厦门大学信息学院 本科选修课
2021-2022 第二学期

模 式 识 别
Pattern Recognition

主 讲：王 程

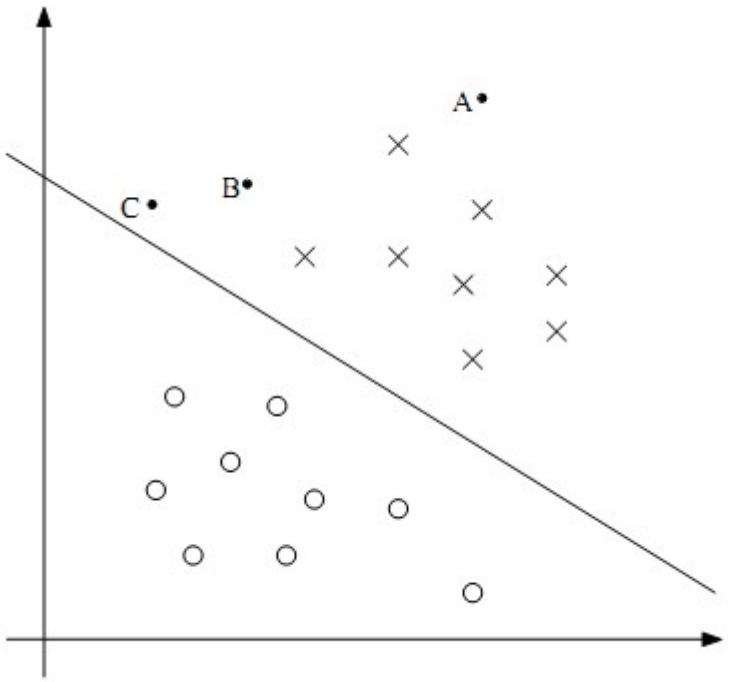
第五章 支持向量机

支持向量机
（support vector machine，SVM）

第五章 支持向量机

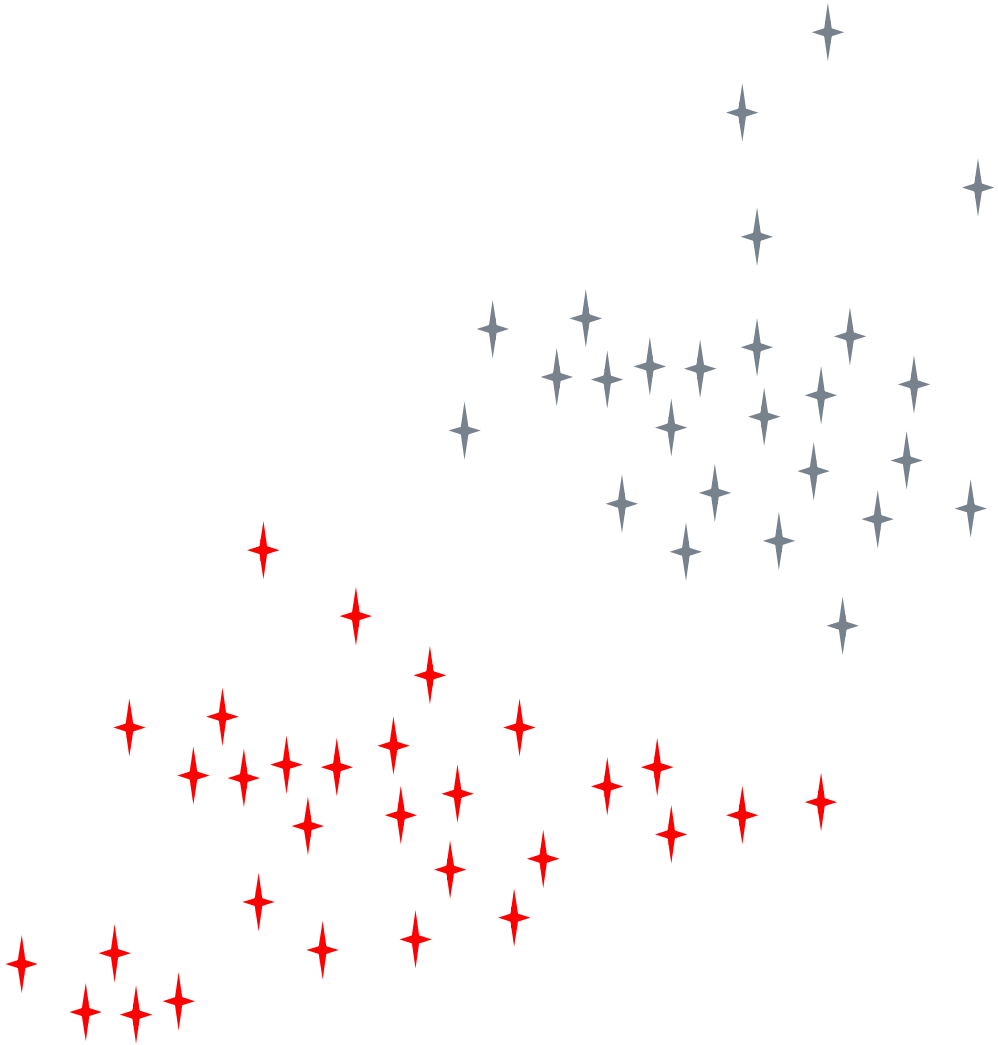
- 5.1 SVM的理论基础
- 5.2 最优分类超平面
- 5.3 支持向量机
- 5.4 SVM的研究与应用

分类间隔的直观理解


A
B
C


Russia
China
Nayfeld
Найфельд
Birofeld
Бирофельд
Krasivoye
Красивое
Lazarevo
Лазарево
Golovino
Головино
Babstovo
Бабстово
Churki
Чурки
Voskresenovka
Воскресеновка
Leninskoye
Ленинское
Novoye
Новое


Russia
China
Nayfeld
Найфельд
Birofeld
Бирофельд
Krasivoye
Красивое
Lazarevo
Лазарево
Babstovo
Бабстово
Churki
Чурки
Leninskoye
Ленинское
Novoye
Новое
Which side to move <- LEFT or -
> RIGHT ?



5.1 SVM的理论基础

- 传统统计模式识别
  - 样本趋向无穷大时，性能才有理论的保证。
  - 这就是经验风险最小化理论(如人工神经网络)

- 统计学习理论（STL）
  - 研究有限样本情况下的机器学习问题
  - SVM 的理论基础就是统计学习理论。

- 推广能力:

  将学习机器(即预测函数,或称学习函数、学习模型)对未来输出进行正确预测的能力。

5.1 SVM的理论基础

- 学习机器风险
  - 经验风险值 + 置信范围值两部分组成
  - 而基于经验风险最小化准则的学习方法只强调了训练样本的经验风险最小误差，没有最小化置信范围值，因此其推广能力较差。

- Vapnik 提出支持向量机(SVM)
  - 以训练误差(经验风险最小)为约束条件，
  - 以置信范围值最小化（推广能力）为优化目标

- SVM: 基于结构风险最小化准则的学习方法
- 1992—1995年

5.1 SVM的理论基础

- SVM 的解是全局唯一的最优解
  - 目标和约束是凸函数

- **SVM在解决小样本、非线性及高维模式识别问题中表现出许多特有的优势**

第五章 支持向量机

- 5.1 SVM的理论基础
- 5.2 最优分类超平面
- 5.3 支持向量机
- 5.4 SVM的研究与应用

5.2 最优分类超平面

- SVM 是从线性可分情况下的最优分类面

➢方形点和圆形点代表两类样本

➢ H为分类线， H1， H2分别为过各类中离分类线最近的样本且平行于分类线的直线，它们之间的距离叫做**分类间隔**(margin)。

➢**最优分类超平面**（optimal seperating hyperplane）

➢如果一个超平面能够将样本正确分开

➢且离超平面最近的样本与超平面之间的距离最大。

➢两类样本中离分类面最近的样本到分类面的距离称作**分类间隔**（margin）。

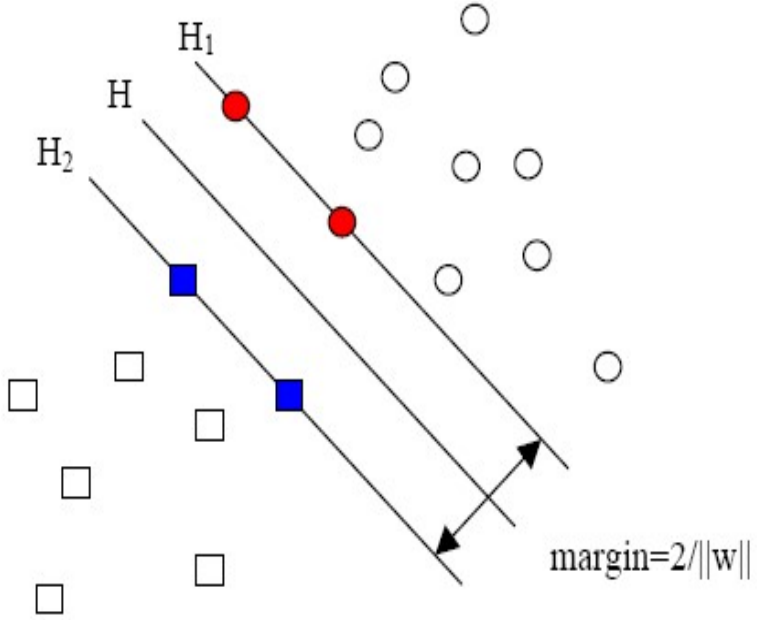


图2 线性可分情况下的最优分类线

5.2 最优分类超平面

设线性可分的样本集(x_i, y_i), $i = 1, \cdots, n, x \in R^d, y \in \{+1, -1\}$。

d维空间中的线性判别函数 $g(x) = w^T x + b$, 分类面方程 $w^T x + b = 0$

所有训练样本都被超平面正确分类，需要满足：

$$\begin{cases} w^T x_i + b > 0, \ y_i = +1 \\ w^T x_i + b < 0, \ y_i = -1 \end{cases}$$

为了避免w和b的尺度调整，将上式规范化为：

$$\begin{cases} w^T x_i + b \geq 1, \ y_i = +1 \\ w^T x_i + b \leq -1, \ y_i = -1 \end{cases}$$

把两个不等式可以合成一个统一的形式：

$$y_i \left[w^T x_i + b \right] \geq 1$$

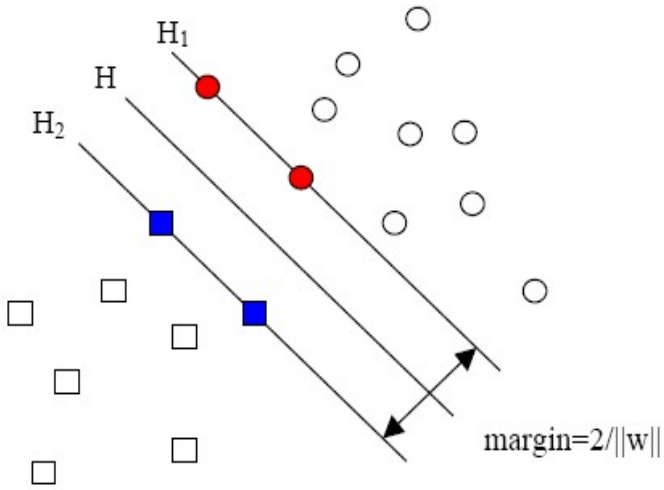


图2 线性可分情况下的最优分类线

这种超平面称为规范化的分类超平面 。 $g(x) = 1$ 和 $g(x) = -1$

就是过两类中各自离分类面最近的样本且与分类面平行的两个边界超平面。

5.2 最优分类超平面

特征空间中任意一点x到超平面的距离为 $r = \frac{g(x)}{\|w\|}$

如果要求离分类面最近的样本满足：$g(x) = 1$

则分类间隔为 $2/\|w\|$

求解最优超平面问题可以表示为：

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

$$s.t.\ y_i\left[w^T x_i + b\right] - 1 \geq 0$$

过两类样本中离分类面最近的点且平行于最优超平面的超平面H1和H2上的训练样本，就是使上式等号成立的样本，也叫支持样本。

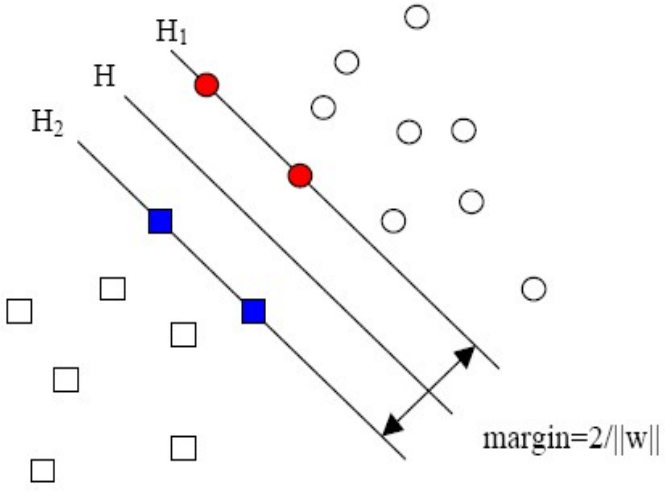


图2 线性可分情况下的最优分类线

5.2 最优分类超平面

这是一个不等式约束的优化问题，可以通过拉格朗日法求解。可以转化为：

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

由于这是一个二次凸规划问题，由于目标函数和约束条件都是凸的，根据最优化理论，这一问题存在唯一全局最小解。这里，它是满足KKT条件：$\alpha_i\left\{y_i\left[w^{T}x_i+b\right]-1\right\}=0$ 。得到的最优解（最优分类面的决策函数）为：

$$f(x)=\mathrm{sgn}\{g(x)\}=\mathrm{sgn}\left\{(w^{*}\bullet x)+b^{*}\right\}=\mathrm{sgn}\left\{\sum_{i=1}^{N}\alpha_i^{*}y_i(x_i\bullet x)+b^{*}\right\}$$

α_i 是支持向量系数，最优超平面的权重向量等于训练样本以一定的系数加权后进行线性组合。只有满足优化式等号成立的样本对应的 α_i 才大于0，其他样本都等于0。求和只对少数支持向量进行。

5.2 最优分类超平面

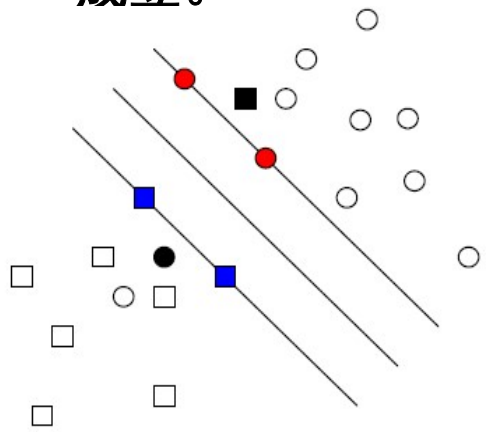
当样本集不是线性可分时，即不等式： $y_i\left[w^T x_i + b\right] \geq 1$ 不能被所有样本同时满足。

既定某个样本 x_k 不满足不等式，即 $y_i\left[w^T x_i + b\right] - 1 < 0$

那么总存在一个正数 ξ_k，使得 $y_i\left[w^T x_i + b\right] - 1 + \xi \geq 0$ 成立。

这时，不等式约束变为：

$$y_i\left[w^T x_i + b\right] - 1 + \xi \geq 0$$



线性不可分下的优化的目标函数转换为：

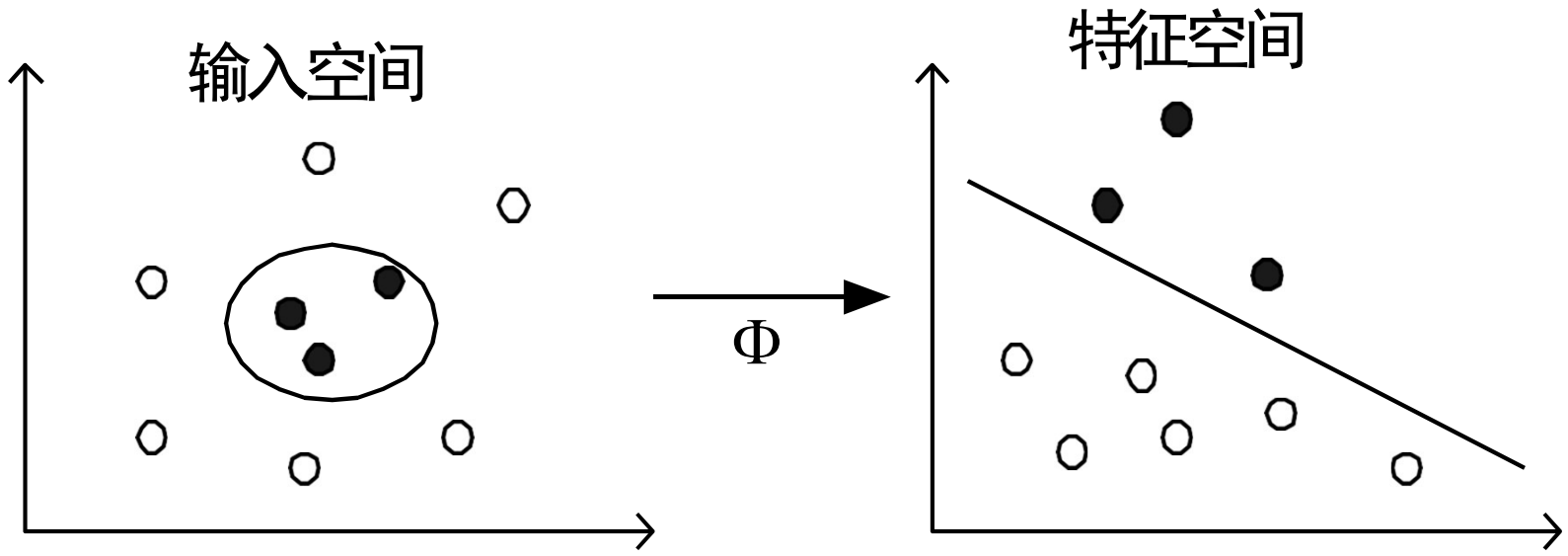
$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{N} \xi_i\right)$$

其中， ξ_i 是松弛因子，$\sum_{i=1}^{N} \xi_i$ 越大，错分的样本越多。此时的目标变成一方面让分类间隔尽可能的大，另一方面让错分的样本尽可能的少且错误率尽可能的低。

第五章 支持向量机

- 5.1 SVM的理论基础
- 5.2 最优分类超平面
- 5.3 支持向量机
- 5.4 SVM的研究与应用

5.3 支持向量机



样本空间经过非线性映射后的特征空间

5.3 支持向量机

线性情况下最优分类面的决策函数为：

$$f(x)=\mathrm{sgn}\{g(x)\}=\mathrm{sgn}\{(w^*\bullet x)+b^*\}=\mathrm{sgn}\left\{\sum_{i=1}^{N}\alpha_i^* y_i(x_i\bullet x)+b^*\right\}$$

如果我们对x进行非线性变换，记作 $z=\varphi(x)$ ，则新的空间里构造的支持向量机决策函数为：

$$f(x)=\mathrm{sgn}\{g(x)\}=\mathrm{sgn}\{(w^{\varphi}\bullet z)+b\}=\mathrm{sgn}\left\{\sum_{i=1}^{N}\alpha_i\, y_i(\varphi(x_i)\bullet\varphi(x))+b\right\}$$

变换对支持向量机的影响是把两个样本在原特征空间中的内积 $(x_i\bullet x)$ 变成了新空间中的内积 $\varphi(x_i)\bullet\varphi(x)$ 。

新空间的内积可以记作： $K(x_i,x_j)=(\varphi(x_i)\bullet\varphi(x_j))$

也称为核函数。

5.3 支持向量机

从计算角度，不论$\varphi(x)$所生成的变换空间维数有多高，这个空间里的线性支持矢量机求解都可以在原空间通过核函数 $K(x_i, x_j)$ 进行，这样就避免了高维空间里的计算，而且计算核函数的复杂度与计算内积并没有实质性增加。

非线性支持矢量机的决策函数可以写为：

$$f(x) = \mathrm{sgn}\left\{\sum_{i=1}^{N} \alpha_i \, y_i K(x_i, x) + b\right\}$$

根据泛函的有关理论，只要一种核函数满足Mercer条件（矩阵正定且可以进行相似性度量），它就对应某一变换空间中的内积。因此，在最优分类面中采用适当的内积函数就可以实现某一非线性变换后的线性分类。

5.3 支持向量机

- SVM通过用内积函数定义的非线性变换将输入空间变换到一个高维空间，在这个空间中求最优分类面
- SVM分类计算的形式类似于神经网络，输出是中间节点的线性组合，每个中间节点对应一个输入样本与一个支持向量的内积，因此也被叫做支持向量网络。

$$f(x)=\operatorname{sgn}\left\{\sum_{i=1}^{N}\alpha_i\, y_i K(x_i,x)+b\right\}$$

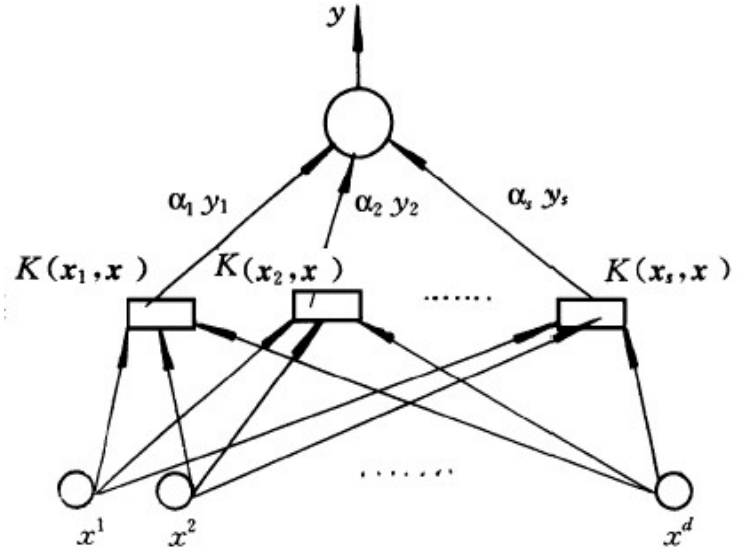


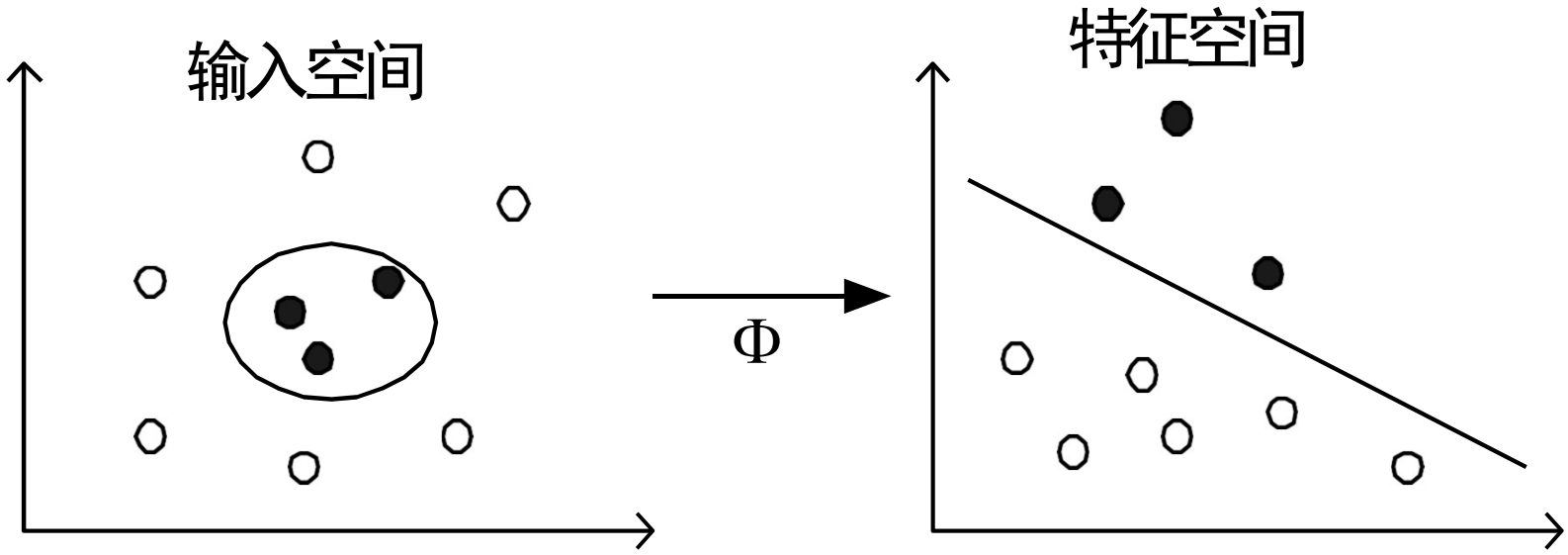
图3　支持向量机示意图

5.3 支持向量机

- 提出
  - 由Vanpik和他的合作者在90年代提出来；
- 特点
  - 解的唯一性且全局最优解；
  - 适用于小样本、高维问题，避免了“维数灾难”问题；
  - 引入核函数，可处理非线性问题，得到的分类或回归超平面在特征空间中是线性的。

5.3 支持向量机

图例：



样本空间经过核映射后的特征空间

5.3 支持向量机

1. 线性支持向量机

我们先考虑线性可分的情况。假设我们有样本：

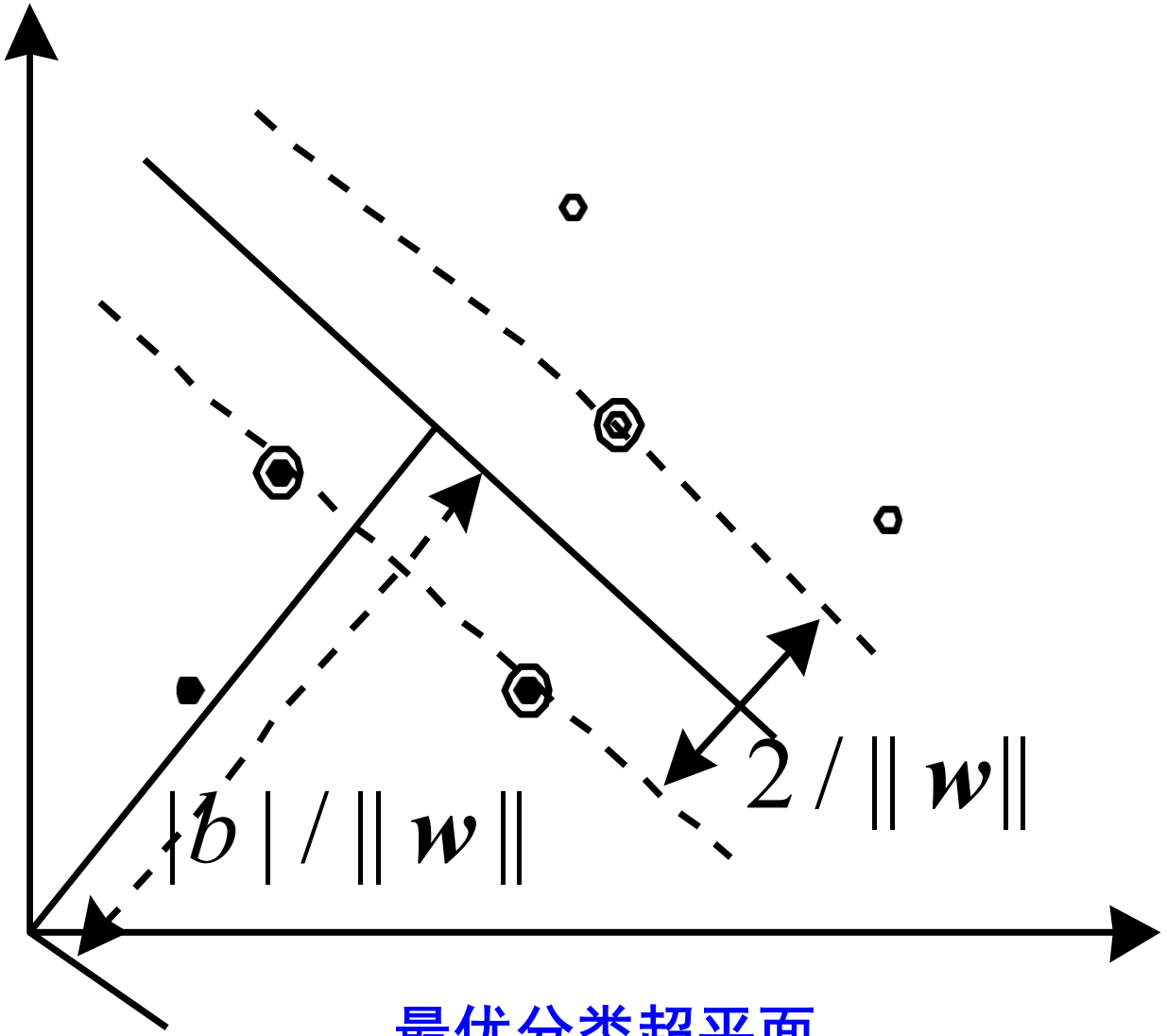
$$\{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in R^N, y_i \in \{-1,1\}, i=1,\ldots,l\}$$

在分离超平面上的点x满足 $\mathbf{w}\mathbf{x} + b = 0$

在超平面H_1上的点x满足 $\mathbf{w}\mathbf{x} + b = 1$

在超平面H_2上的点x满足 $\mathbf{w}\mathbf{x} + b = -1$

5.3 支持向量机



最优分类超平面

5.3 支持向量机

- 那么对所有线性可分的样本都满足下面的约束

$$\begin{cases} \mathbf{w}\mathbf{x}_i + b \geq +1 & y_i = +1 \\ \mathbf{w}\mathbf{x}_i + b \leq -1 & y_i = -1 \end{cases} \qquad (2) \quad r_0 = \frac{w_0}{\|\mathbf{w}\|}$$

- 超平面H_1到原点的距离为 $|1-b|/\|\mathbf{w}\|$
- 超平面H_2到原点的距离为$|-1-b|/\|\mathbf{w}\|$
- 超平面H_1和H_2之间的距离为$2/\|\mathbf{w}\|$

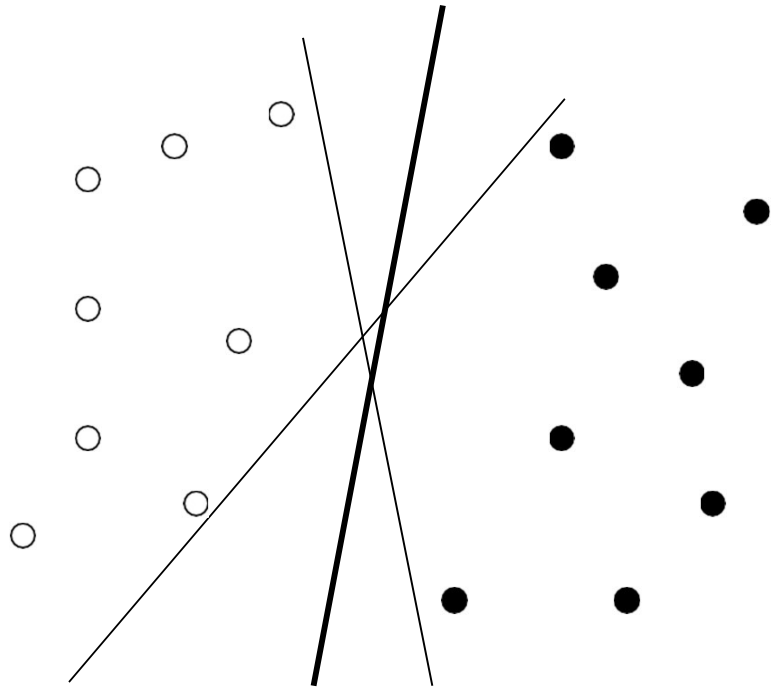
5.3 支持向量机

- 支持向量机算法就是要使得两个超平面H_1和H_2之间的距离最大，也就是分离超平面的间隔(margin)最大，即

$$\min \quad \| \mathbf{w} \|^2 / 2$$

$$s.t. \quad y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1$$

$$i = 1, \cdots, l$$



5.3 支持向量机

- 线性不可分情况：如果两个超平面H_1和H_2之间存在着样本点的话，上面的优化问题得不到一个好的分类结果。Vanpik 等(1995年)引入了一组正的松弛变量ξ，优化问题变为：

不等式约束变为：

$$y_i\left[w^T x_i + b\right] - 1 + \xi \geq 0$$

线性不可分下的优化的目标函数转换为：

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

5.3 支持向量机

2. 非线性支持向量机

- 定义：核映射 Φ

$$\mathbf{x} \in X \xrightarrow{\Phi} \Phi(\mathbf{x}) \in H$$

其中 $$\Phi(\mathbf{x}) = \left(\sqrt{\lambda_1}\phi_1(\mathbf{x}), \sqrt{\lambda_2}\phi_2(\mathbf{x}), \cdots, \sqrt{\lambda_n}\phi_n(\mathbf{x}), \cdots\right)^T$$

- 在样本空间 X 上的线性函数集可以表示为

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$$

与样本空间对应的高维特征空间中线性函数集的表示 $$f(\mathbf{x}) = \sum_{n=1}^{\infty} v_n\phi_n(\mathbf{x}) + b = \mathbf{v}^T\Phi(\mathbf{x}) + b$$

5.3 支持向量机

那么高维空间表示的是再生核希尔伯特空间

RKHS, reproducing kernel Hilbert space

$$f(\mathbf{x}) \in RKHS\, H_K$$

- 核映射把低维样本映射到高维空间
- 原空间线性不可分 -> 高维空间线性可分
- 高维空间决策函数只与核函数和样本集有关，而不必知道具体的映射 Φ 。

5.3 支持向量机

类似线性支持向量机，我们给出非线性支持向量机的优化形式：

- 原规划：

$$\min \ \frac{1}{2}\| \mathbf{v} \|^2 + C\sum_{i=1}^{l}\xi_i$$

- 约束：

$$y_i(\mathbf{v}^T\Phi(\mathbf{x}_i)+b) \geq 1-\xi_i$$

$$\xi_i \geq 0; \quad i = 1,\ldots,l$$

5.3 支持向量机

- 这样线性支持向量机通过简单地引入正定核函数实现了非线性特性。
- 目前常用核函数有

（1）高斯核函数：$K(\mathbf{x}_i,\mathbf{x}_j)=\exp\left(-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|^2}{\sigma^2}\right)$

（2）多项式核函数：$K(\mathbf{x}_i,\mathbf{x}_j)=\left((\mathbf{x}_i\cdot\mathbf{x}_j)+1\right)_q$

（3）Sigmoid函数：$K(\mathbf{x}_i,\mathbf{x}_j)=\tanh\left(v(\mathbf{x}_i\cdot\mathbf{x}_j)+c\right)$

第五章 支持向量机

- 5.1 SVM的理论基础
- 5.2 最优分类超平面
- 5.3 支持向量机
- 5.4 SVM的研究与应用

5.4 SVM的研究与应用

- 多类SVM算法
  - **One-against-all方法**
    把k类问题分解成k个两类问题；
    有k个决策函数
  - **One-against-one方法**
    把k类问题分解成k(k-1)/2个两类问题; 有k个决策函数
  - **其他的整体算法(Weston,Bredensteiner, 和Guermeur)**

5.4 SVM的研究与应用

- SVM 通过核函数实现到高维空间的非线性映射，所以适合于解决本质上非线性的分类、回归和密度函数估计等问题。
- 支持向量方法也为样本分析、因子筛选、信息压缩、知识挖掘和数据修复等提供了新工具。

5.4 SVM的研究与应用

https://www.csie.ntu.edu.tw/~cjlin/libsvm/

LIBSVM -- A Library for Support Vector Machines

Chih-Chung Chang and Chih-Jen Lin

NEW Version 3.25 released on April 14, 2021. Installing the Python interface through PyPI is supported

```
> pip install -U libsvm-official
```

The python directory is re-organized so

```
>>> from libsvm.svmutil import *
```

instead of

```
>>> from svmutil import *
```

should be used.

NEW LIBSVM tools provides **many extensions** of LIBSVM. Please check it if you need some functions not supported in LIBSVM.

NEW We now have a nice page LIBSVM data sets providing problems in LIBSVM format.

NEW A practical guide to SVM classification is available now! (mainly written for beginners)

We now have an easy script (easy.py) for users who know NOTHING about SVM. It makes everything automatic--from data scaling to parameter selection.

The parameter selection tool grid.py generates the following contour of cross-validation accuracy. To use this tool, you also need to install python and gnuplot.

End