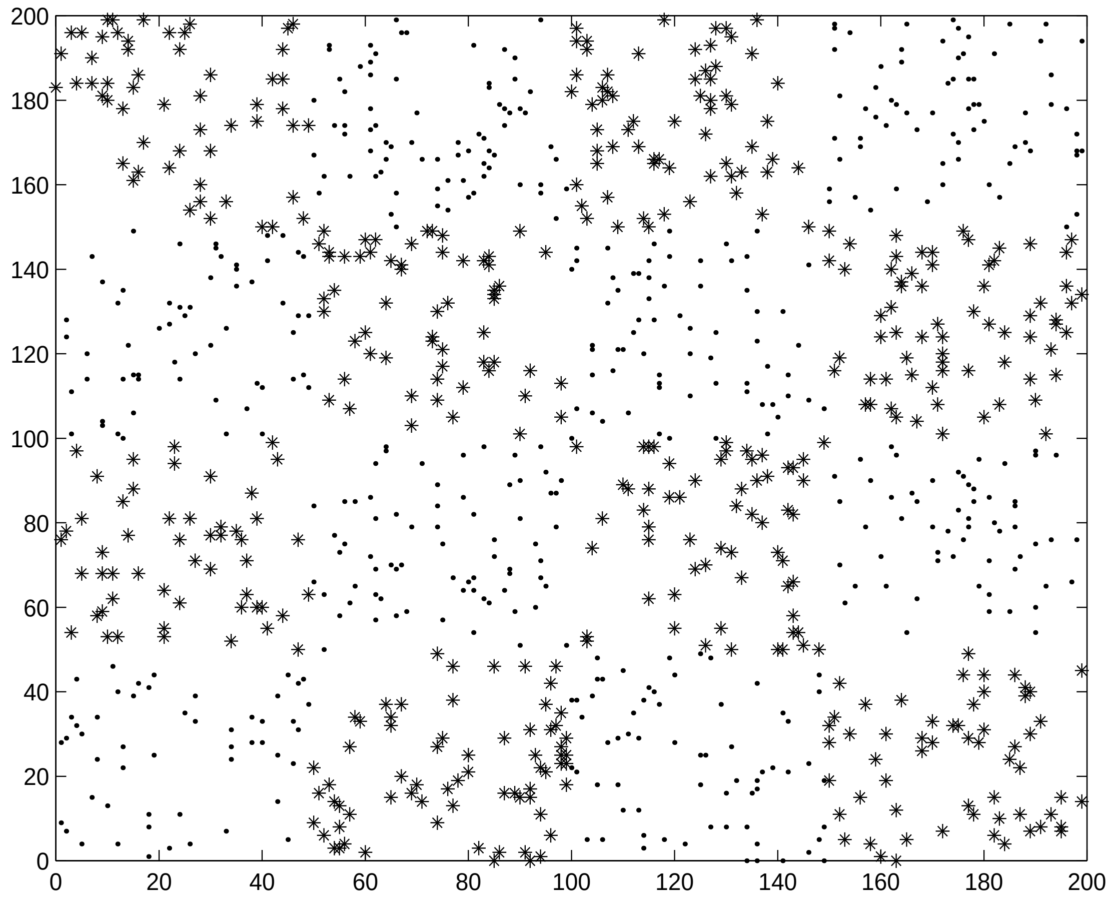
2 应用举例

例： A Nonlinear Kernel Application

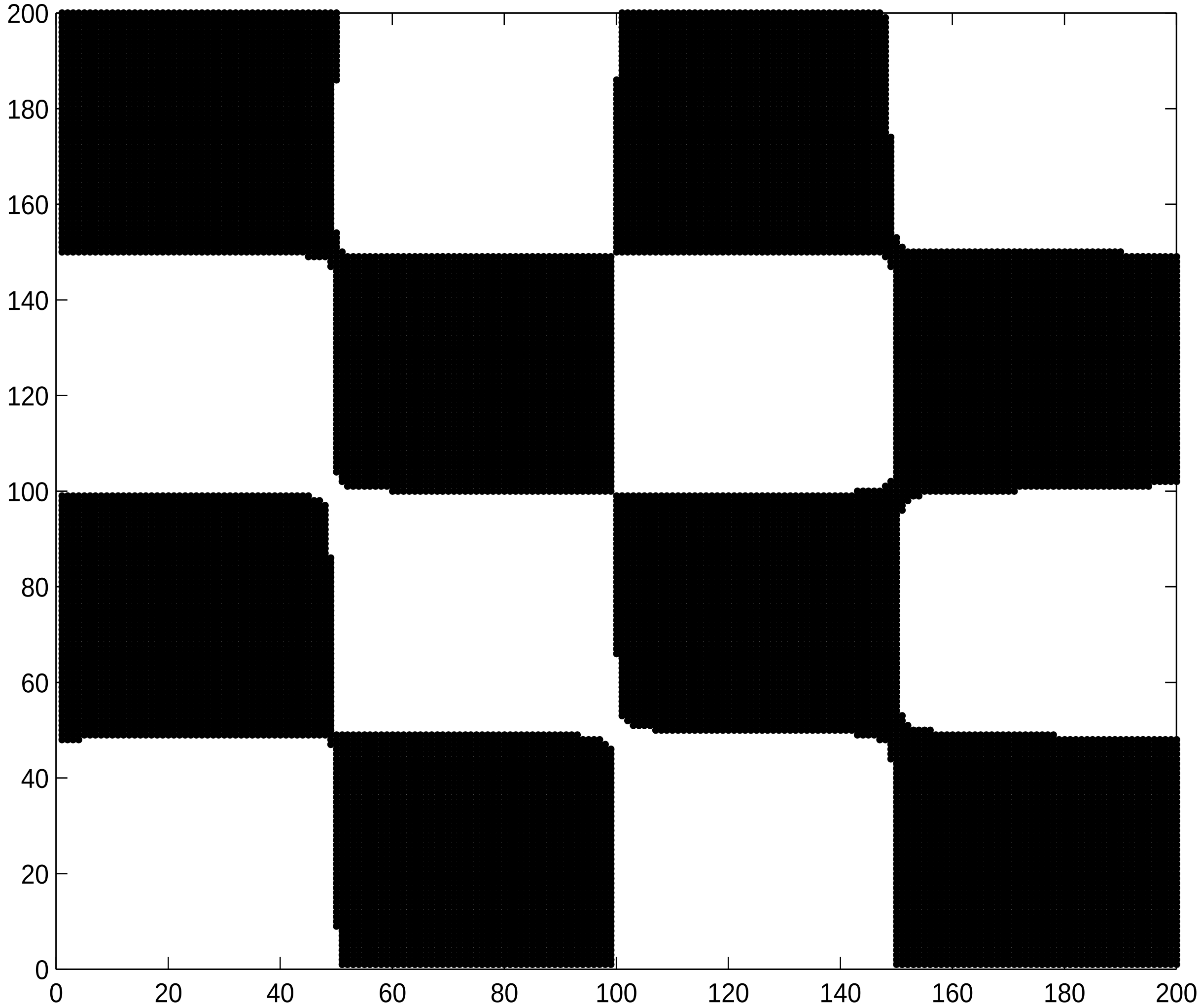
Checkerboard Training Set: 1000 Points in R^2

Separate 486 Asterisks(*) from 514 Dots(.)

两类问题



Polynomial Kernel:

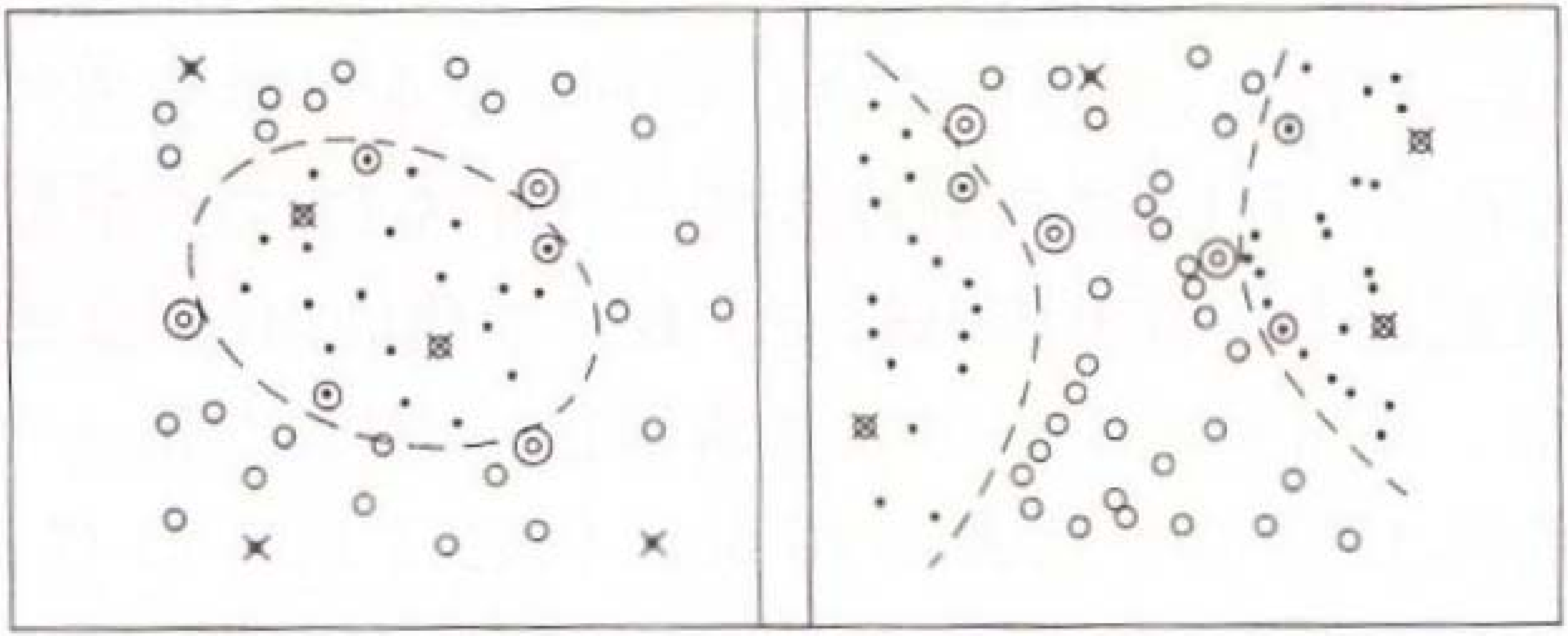


例.两类问题--两组二维合成数据分类

图中：　　　　划叉样本点----错分样本；

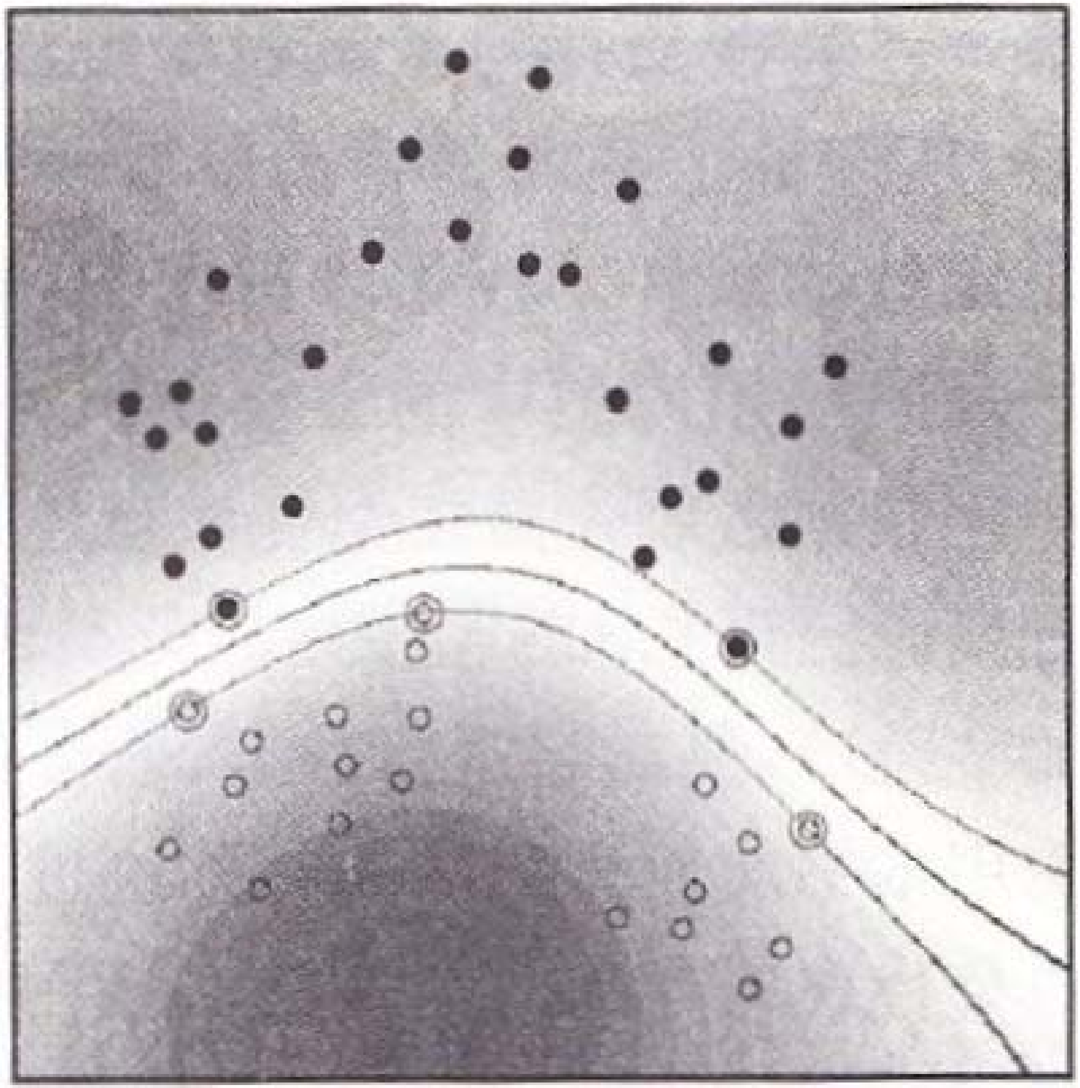
　　　　　　　被圈样本点----支持向量

核函数----q=2阶多项式核函数



例：两类问题。

- 核函数：高斯径向基核函数
- 训练样本
- 支持向量
- 决策边界



例：多类问题

美国邮政标准手写库的数字识别对比实验

部分样本

- 可识别性比较差
- 样本16*16点阵256维特征

训练集：7300个
测试集：2000个

数字识别实质为“10类问题”的分类

第1. 人工及传统方法分类的性能比较

分类器	测试错误率
人工分类	2.5%
决策树C4.5	16.2%
最好的两层神经网络	5.9%
LeNet1(五层神经网络)	5.1%

第2.基于核SVM的数字识别

(1) **多项式核函数** $K(\boldsymbol{x},\boldsymbol{x}_i)=\left[\frac{1}{256}(\boldsymbol{x}\cdot\boldsymbol{x}_i)\right]^q$，参数$q=1,..,6$

(2) **径向基核函数** $K(\boldsymbol{x},\boldsymbol{x}_i)=\exp\left\{-\frac{\left|\boldsymbol{x}-\boldsymbol{x}_i\right|^2}{256\sigma^2}\right\}$

参数$\sigma=0.1\sim 4.0$

(3) ***Sigmoid*型核函数**

$$K(\boldsymbol{x},\boldsymbol{x}_i)=\tanh\left(\frac{b(\boldsymbol{x}\cdot\boldsymbol{x}_i)}{256}-c\right)$$

b,c参数

第2：基于核SVM的数字识别

支持向量机类型	内积函数中的参数	支持向量个数	测试错误率
多项式内积	$q=3$	274	4.0%
径向基函数内积	$\sigma^2=0.3$	291	4.1%
Sigmoid 内积	$b=2, c=1$	254	4.2%

本例结论

- SVM方法比传统方法更有优势
- 不同的SVM方法可以得到性能相近的结果
- SVM分类器所用支持向量数量相对于训练样本总数很少
- 支持向量本身对不同核具有一定的不敏感性

SVM的软件实现

- SVM 资源及实现：

http://www.support-vector.net/index.html

http://www.support-vector.net/software.html

http://www.kernel-machines.org/

http://www.kernel-machines.org/software.html

- 典型工具箱：C 或 MATLB

$\text{SVM}^{\text{light}}$

$\text{SVM}^{\text{Torch}}$

LIBSVM

…

思考：哪些因素会影响SVM分类模型的使用效果？

- 模型的学习过程中是否存在类别不均衡问题？
- 样本是否进行了规范化预处理？
- 分类模型的学习是否进行了超参数的优选？

编程作业：

1. 以鸢尾花数据集为例，分别实现基于线性SVM、RBF-SVM的分类模型。

(1)取前两类，实现两类别分类

(2) 多类别分类