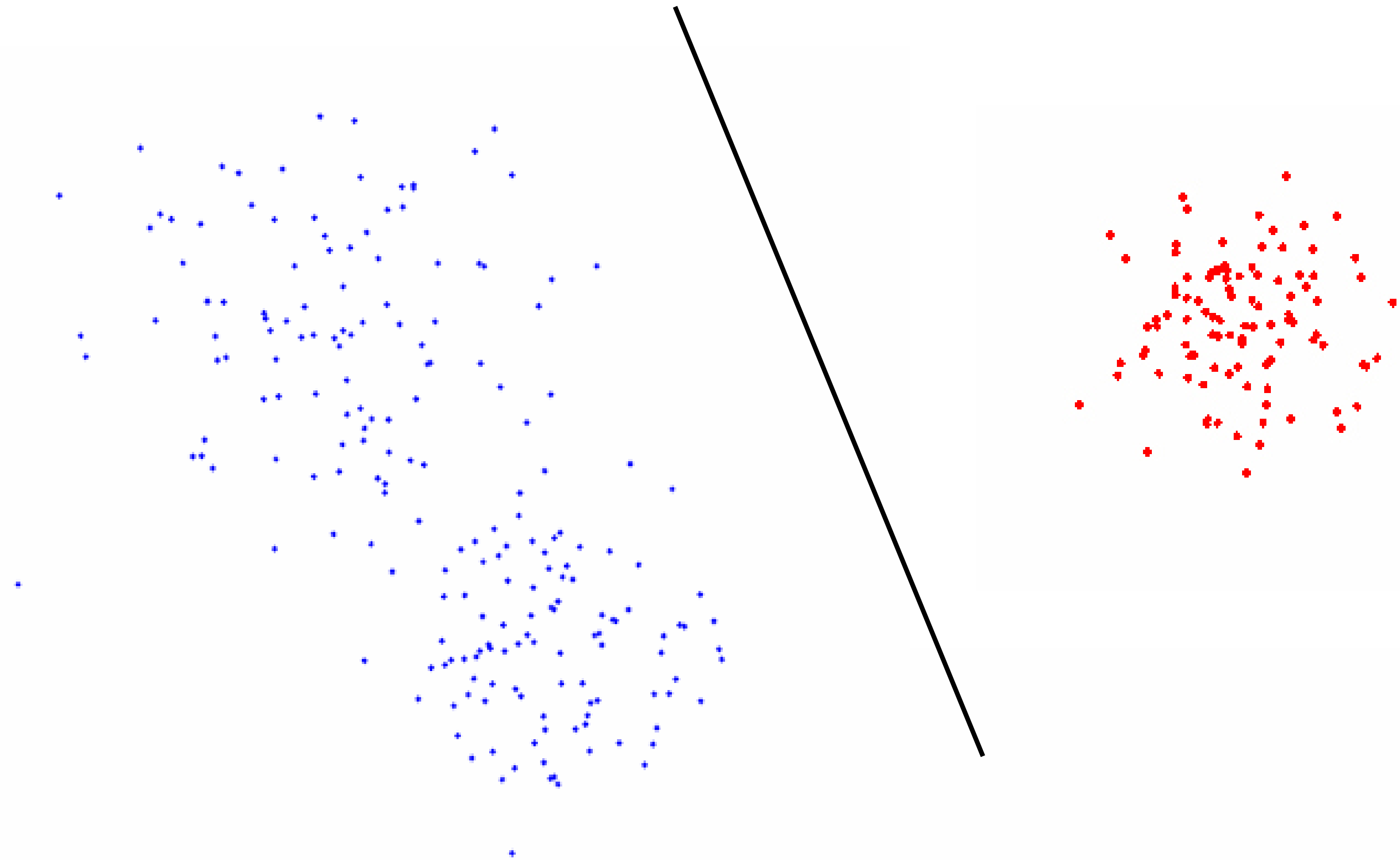
工 业 和 信 息 化 部 “ 十 二 五 ” 规 划 教 材
“十二五”国家重点图书出版规划项目

模式识别

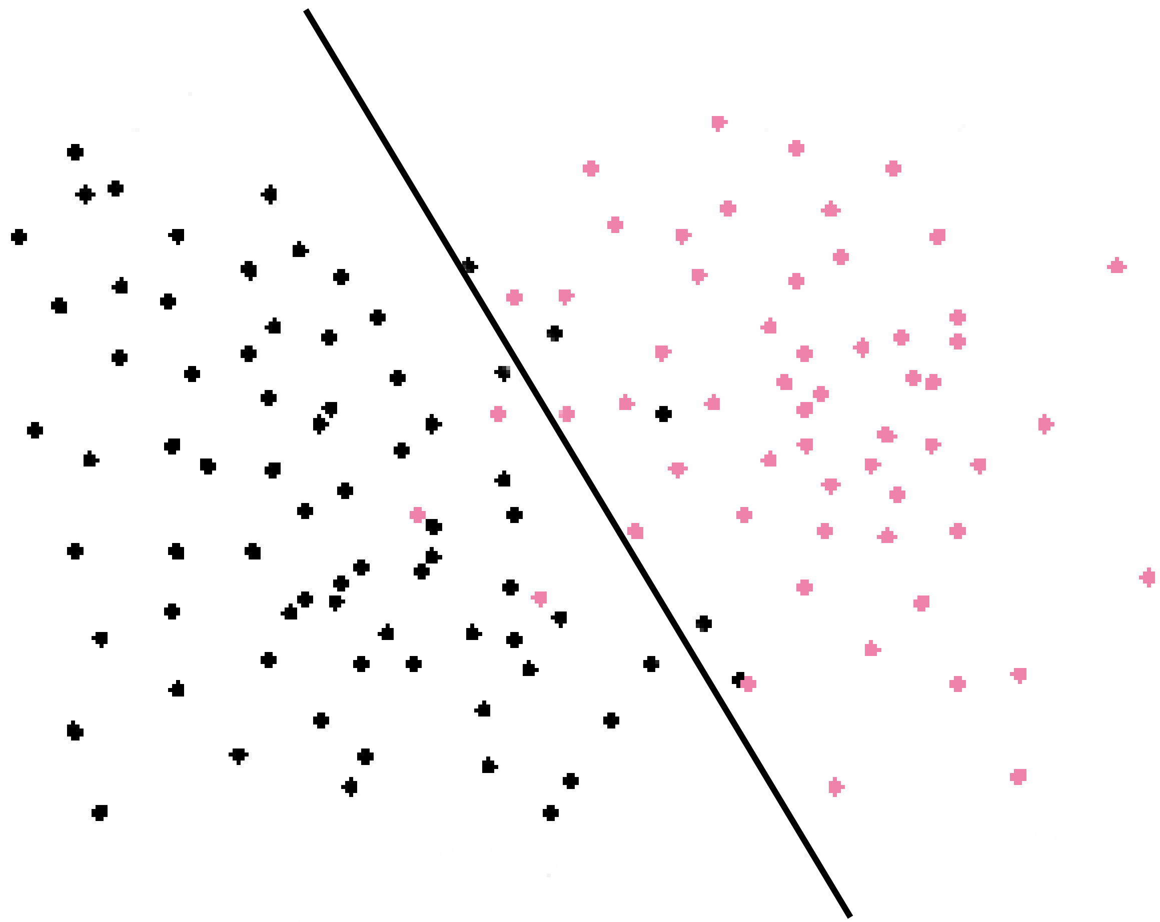
Pattern Recognition

第3讲 线性分类器与支持向量机

哈尔滨工业大学 计算机学院
模式识别与智能系统研究中心
ISBN 978-7-5603-4763-9



ISBN 978-7-5603-4763-9



ISBN 978-7-5603-4763-9

4.1线性判别函数和线性分类界面

线性判别函数

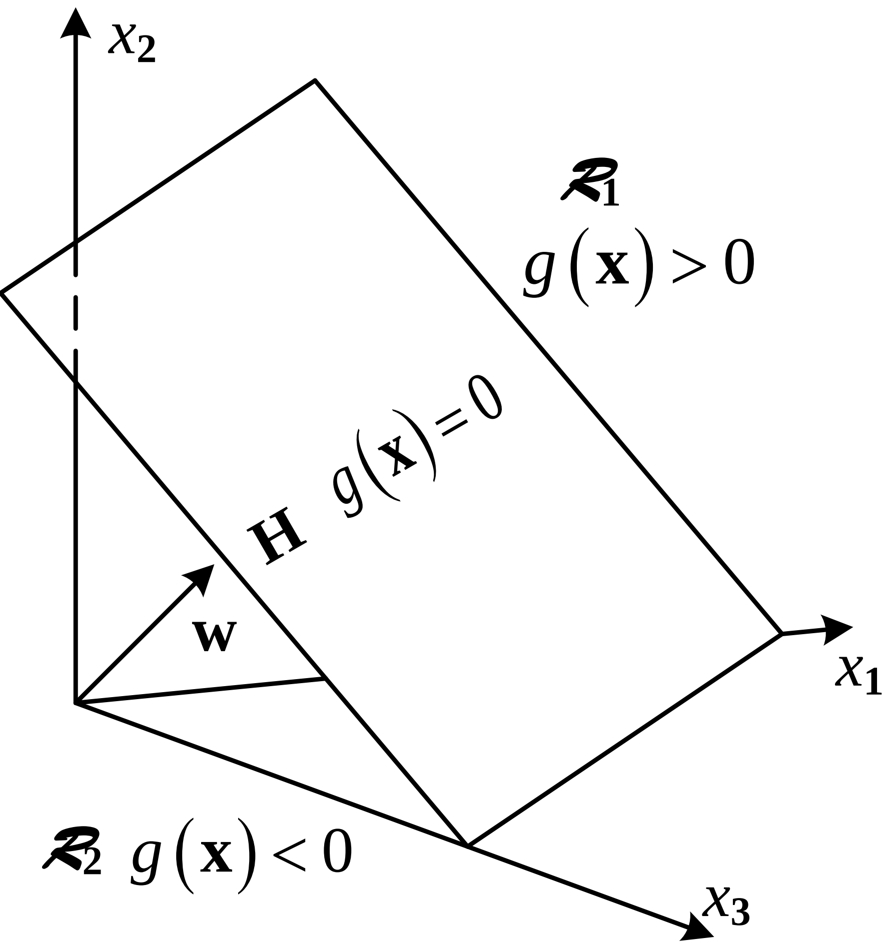
	3维空间平面	d维空间超平面
代数形式	$w_1x_1 + w_2x_2 + w_3x_3 + w_0 = 0$	$w_1x_1 + w_2x_2 + \cdots + w_dx_d + w_0 = 0$
向量形式	$(w_1, w_2, w_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + w_0 = 0$	$(w_1, w_2, \cdots, w_d)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} + w_0 = 0$
	$\mathbf{w}^T\mathbf{x} + w_0 = 0$	其中 $\mathbf{W}$:权值矢量 w_0 :偏置

d维特征空间中的超平面方程 $H: g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = 0$

- 特征矢量：**x**=(x1, x2,…, xd)^T
- 权矢量**w**=(w1, w2, …, wd)^T:
- **偏置(bias)：w0**

4.1线性判别函数和线性分类界面

线性判别函数



- H将特征空间划分为两区域R1,R2

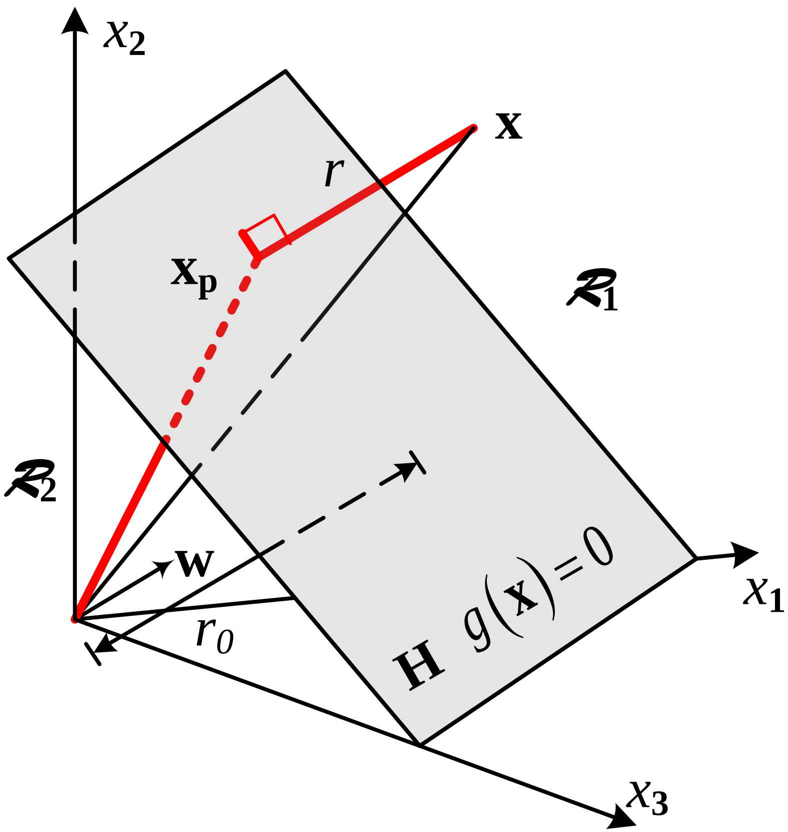
$$g(\mathbf{x})\begin{cases} >0, & \mathbf{x}\text{处于H上方R2区域} \\ =0, & \mathbf{x}\text{在H上} \\ <0, & \mathbf{x}\text{处于H下方R2区域} \end{cases}$$

- 权矢量**w**垂直于分类面H，指向R1区域

ISBN 978-7-5603-4763-9

4.1线性判别函数和线性分类界面

线性判别函数——断言



➢点到平面的距离

平面H方程：$g(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0 = 0$

点到平面H距离：$r = \dfrac{g(\mathbf{x})}{\|\mathbf{w}\|}$

原点到平面距离：

$$r_0 = \frac{g(\mathbf{0})}{\|\mathbf{w}\|} = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{0} + w_0}{\|\mathbf{w}\|} = \frac{w_0}{\|\mathbf{w}\|}$$

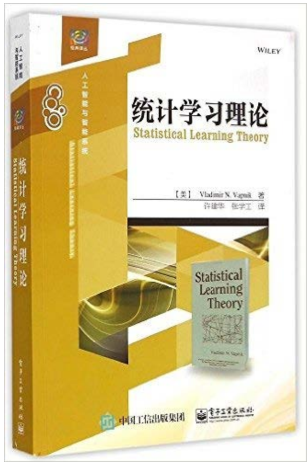
ISBN 978-7-5603-4763-9

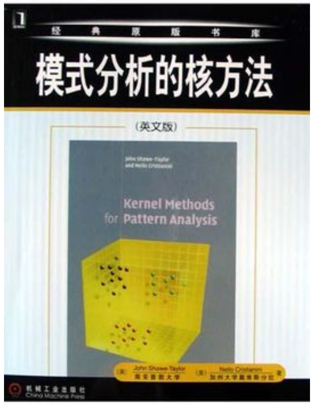
Vapnik VS Yann LeCun

ISBN 978-7-5603-4763-9

支持向量机
(SVM, Support Vector Machine)

曾经，最常用的分类方法！





统计学习理论，Vapnik　　　**模式分析的核方法，泰勒**

ISBN 978-7-5603-4763-9

支持向量机
(SVM, Support Vector Machine)

基本思想：

将特征“映射”到高维空间中去，寻找最优分类面。

关键问题：

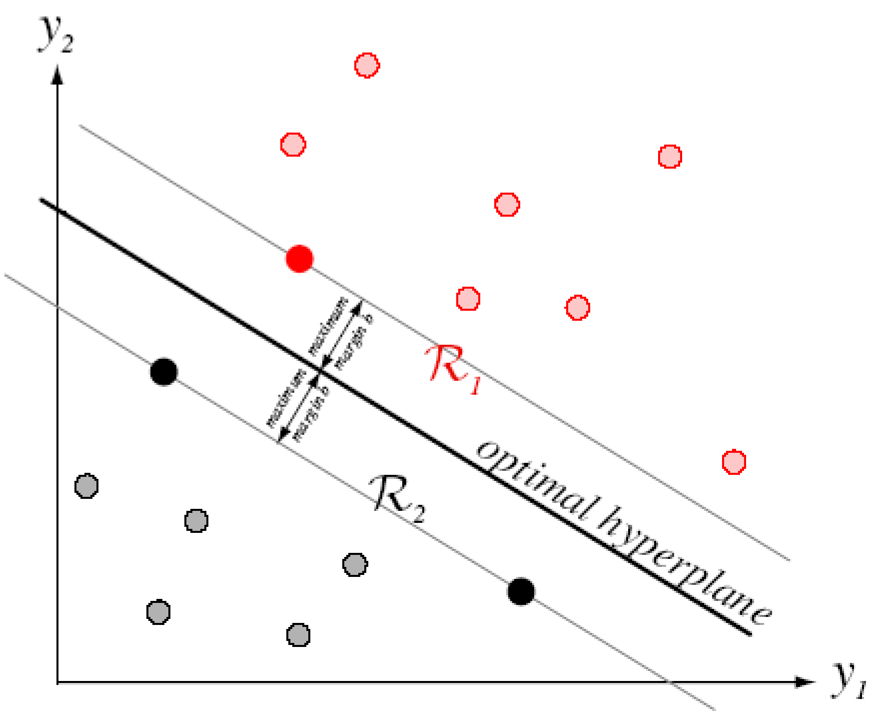
最优分类面：定义和计算最优分类界面、保证泛化能力。

核方法：利用核函数进行隐式“映射”。

线性不可分问题：软间隔与折中。

优化方法：如何搜索最优解

1最优线性判别函数



最优分类面：

距离样本点“越远”越安全、泛化能力越强；

在两类样本中间“不偏不倚”，错误率最低。

支持向量：

到分类面由距离最近的样本“支持”，称之为支持向量；

各支持向量到分类面的距离相等。

ISBN 978-7-5603-4763-9

1最优线性判别函数

分类面方程：

$$g(\mathbf{x}) = \mathbf{w}^t\mathbf{x} + w_0 = 0$$

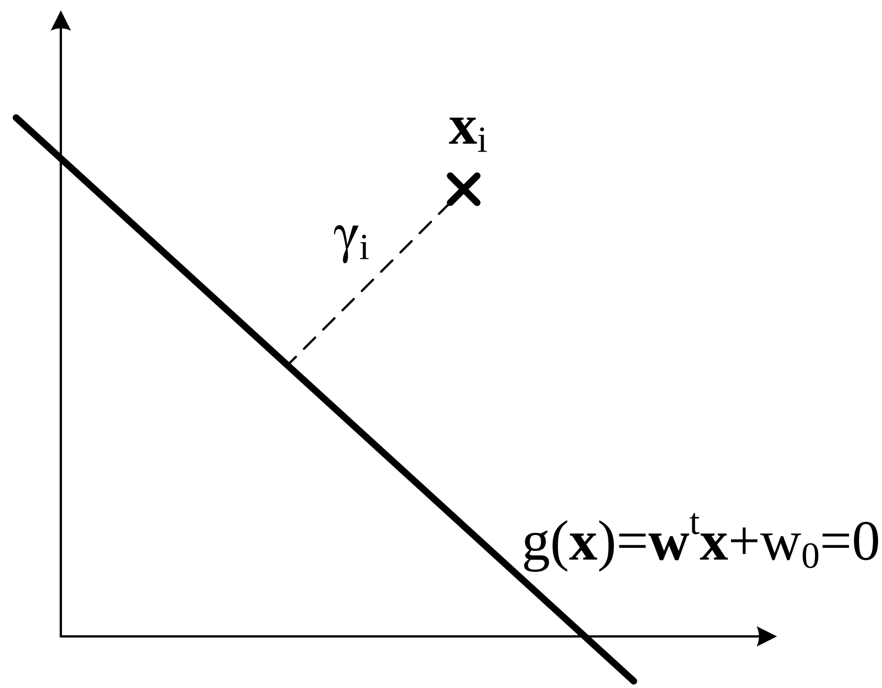
样本点 $\mathbf{x}_i$ 到分类面的间隔：

函数间隔：

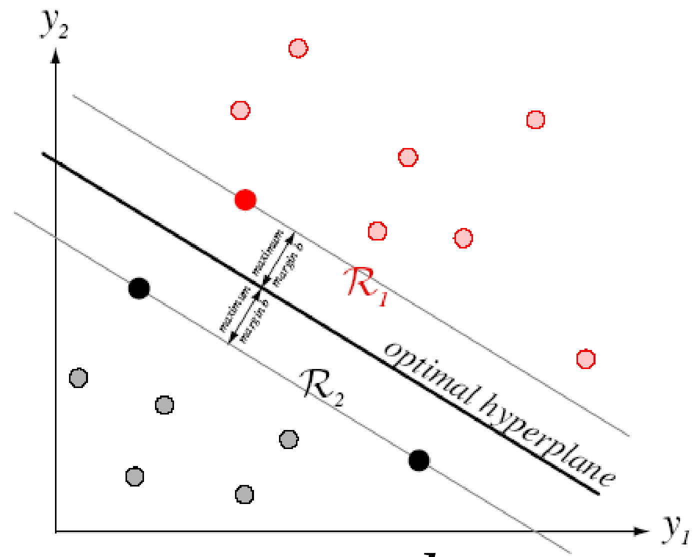
$$b_i = |g(\mathbf{x}_i)| = |\mathbf{w}^t\mathbf{x}_i + w_0|$$

几何间隔：

$$\gamma_i = \frac{b_i}{\|\mathbf{w}\|}$$



2支持向量机的学习



样本：$\{(\mathbf{x}, z)_n\}$，其中$\mathbf{x}$为特征，

$z \in \{-1,1\}$为类别标签

目标：$z\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}+w_0\right) \geq b_{\min}$的前提下，最大化$\gamma = \dfrac{b_{\min}}{\|\mathbf{w}\|}$

推导：$\mathbf{w}, w_0$同乘$b_{\min}$不改变分类面，有：$z\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}+w_0\right) \geq 1$

此时样本到分类面的长度为：$\dfrac{\left|\mathbf{w}^{\mathrm{T}}\mathbf{x}+w_0\right|}{\|\mathbf{w}\|} \geq \dfrac{1}{\|\mathbf{w}\|}$

优化：在$z\left(\mathbf{w}^{t}\mathbf{x}+w_0\right) \geq 1$的条件下，最小化$\|\mathbf{w}\|$

使得等号成立的x，即为支持向量

ISBN 978-7-5603-4763-9

2支持向量机的学习

样本集到分类界面的几何间隔：

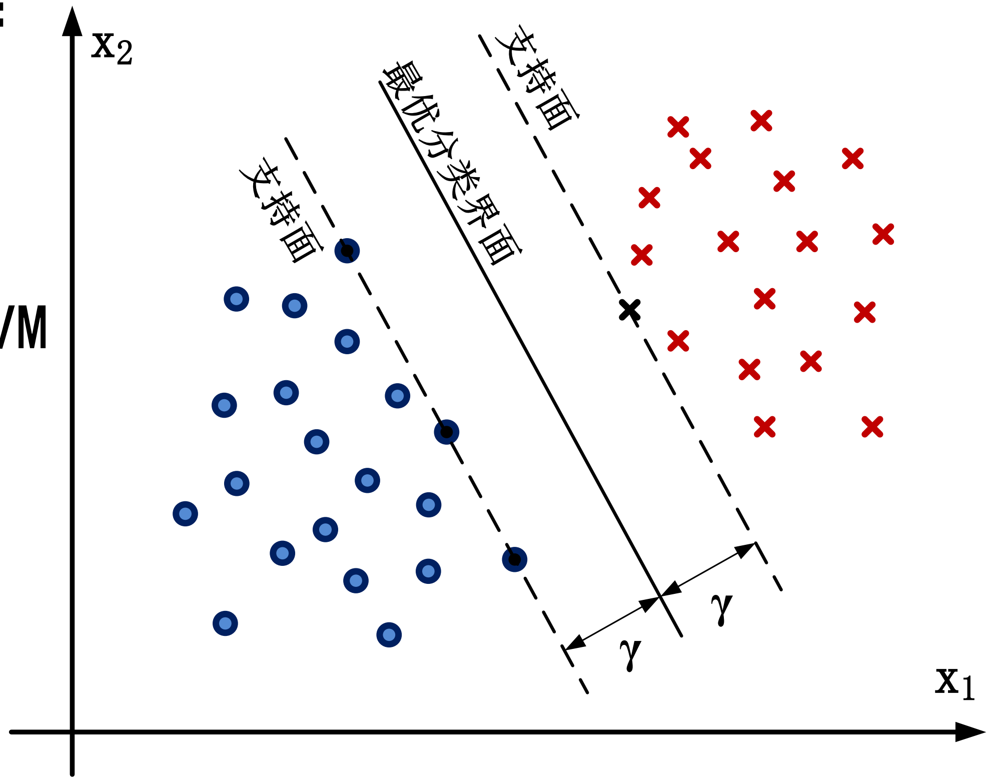
$$\gamma = \frac{1}{\|\mathbf{w}\|}$$

γ 最大，亦即$\|\mathbf{w}\|$最小，所以SVM可以变为如下优化问题：

在 $z_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + w_0\right) \geq 1$ 的条件下

最小化准则函数：

$$J_{SVM} = \frac{1}{2}\|\mathbf{w}\|^2$$



不等式约束下的函数极值：

Kuhn-Tucker构造法

转化为对偶问题

ISBN 978-7-5603-4763-9

函数极值

1629 Fermat 定理： 无约束条件下的函数极值

1788 Lagrange乘子法：等式约束下的函数极值

1951 Kuhn-Tucker定理：

（凸）不等式约束下，最小化（凸）目标函数

$$z_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + w_0\right) \geq 1$$

$$J_{SVM} = \frac{1}{2}\|\mathbf{w}\|^2$$

ISBN 978-7-5603-4763-9

凸优化——Kuhn-Tucker定理

凸集合：集合中任意两点的连线都属于该集合

凸函数：对于任意两点 x, y, Jensen不等式成立：

$$f\left(\alpha x+(1-\alpha)x\right)y \le \alpha f(x)+(1-\alpha)f(y)$$

凸优化： X是一个线性空间，A是空间的一个凸子集，

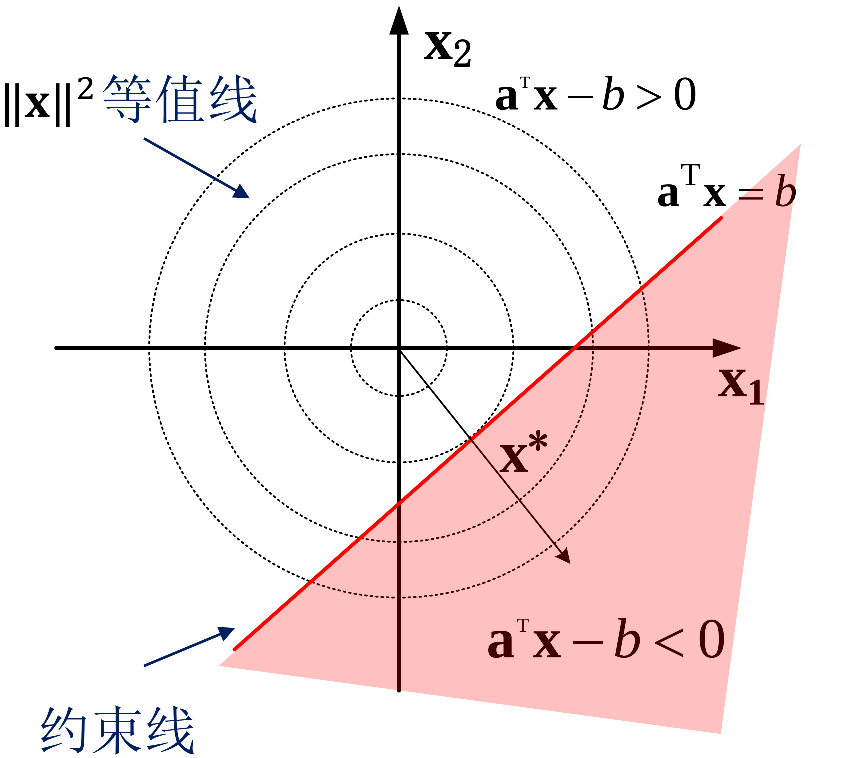
$f_k(\mathbf{x})$, k=0,1,…,m 是凸函数

最小化泛函: $f_0(x) \to \inf$

约束条件为: $x \in A$,

$$f_k(x) \le 0, k = 1, \cdots, m$$

例： $\min\limits_{\mathbf{x}\in R^2} f(\mathbf{x})=\|\mathbf{x}\|^2=x_1^2+x_2^2$， s.t. $g(\mathbf{x})=\mathbf{a}^{\mathrm{T}}\mathbf{x}-b\le 0$



等值线与约束线相切时，取得最小值。等值线的法向量 $\nabla f(\mathbf{x})$ 与约束线的法向量 $\nabla g(\mathbf{x})$ 方向相反

$$\begin{cases}\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}}=-\lambda\dfrac{\partial g(\mathbf{x})}{\partial \mathbf{x}}=-\lambda\mathbf{a}, & \lambda>0\\ g(\mathbf{x})=0\end{cases}$$

构造Lagrange函数

$$L(\mathbf{x},\lambda)=f(\mathbf{x})+\lambda g(\mathbf{x})$$

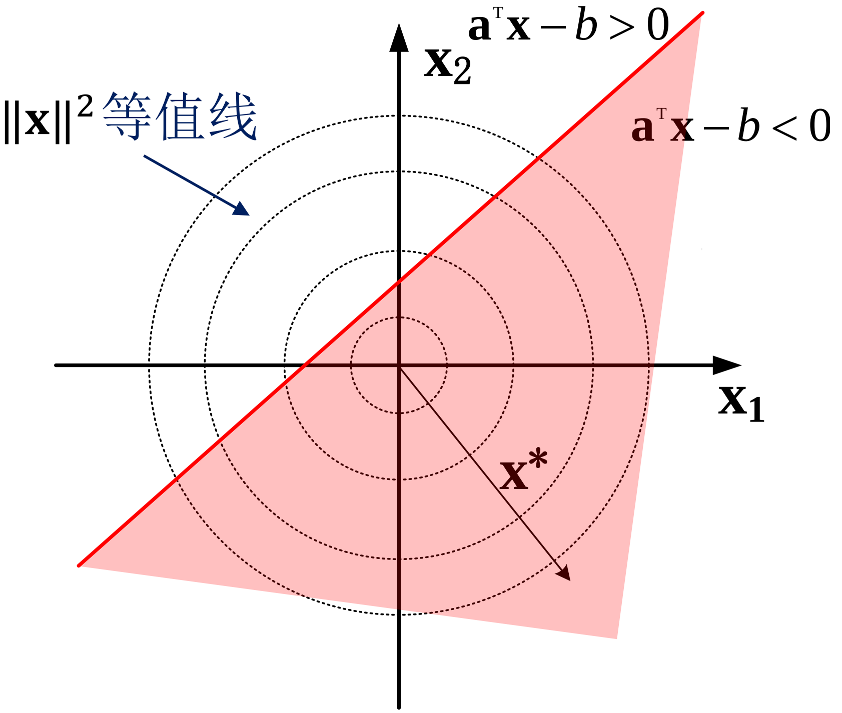
分别对 $\mathbf{x},\lambda$ 求导有

$$\begin{cases}\dfrac{\partial L(\mathbf{x},\lambda)}{\partial \mathbf{x}}=\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}}+\lambda\dfrac{\partial g(\mathbf{x})}{\partial \mathbf{x}}=0\\ \dfrac{\partial L(\mathbf{x},\lambda)}{\partial \lambda}=g(\mathbf{x})=0\end{cases}$$

如果 $f(\mathbf{x})$极值点在约束区域外，条件极值一定在约束边界上。可以通过构造Lagrange函数转化为无约束问题

ISBN 978-7-5603-4763-9

例： $\min_{\mathbf{x}\in R^2} f(\mathbf{x})=\|\mathbf{x}\|^2=x_1^2+x_2^2$， s.t. $g(\mathbf{x})=\mathbf{a}^{\mathrm{T}}\mathbf{x}-b\le 0$



如果 $f(\mathbf{x})$ 极值点在约束区域内，
也可构造Lagrange函数，

此时 $g(\mathbf{x})<0$， $\lambda=0$

如果 $f(\mathbf{x})$ 极值点在约束区域内，

$$\begin{cases}\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}}=0\\ g(\mathbf{x})<0\end{cases}$$

构造Lagrange函数

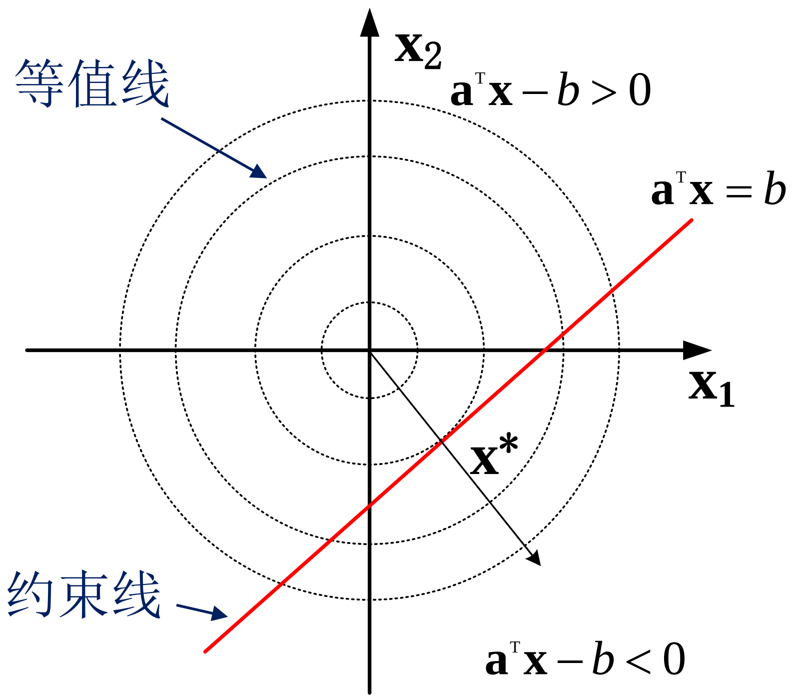
$$L(\mathbf{x},\lambda)=f(\mathbf{x})+\lambda g(\mathbf{x})$$

分别对 $\mathbf{x},\lambda$ 求导有

$$\begin{cases}\dfrac{\partial L(\mathbf{x},\lambda)}{\partial \mathbf{x}}=\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}}+\lambda\dfrac{\partial g(\mathbf{x})}{\partial \mathbf{x}}=0\\ \lambda=0\\ g(\mathbf{x})<0\end{cases}$$

ISBN 978-7-5603-4763-9

例： $\min\limits_{\mathbf{x}\in R^2} f(\mathbf{x})=\|\mathbf{x}\|^2=x_1^2+x_2^2,\quad \text{s.t.}\quad g(\mathbf{x})=\mathbf{a}^{\mathrm{T}}\mathbf{x}\text{-}b\le 0$



综合以上两种情况，构造Lagrange函数

$$L(\mathbf{x},\lambda)=f(\mathbf{x})+\lambda g(\mathbf{x})$$

有：

$$\begin{cases}\partial L(\mathbf{x},\boldsymbol{\lambda})/\partial\mathbf{x}=\mathbf{0}\\ \lambda\ge 0\\ \lambda g(\mathbf{x})=0\end{cases}$$

当不等式在极值点 $\mathbf{x}^*$ 上是以 $g(\mathbf{x})<0$ 的方式得到满足时，对应 $\lambda=0$；当以 $g(\mathbf{x})=0$ 的形式满足时，$\lambda>0$。

凸优化——Kuhn-Tucker定理

构建拉格朗日函数

令 $L=L(x,\lambda_0,\boldsymbol{\lambda})=\sum_{k=1}^{m}\lambda_k f_k(x),\quad \boldsymbol{\lambda}=(\lambda_1,\cdots\lambda_m)$

如果 x^* 在满足约束条件下最小化 $f_0(x)$，则存在不同时为0的 $\lambda_0^*,\boldsymbol{\lambda}^*=(\lambda_1^*,\cdots\lambda_m^*)$ 使下列三条件成立：

(a):最小值原理 $\min\limits_{x\in A} L(x,\lambda^*{}_0,\boldsymbol{\lambda}^*)=L(x^*,\lambda^*{}_0,\boldsymbol{\lambda}^*)$

如果L在$x*$处可导

则$\nabla L(x*)=0$

(b):非负性条件 $\lambda_k^*\geq 0$

(c):Kuhn-Tucker条件 $\lambda_k^* f_k(x^*)=0,k=1,\cdots m$

要么系数 λ_k^* 为0、要么 $f_k(x^*)=0$ 等号成立！

SVM的对偶问题

代价函数 $J_{SVM}=\frac{1}{2}\|\mathbf{w}\|^2$

约束条件 $z_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i+w_0\right)\geq 1,\quad i=1,2,\cdots,n$

构造Lagrange函数 $L\left(\mathbf{w},w_0,\boldsymbol{\alpha}\right)=\frac{1}{2}\|\mathbf{w}\|^2-\sum_{i=1}^{n}\alpha_i\left[z_i\left(\mathbf{w}^t\mathbf{x}_i+w_0\right)-1\right]$

Lagrange系数$\alpha_i\geq 0$

ISBN 978-7-5603-4763-9

约束优化的对偶问题

min-max问题
$$\begin{cases} f^*(\mathbf{u}) = \max_{\mathbf{v}} f(\mathbf{u},\mathbf{v}) \\ \min_{\mathbf{u}} f^*(\mathbf{u}) = \min_{\mathbf{u}} \max_{\mathbf{v}} f(\mathbf{u},\mathbf{v}) \end{cases}$$

max-min问题
$$\begin{cases} f^*(\mathbf{v}) = \min_{\mathbf{u}} f(\mathbf{u},\mathbf{v}) \\ \max_{\mathbf{v}} f^*(\mathbf{v}) = \max_{\mathbf{v}} \min_{\mathbf{u}} f(\mathbf{u},\mathbf{v}) \end{cases}$$

上述两个问题如果有解存在的话，必在同一点取得最优解

$$\min_{\mathbf{u}} \max_{\mathbf{v}} f(\mathbf{u},\mathbf{v}) = \max_{\mathbf{v}} \min_{\mathbf{u}} f(\mathbf{u},\mathbf{v}) = f(\mathbf{u}^*,\mathbf{v}^*)$$

两个优化问题的求解顺序可以颠倒！

SVM的对偶问题

原始Lagrange函数

$$L(\mathbf{w},w_0,\boldsymbol{\alpha})=\frac{1}{2}\|\mathbf{w}\|^2-\sum_{i=1}^{n}\alpha_i\left[z_i\left(\mathbf{w}^t\mathbf{x}_i+w_0\right)-1\right]$$

J_{svm}是Lagrange函数关于α的最大化

$$\max_{\boldsymbol{\alpha}}L(\mathbf{w},w_0,\boldsymbol{\alpha})=J_{SVM}(\mathbf{w},w_0)=\frac{1}{2}\|\mathbf{w}\|^2$$

J_{svm}函数的max-min对偶问题

$$\min_{\mathbf{w},w_0}J_{SVM}(\mathbf{w},w_0)$$

$$=\min_{\mathbf{w},w_0}\max_{\boldsymbol{\alpha}}L(\mathbf{w},w_0,\boldsymbol{\alpha})$$

$$=\max_{\boldsymbol{\alpha}}\min_{\mathbf{w},w_0}L(\mathbf{w},w_0,\boldsymbol{\alpha})$$

①构建Lagrange函数

②针对$\mathbf{w}$，w_0的最小值问题

③计算针对α的最大值问题

1，构造Lagrange函数：

$$L(\mathbf{w},w_0,\boldsymbol{\alpha})=\frac{1}{2}\|\mathbf{w}\|^2-\sum_{i=1}^{n}\alpha_i\left[z_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i+w_0\right)-1\right],\alpha_i\geq 0$$

2，针对w，w0的最小值问题：分别对参数w和w_0求导，令导数为0：

$$\frac{\partial L(\mathbf{w},w_0,\boldsymbol{\alpha})}{\partial\mathbf{w}}=\mathbf{w}-\sum_{i=1}^{n}z_i\alpha_i\mathbf{x}_i=0 \quad\longrightarrow\quad \mathbf{w}=\sum_{i=1}^{n}z_i\alpha_i\mathbf{x}_i \quad (1)$$

$$\frac{\partial L(\mathbf{w},w_0,\boldsymbol{\alpha})}{\partial w_0}=\sum_{i=1}^{n}z_i\alpha_i=0 \quad\longrightarrow\quad \sum_{i=1}^{n}z_i\alpha_i=0 \quad (2)$$

3，针对α的最大值问题：将(1),(2)代人 $L(\mathbf{w},w_0,\boldsymbol{\alpha})$

$$\max_{\boldsymbol{\alpha}} L(\boldsymbol{\alpha})=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j z_i z_j \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j$$

约束条件： $\sum_{i=1}^{n}z_i\alpha_i=0,\ \ \alpha_i\geq 0,\ \ i=1,\cdots,n$

SVM约束优化的对偶问题
——凸二次规划

ISBN 978-7-5603-4763-9

转化为对偶问题的优点

$$\max_{\boldsymbol{\alpha}} L(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j z_i z_j \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j$$

不直接优化权值矢量：

与样本的特征维数无关，只与样本的数量有关。
当样本的特征维数很高时，对偶问题更容易求解

只用内积，不要坐标：

训练样本只以内积的形式出现
优化求解过程中，只需计算内积，不需要知道每一维特征

样本的内积矩阵：

n个样本，两两内积，构成 n x n 矩阵G

$$\mathbf{G}_{i,j} = \left(z_i \mathbf{x}_i\right)^{\mathrm{T}} \left(z_j \mathbf{x}_j\right)$$

仅根据内积矩阵能解决多少模式识别问题?

内积符号 $$<\mathbf{x},\mathbf{y}>\triangleq \mathbf{x}^{\mathrm{T}}\mathbf{y}\triangleq \sum_n x_n y_n$$

内积加权和 $$<\sum_i \alpha_i \mathbf{x}_i, \sum_j \beta_j \mathbf{y}_j> = \sum_i \sum_j \alpha_i \beta_j <\mathbf{x}_i, \mathbf{y}_j>$$

训练样本的内积矩阵 $$G \triangleq \left(< z_i \mathbf{x}_i, z_j \mathbf{x}_j >\right)_{n\times n}$$

信息瓶颈！

二次规划
$$\max_{\boldsymbol{\alpha}} L(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j z_i z_j \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j$$
$$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j \left\langle z_i \mathbf{x}_i, z_j \mathbf{x}_j \right\rangle$$
$$= \mathbf{c}\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{G}\boldsymbol{\alpha}$$

仅利用样本间的内积信息进行识别！

G 为对称矩阵时，叫做二次规划(Quadratic Programming, QP)

ISBN 978-7-5603-4763-9

凸二次规划

$$L(\boldsymbol{\alpha})=\mathbf{c}\boldsymbol{\alpha}-\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}G\boldsymbol{\alpha}$$

G正定时目标函数为凸函数，线性约束下可行域又是凸集，称为凸二次规划

$$v^{t}Gv=\sum_{i,j=1}^{n}v_{i}v_{j}G_{ij}=\sum_{i,j=1}^{n}v_{i}v_{j}<\mathbf{x}_{i},\mathbf{x}_{j}>$$

$$=\left\langle\sum_{i=1}^{n}v_{i}\mathbf{x}_{i},\sum_{j=1}^{n}v_{j}\mathbf{x}_{j}\right\rangle=\left\|\sum_{j=1}^{n}v_{j}\mathbf{x}_{j}\right\|^{2}\geq 0$$

凸二次规划性质：

K-T条件是最优解的充要条件；

局部最优解就是全局最优解。

哈尔滨工业大学 计算机学院
模式识别与智能系统研究中心
ISBN 978-7-5603-4763-9

SVM解的讨论

不等式约束条件下的凸二次优化问题，解法基础是Kuhn-Tucker定理；

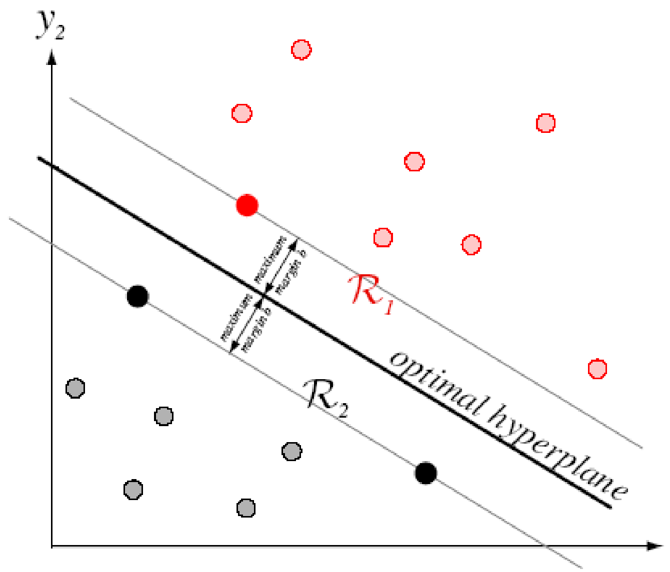
首先求解n个Lagrange乘子，n为训练样本数。根据Kuhn-Tucker定理，有：

$$\lambda_k^* f_k(x^*) = 0 \quad \rightarrow \quad \alpha_i \left[z_i \left(\mathbf{w}^{\mathrm{T}} \mathbf{x}_i + w_0 \right) - 1 \right] = 0$$

$$\begin{cases} \alpha_i = 0, & z_i \left(\mathbf{w}^{\mathrm{T}} \mathbf{x}_i + w_0 \right) - 1 > 0 \\ \alpha_i > 0, & \underline{z_i \left(\mathbf{w}^{\mathrm{T}} \mathbf{x}_i + w_0 \right) - 1 = 0} \end{cases}$$

满足该条件的$\mathbf{x}_i$称为支持矢量。

只有支持向量的 α_i 才不为0



SVM解的讨论

权矢量 w：根据找到的支持矢量 $\mathbf{x}_i$ 以及相应的Lagrange乘子 α_i，计算权矢量 $\mathbf{w}$：

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i z_i \mathbf{x}_i$$

只有支持向量的α_i不为0
对W起到“支持”，
其他向量对W不产生影响

偏置w_0：可以用支持矢量满足的条件求得

$$z_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + w_0\right) = 1$$

ISBN 978-7-5603-4763-9

练习：SVM分类器训练结果如下：

第一类样本：

$$\mathbf{x}_1=(1,1)^{\mathrm{T}},\mathbf{x}_2=(2,0)^{\mathrm{T}}$$

第二类样本：

$$\mathbf{x}_3=(1,0)^{\mathrm{T}},\mathbf{x}_4=(0,1)^{\mathrm{T}}$$

对应Lagrange乘子：

$$\alpha_1=3,\alpha_2=1,\alpha_3=3,\alpha_4=1$$

求线性判别函数，并判别(2, 2)的类别属性

根据找到的支持矢量$\mathbf{x}_i$及相应的Lagrange乘子α_i，计算权矢量$\mathbf{w}$：

$$\mathbf{w}=\sum_{i=1}^{n}\alpha_i z_i \mathbf{x}_i$$

偏置$\mathbf{w}_0$可以用支持矢量满足的条件求得：

$$z_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i+w_0\right)=1$$

ISBN 978-7-5603-4763-9

新手作业：SVM具有重要的理论及应用价值，请完成：

1，SVM 中“最优分类面”的基本思想及数学模型 。
2，什么是对偶问题？如何理解以下公式：

$$\max_{\boldsymbol{\alpha}} L(\mathbf{w}, w_0, \boldsymbol{\alpha}) = J_{SVM}(\mathbf{w}, w_0) = \frac{1}{2}\|\mathbf{w}\|^2$$

3，参考PPT及相关书籍，完成从原始优化问题到对偶问题的推导。

$$\min_{\mathbf{w}, w_0} J_{SVM}(\mathbf{w}, w_0) = \frac{1}{2}\|\mathbf{w}\|^2$$

$$\text{s.t.} \quad z_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + w_0\right) \geq 1, i = 1, \cdots, n$$

$\longrightarrow$

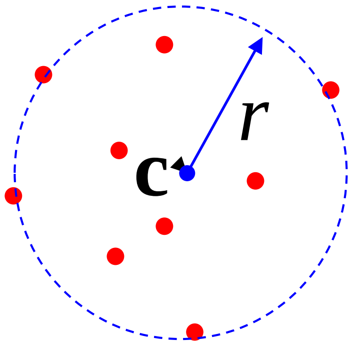
$$\max_{\boldsymbol{\alpha}} L(\boldsymbol{\alpha}) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j z_i z_j \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j$$

$$\text{s.t.} \quad \alpha_i \geq 0,\ i = 1, \cdots, n$$

$$\sum_{i=1}^{n}\alpha_i z_i = 0$$

4，转化为对偶问题后，有哪些优点?

高手作业：包含点集合的最小超球体



样本集：$S=\{\mathbf{x}_1,\cdots\mathbf{x}_\ell\}$

寻找一个包含所有样本的最小超球体：

$$\min_{c,r}\ r^2$$

$$s.t.\ \|\mathbf{x}_i-\mathbf{c}\|^2-r^2\le 0$$

1，仿照本章SVM的推导，构建Lagrange函数

$$L(\mathbf{c},r,\boldsymbol{\alpha})=r^2+\sum_{i=1}^{\ell}\alpha_i\left[\|\mathbf{x}_i-\mathbf{c}\|^2-r^2\right]$$

推导其关于$\boldsymbol{\alpha}$的凸二次规划形式。

2，设$\boldsymbol{\alpha}$最优解为 $\boldsymbol{\alpha}^*=\{\alpha_1^*,\cdots,\alpha_\ell^*\}$ 求半径 r 与球心坐标 $\mathbf{c}$。

ISBN 978-7-5603-4763-9

SVM应用过程中需要解决的问题

不可分问题：增强分类面的稳定性、处理不可分情况

软间隔： $z_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i+w_0\right)\geq 1-\xi_i$

非线性问题：如何映射到更高维度的空间中，增强分类能力

“到”高维空间中做内积——核方法

大规模优化问题：大规模数据集应用中，如何进行寻优

根据应用规模选择不同优化方法

多类别问题：如何将2分类推广到多分类

间接应用：1对多，1对1， 直接推广：多分类SVM

ISBN 978-7-5603-4763-9

软间隔SVM

$$[\mathbf{w}, w_0] = \arg\min_{\mathbf{w}, w_0} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi^p$$

$$z_i\left(\mathbf{w}^t\mathbf{x}_i + w_0\right) \geq 1 - \xi_i, \qquad \xi_i > 0$$

松弛变量

惩罚系数C：

C越大，对错分的惩罚力度越大

C越小，获得的分类面间隔越大

软间隔系数p：

p=1 L1-soft margin

p=2 L2-soft margin

ISBN 978-7-5603-4763-9

	硬间隔SVM	软间隔SVM
目标	$[\mathbf{w}, w_0] = \arg\min_{\mathbf{w}, w_0} \frac{1}{2}\lVert\mathbf{w}\rVert^2$	$[\mathbf{w}, w_0] = \arg\min_{\mathbf{w}, w_0} \frac{1}{2}\lVert\mathbf{w}\rVert^2 + C\sum_i \xi^p$
条件	$z_i\left(\mathbf{w}^t\mathbf{x}_i + w_0\right) \geq 1$	$z_i\left(\mathbf{w}^t\mathbf{x}_i + w_0\right) \geq 1 - \xi_i$ $\xi_i \geq 0$
L函数	$L(\mathbf{w}, w_0, \boldsymbol{\alpha}) = \frac{1}{2}\lVert\mathbf{w}\rVert^2$ $-\sum_{i=1}^{n}\alpha_i\left[z_i\left(\mathbf{w}^t\mathbf{x}_i + w_0\right) - 1\right]$	$L(\mathbf{w}, w_0, \boldsymbol{\alpha}) = \frac{1}{2}\lVert\mathbf{w}\rVert^2 + C\sum_i \xi^p$ $-\sum_{i=1}^{n}\alpha_i\left[z_i\left(\mathbf{w}^t\mathbf{x}_i + w_0\right) - 1 + \xi_i\right]$ $-\sum_{i=1}^{n}\beta_i\xi_i$

1，构建Lagrange函数

$$L\left(\mathbf{w}, w_0, \boldsymbol{\alpha}, \boldsymbol{\beta}\right) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i} \xi^p - \sum_{i=1}^{n} \alpha_i \left[z_i \left(\mathbf{w}^{\mathrm{T}} \mathbf{x}_i + w_0 \right) - 1 + \xi_i \right] - \sum_{i=1}^{n} \beta_i \xi_i$$

2，针对w，w_0，ξ_i 的最小值问题：对参数w、w_0、ξ_i 求导，令导数为0：

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{n} \alpha_i z_i \mathbf{x}_i = 0 \quad \rightarrow \quad \mathbf{w} = \sum_{i=1}^{n} \alpha_i z_i \mathbf{x}_i \qquad (1)$$

$$\frac{\partial L}{\partial w_0} = -\sum_{i=1}^{n} \alpha_i z_i = 0 \quad \rightarrow \quad \sum_{i=1}^{n} \alpha_i z_i = 0 \qquad (2)$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \beta_i = 0 \quad \rightarrow \quad \beta_i = C - \alpha_i > 0 \qquad (3)$$

3，针对α的最大值问题：将(1),(2), (3)代人 $L\left(\mathbf{w}, w_0, \boldsymbol{\alpha}, \boldsymbol{\beta}\right)$

$$L\left(\mathbf{w}, w_0, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}\right) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j z_i z_j \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j$$

约束条件：$\sum_{i=1}^{n} z_i \alpha_i = 0, \quad 0 \le \alpha_i \le C$ → 与硬间隔的唯一不同

软间隔SVM解的讨论

α取值：Kuhn-Tucker定理可以证明

$$\begin{cases} z_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i+w_0\right)>1, & \alpha_i=0 \quad \text{支持面内} \\ z_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i+w_0\right)=1, & C>\alpha_i>0 \quad \text{支持面上} \\ z_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i+w_0\right)<1, & \alpha_i=C \quad \text{支持面外} \end{cases}$$

此时 $\xi_i>0$，根据KT条件 $\beta_i=C-\alpha_i=0$

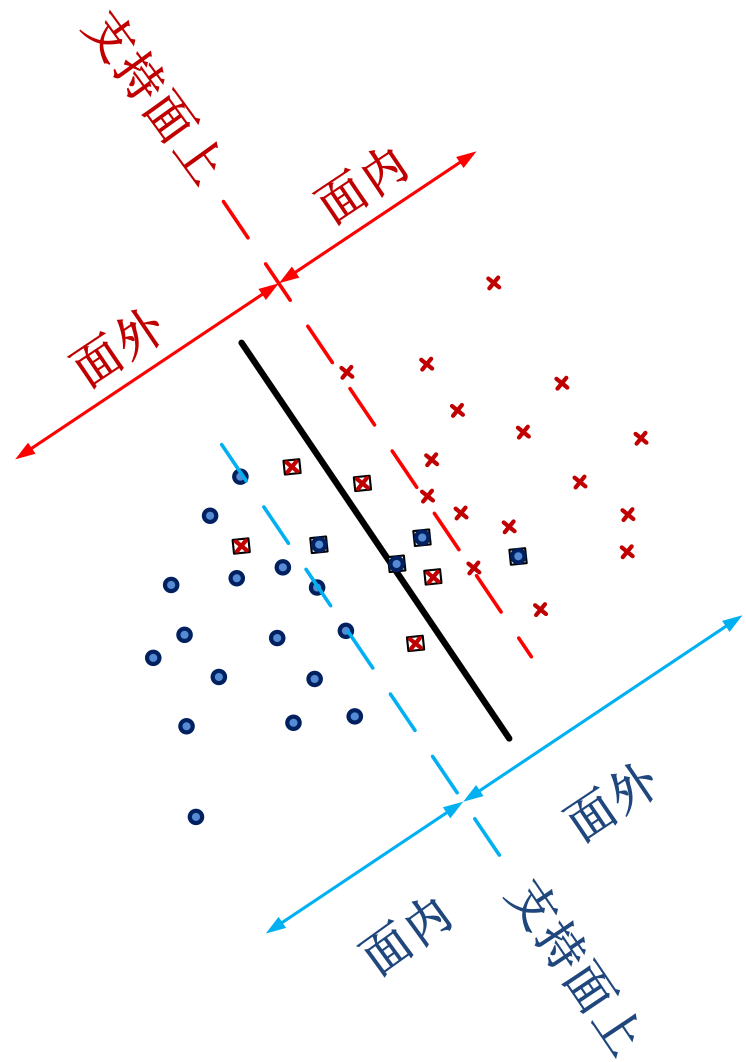
支持向量：所有对应 $\alpha_i>0$ 的训练样本

分类面权向量w：训练样本加权和

$$\mathbf{w}=\sum_{i=1}^{n}\alpha_i z_i \mathbf{x}_i$$

偏置w_0：可以用支持面上的支持向量求得

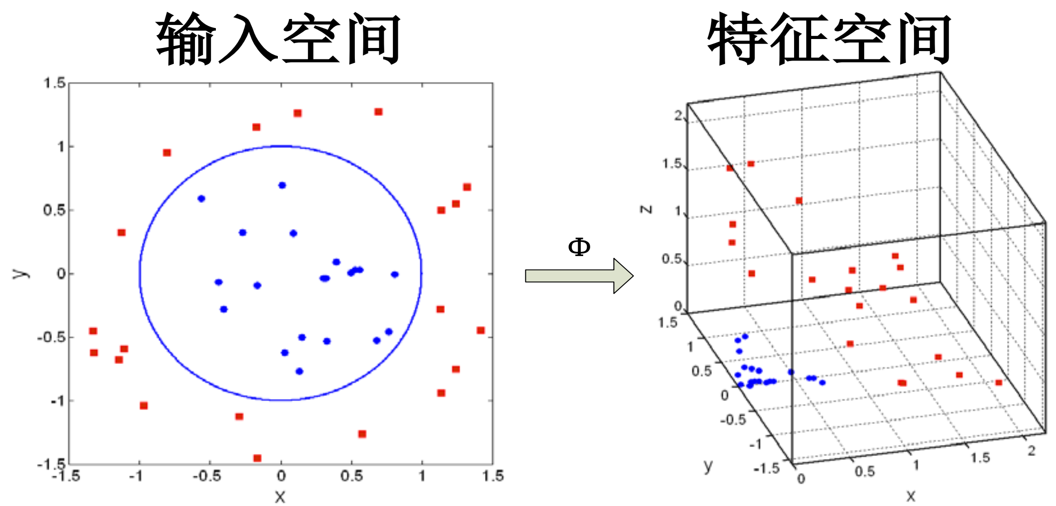
$$z_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i+w_0\right)=1$$



哈尔滨工业大学 计算机学院
模式识别与智能系统研究中心
ISBN 978-7-5603-4763-9

3核函数与非线性支持向量机

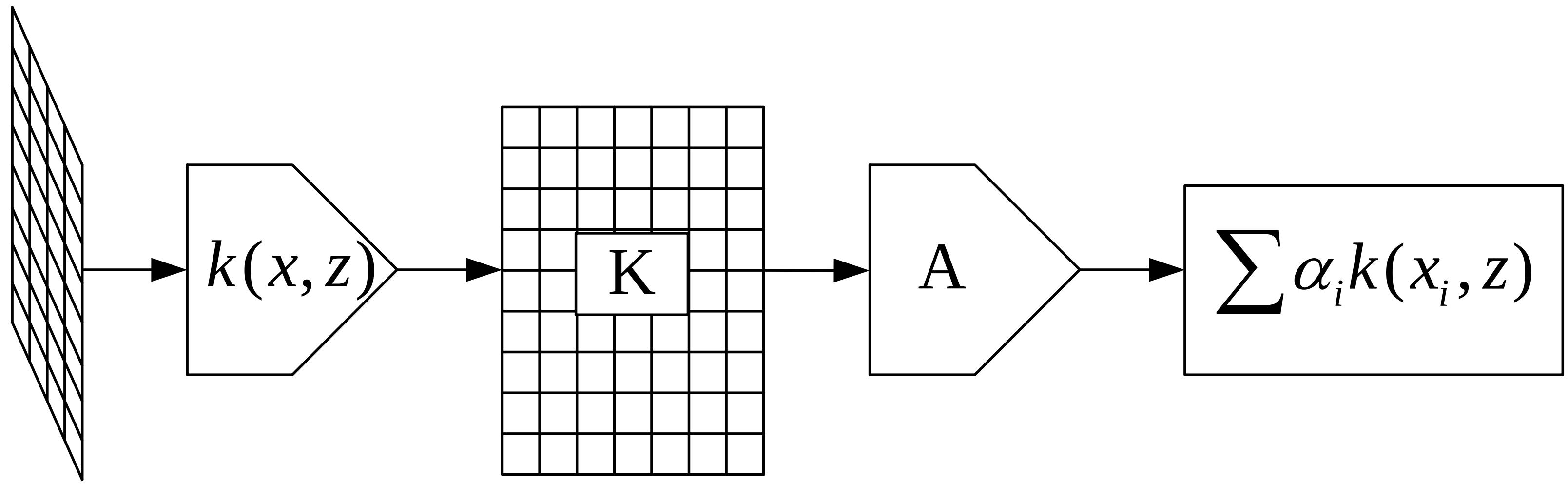
广义线性判别函数： 将低维特征向量映射到高维空间中，学习线性判别函数



核函数(Kernel Function)： 内积的推广，低维空间中的核函数，相当于高维空间中的内积

核方法： 利用核函数，避免直接在高维空间中计算，实现非线性识别方法。

核方法概述



数据　核函数　核矩阵　PA算法　模式函数

信息瓶颈，
滤去坐标，只剩内积

核方法的4个关键：

数据嵌入特征空间

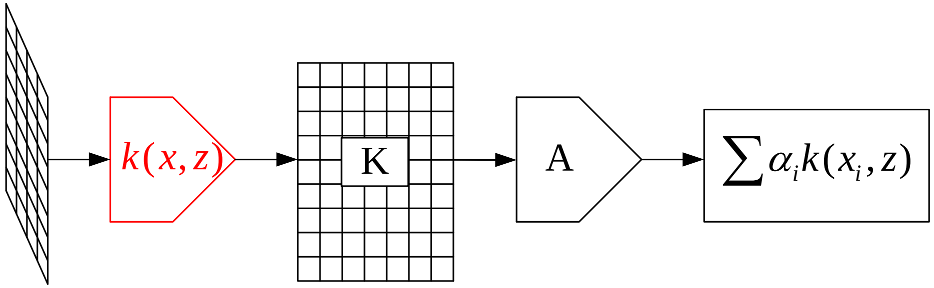
在特征空间中寻找线性模式

在嵌入空间中，不需要计算点的坐标，只用两两内积

利用核函数，可以直接从初始数据高效地计算内积。

ISBN 978-7-5603-4763-9

核方法概述



1）核函数：

核函数定义、性质、常用形式

2）核矩阵

3）核方法的基础算法

4）核方法举例

核函数

函数作用：特征空间中两个矢量之间的内积可以通过定义输入空间中的核函数直接计算得到。

$$\kappa(\mathbf{x},\mathbf{z}) = <\phi(\mathbf{x}),\phi(\mathbf{z})>$$

实现方法：不必定义非线性映射Φ，而直接在输入空间中定义核函数K来完成非线性映射。

应用条件：

1. 定义的核函数K能够对应于特征空间中的内积；
2. 识别方法中不需要计算特征空间中的矢量本身，而只须计算特征空间中两个矢量的内积。

ISBN 978-7-5603-4763-9

例：
$$K(\mathbf{x},\mathbf{y})=\left(\mathbf{x}^{\mathrm{T}}\mathbf{y}\right)^{2}=\langle\mathbf{x},\mathbf{y}\rangle^{2}$$

R^2→R^3非线性映射：

$$\Phi:\mathbf{x}=(x_1,x_2)\mapsto\Phi(\mathbf{x})=(x_1^2,x_2^2,\sqrt{2}x_1x_2)\in F\subseteq\mathbb{R}^3$$

R^2中核函数K，相当于Φ映射到R^3后再计算内积：

$$\langle\Phi(\mathbf{x}),\Phi(\mathbf{y})\rangle=\left(x_1^2,\sqrt{2}x_1x_2,x_2^2\right)^{\mathrm{T}}\left(y_1^2,\sqrt{2}y_1y_2,y_2^2\right)$$
$$=x_1^2y_1^2+\sqrt{2}x_1x_2\sqrt{2}y_1y_2+x_2^2y_2^2=\left(x_1y_1+x_2y_2\right)^2=\langle\mathbf{x},\mathbf{y}\rangle^2$$

特征空间不由核函数唯一确定：同一核函数，对应的映射并不唯一，例如$\langle\mathbf{x},\mathbf{y}\rangle^2$还可以对应：

$$\Phi:\mathbf{x}=(x_1,x_2)\mapsto\Phi(\mathbf{x})=(x_1^2,x_2^2,x_1x_2,x_2x_1)\in F\subseteq\mathbb{R}^4$$

核函数性质

Mercer定理：如果一个对称函数$K(\mathbf{x},\mathbf{y})$针对任意的平方可积函数 $g(\mathbf{y})$满足半正定条件，即为核函数。

$$\iint K(\mathbf{x},\mathbf{y})g(\mathbf{x})g(\mathbf{y})d\mathbf{x}d\mathbf{y} > 0 \qquad \int g^2(\mathbf{x})d\mathbf{x} < \infty$$

核的基本运算：如果κ_1,κ_2是核，$\mathbf{B}$是一个半正定矩阵，$p(x)$是一个正系数多项式，那么下面都是核:

$$\kappa(\mathbf{x},\mathbf{z}) = \kappa_1(\mathbf{x},\mathbf{z}) + \kappa_2(\mathbf{x},\mathbf{z}) \qquad \kappa(\mathbf{x},\mathbf{z}) = \mathbf{x}^{\mathrm{T}} B\mathbf{x}$$

$$\kappa(\mathbf{x},\mathbf{z}) = \alpha\kappa_1(\mathbf{x},\mathbf{z}) \qquad \kappa(\mathbf{x},\mathbf{z}) = p\left(\kappa_1(\mathbf{x},\mathbf{z})\right)$$

$$\kappa(\mathbf{x},\mathbf{z}) = \kappa_1(\mathbf{x},\mathbf{z})\kappa_2(\mathbf{x},\mathbf{z}) \qquad \kappa(\mathbf{x},\mathbf{z}) = \exp\left(\kappa_1(\mathbf{x},\mathbf{z})\right)$$

常用的核函数

Gaussian RBF：

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

Polynomial：

$$K(\mathbf{x},\mathbf{y})=\left(\left(\mathbf{x}^{\mathrm{T}}\mathbf{y}\right)+1\right)^d$$

Sigmoidal：

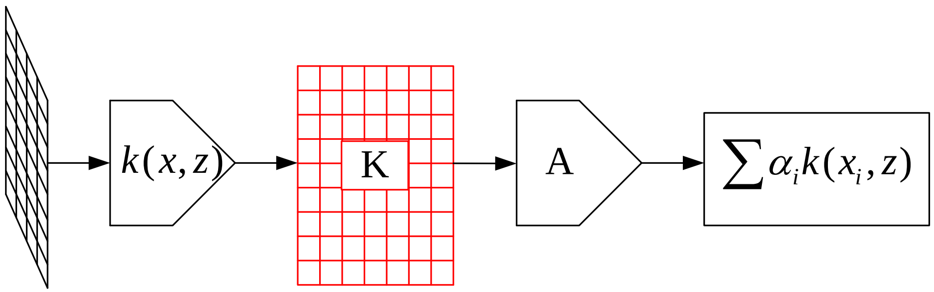
$$K(\mathbf{x},\mathbf{y})=\tanh\left(\alpha\left(\mathbf{x}^{\mathrm{T}}\mathbf{y}\right)+\theta\right)$$

Inv. Multiquardric：

$$K(\mathbf{x},\mathbf{y})=\frac{1}{\sqrt{\left\|\mathbf{x}-\mathbf{y}\right\|^2+c^2}}$$

ISBN 978-7-5603-4763-9

核方法概述



1）核函数

2）核矩阵

3）核方法的基础算法

4）核方法举例

训练样本的核矩阵

核函数： $$\kappa(\mathbf{x},\mathbf{z}) = <\phi(\mathbf{x}),\phi(\mathbf{z})>$$

其中 ϕ 是从 X 到（内积）特征空间 F 的一个映射：

$$\phi : \mathbf{x} \mapsto \phi(\mathbf{x}) \in F$$

特征矩阵： n 个训练样本在特征空间 F 中映射，记为

$$\mathbf{X} = \left[\phi(\mathbf{x}_1),\cdots,\phi(\mathbf{x}_n)\right]^{\mathrm{T}}$$

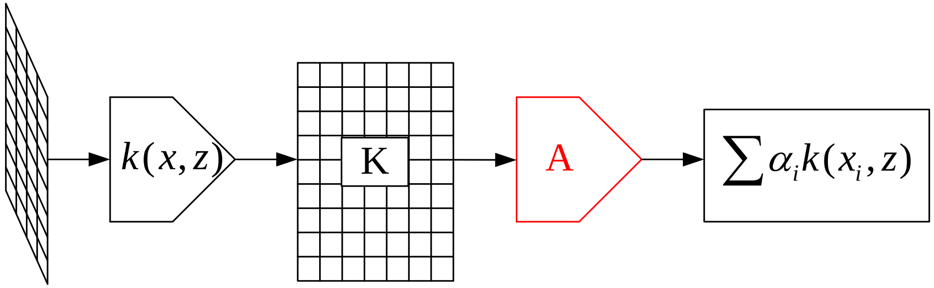
核矩阵： $\mathbf{X}$ 的 $n\times n$ Gram矩阵为核矩阵

$$\mathbf{K} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$$

$$\mathbf{K}_{ij} = \left\langle \phi(\mathbf{x}_i),\phi(\mathbf{x}_j)\right\rangle = \kappa(\mathbf{x}_i,\mathbf{x}_j)$$

——内积矩阵的推广，在原始样本中提取到的全部信息

核方法概述



1）核函数

2）核矩阵

3）核方法的基础算法

利用核函数，间接计算特征空间中范数、距离、均值、方差…

4）核方法举例

核方法的基础算法

特征向量的范数:

$$\left\|\phi(\mathbf{x})\right\|_2 = \sqrt{\left\langle \phi(\mathbf{x}), \phi(\mathbf{x}) \right\rangle} = \sqrt{\kappa(\mathbf{x}, \mathbf{x})}$$

特征向量的规范化内积:

$$\hat{\phi}(\mathbf{x}) = \frac{\phi(\mathbf{x})}{\left\|\phi(\mathbf{x})\right\|}$$

$$\hat{\kappa}(\mathbf{x}, \mathbf{z}) = \left\langle \hat{\phi}(\mathbf{x}), \hat{\phi}(\mathbf{z}) \right\rangle = \left\langle \frac{\phi(\mathbf{x})}{\left\|\phi(\mathbf{x})\right\|}, \frac{\phi(\mathbf{z})}{\left\|\phi(\mathbf{z})\right\|} \right\rangle$$

$$= \frac{\left\langle \phi(\mathbf{x}), \phi(\mathbf{z}) \right\rangle}{\left\|\phi(\mathbf{x})\right\| \left\|\phi(\mathbf{z})\right\|} = \frac{\kappa(\mathbf{x}, \mathbf{z})}{\sqrt{\kappa(\mathbf{x}, \mathbf{x})\kappa(\mathbf{z}, \mathbf{z})}}$$

核方法的基础算法

特征向量线性组合的范数:

$$
\begin{aligned}
\left\|\sum_{i=1}^{\ell}\alpha_i\phi(\mathbf{x}_i)\right\|^2 &= \left\langle \sum_{i=1}^{\ell}\alpha_i\phi(\mathbf{x}_i), \sum_{j=1}^{\ell}\alpha_j\phi(\mathbf{x}_j) \right\rangle \\
&= \sum_{i=1}^{\ell}\alpha_i\sum_{j=1}^{\ell}\alpha_j\left\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \right\rangle \\
&= \sum_{i,j=1}^{\ell}\alpha_i\alpha_j\kappa(\mathbf{x}_i, \mathbf{x}_j)
\end{aligned}
$$

核方法的基础算法

特征向量之间的距离:

$$\left\|\phi(\mathbf{x})-\phi(\mathbf{z})\right\|^2=\left\langle\phi(\mathbf{x})-\phi(\mathbf{z}),\phi(\mathbf{x})-\phi(\mathbf{z})\right\rangle$$

$$=\left\langle\phi(\mathbf{x}),\phi(\mathbf{x})\right\rangle-2\left\langle\phi(\mathbf{x}),\phi(\mathbf{z})\right\rangle+\left\langle\phi(\mathbf{z}),\phi(\mathbf{z})\right\rangle$$

$$=\kappa(\mathbf{x},\mathbf{x})-2\kappa(\mathbf{x},\mathbf{z})+\kappa(\mathbf{z},\mathbf{z})$$

ISBN 978-7-5603-4763-9

核方法的基础算法

质心（特征均值）的范数

$$\phi_s = \frac{1}{\ell}\sum_{i=1}^{\ell}\phi(\mathbf{x}_i)$$

$$\|\phi_s\|^2 = \left\langle \frac{1}{\ell}\sum_{i=1}^{\ell}\phi(\mathbf{x}_i), \frac{1}{\ell}\sum_{j=1}^{\ell}\phi(\mathbf{x}_j) \right\rangle$$

$$= \frac{1}{\ell^2}\sum_{i,j=1}^{\ell}\left\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \right\rangle = \frac{1}{\ell^2}\sum_{i,j=1}^{\ell}\kappa(\mathbf{x}_i, \mathbf{x}_j)$$

质心的范数的平方=核矩阵元素的平均值

ISBN 978-7-5603-4763-9

核方法的基础算法

点到质心（特征均值）的距离

$$\phi_s = \frac{1}{\ell}\sum_{i=1}^{\ell}\phi(\mathbf{x}_i)$$

$$\left\|\phi(\mathbf{x})-\phi_s\right\|^2 = \left\langle\phi(\mathbf{x}),\phi(\mathbf{x})\right\rangle + \left\langle\phi_s,\phi_s\right\rangle - 2\left\langle\phi(\mathbf{x}),\phi_s\right\rangle$$

$$= \kappa(\mathbf{x},\mathbf{x}) + \frac{1}{\ell^2}\sum_{i,j=1}^{\ell}\kappa(\mathbf{x}_i,\mathbf{x}_j) - \frac{2}{\ell}\sum_{i=1}^{\ell}\kappa(\mathbf{x},\mathbf{x}_j)$$

哈尔滨工业大学 计算机学院
模式识别与智能系统研究中心
ISBN 978-7-5603-4763-9

核方法的基础算法

特征方差

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\left\|\phi(\mathbf{x}_i)-\phi_s\right\|^2$$

$$=\frac{1}{\ell}\sum_{k=1}^{\ell}\kappa(\mathbf{x}_k,\mathbf{x}_k)+\frac{1}{\ell^2}\sum_{i,j=1}^{\ell}\kappa(\mathbf{x}_i,\mathbf{x}_j)-\frac{2}{\ell^2}\sum_{i,k=1}^{\ell}\kappa(\mathbf{x}_i,\mathbf{x}_k)$$

$$=\frac{1}{\ell}\sum_{k=1}^{\ell}\kappa(\mathbf{x}_k,\mathbf{x}_k)-\frac{1}{\ell^2}\sum_{i,j=1}^{\ell}\kappa(\mathbf{x}_i,\mathbf{x}_j)$$

核矩阵对角线元素平均值 - 全体元素平均值

核方法的基础算法

中心化数据: 把原点移到质心,使平均特征值最小化

$$\hat{\phi}(\mathbf{x}) = \phi(\mathbf{x}) - \phi_s = \phi(\mathbf{x}) - \frac{1}{\ell}\sum_{i=1}^{\ell}\phi(\mathbf{x}_i)$$

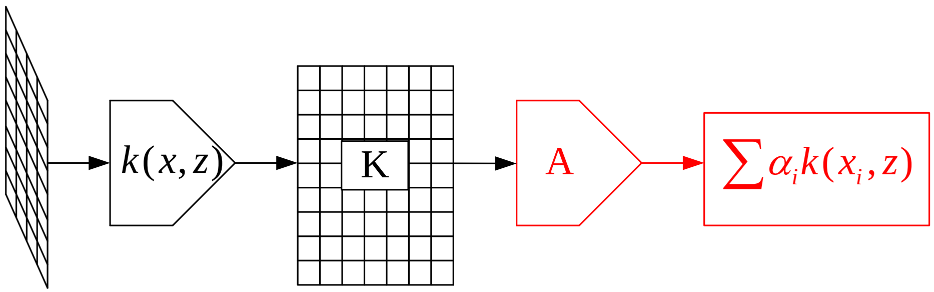
移动后，新的核函数为

$$\hat{\kappa}(\mathbf{x},\mathbf{z}) = \left\langle \hat{\phi}(\mathbf{x}), \hat{\phi}(\mathbf{z}) \right\rangle$$

$$= \left\langle \phi(\mathbf{x}) - \frac{1}{\ell}\sum_{i=1}^{\ell}\phi(\mathbf{x}_i), \phi(\mathbf{z}) - \frac{1}{\ell}\sum_{i=1}^{\ell}\phi(\mathbf{x}_i) \right\rangle$$

$$= \kappa(\mathbf{x},\mathbf{z}) - \frac{1}{\ell}\sum_{i=1}^{\ell}\kappa(\mathbf{x},\mathbf{x}_i) - \frac{1}{\ell}\sum_{j=1}^{\ell}\kappa(\mathbf{x}_j,\mathbf{z}) + \frac{1}{\ell^2}\sum_{i,j=1}^{\ell}\kappa(\mathbf{x}_i,\mathbf{x}_j)$$

核方法概述



1）核函数

2）核矩阵

3）核方法的基础算法

4）核方法举例

距离分类器

核PCA

非线性SVM

ISBN 978-7-5603-4763-9

距离分类器

将训练集S划分为两个正例、负例子集：S_-，S_+

计算测试点 **x** 到两子集质心的距离：

$$d_+(\mathbf{x}) = \left\| \phi(\mathbf{x}) - \phi_{S+} \right\| \qquad d_-(\mathbf{x}) = \left\| \phi(\mathbf{x}) - \phi_{S-} \right\|$$

分类规则： $$h(\mathbf{x}) = \operatorname{sgn}\left(\left\| \phi(\mathbf{x}) - \phi_{S-} \right\|^2 - \left\| \phi(\mathbf{x}) - \phi_{S+} \right\|^2 \right)$$

$$= \operatorname{sgn}\left(\kappa(\mathbf{x},\mathbf{x}) + \underbrace{\frac{1}{\ell_-^2} \sum_{i,j=1}^{\ell-} \kappa(\mathbf{x}_i,\mathbf{x}_j)}_{b+} - \frac{2}{\ell_-} \sum_{i=1}^{\ell-} \kappa(\mathbf{x},\mathbf{x}_i) \right.$$

$$\left. -\kappa(\mathbf{x},\mathbf{x}) - \underbrace{\frac{1}{\ell_+^2} \sum_{i,j=\ell_-+1}^{\ell} \kappa(\mathbf{x}_i,\mathbf{x}_j)}_{b-} + \frac{2}{\ell_+} \sum_{i=\ell_-+1}^{\ell} \kappa(\mathbf{x},\mathbf{x}_i) \right)$$

$$= \operatorname{sgn}\left(\frac{1}{\ell_+} \sum_{i=\ell_-+1}^{\ell} \kappa(\mathbf{x},\mathbf{x}_i) - \frac{1}{\ell_-} \sum_{i=1}^{\ell-} \kappa(\mathbf{x},\mathbf{x}_i) - b \right)$$

ISBN 978-7-5603-4763-9

非线性SVM

原目标函数：$L(\boldsymbol{\alpha})=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j\underline{\left\langle z_i\mathbf{x}_i, z_j\mathbf{x}_j\right\rangle}$

内积矩阵

非线性映射：$L(\boldsymbol{\alpha})=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\left(\alpha_i\alpha_j\left\langle\Phi(z_i\mathbf{x}_i),\Phi(z_j\mathbf{x}_j)\right\rangle\right)$

$$=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\left(\alpha_i\alpha_j\underline{K(z_i\mathbf{x}_i, z_j\mathbf{x}_j)}\right)$$

核矩阵

判别函数：$g(\mathbf{x})=\mathbf{w}^t\Phi(\mathbf{x})+w_0$

$$=\underline{\sum_{i=1}^{n}\left(z_i\alpha_i\Phi(\mathbf{x}_i)\right)^t}\,\Phi(\mathbf{x})+w_0$$

$$=\sum_{i=1}^{n}z_i\alpha_i\underline{K(\mathbf{x}_i,\mathbf{x})}+w_0$$

$\mathbf{w}=\sum_{i=1}^{n}z_i\alpha_i\Phi(\mathbf{x}_i)$

判别面权向量

不需要计算$\Phi(x)$

但需保留所有支持向量

ISBN 978-7-5603-4763-9

SVM的使用、配置

matlab（2012a）自带函数svmtrain svmclassify：

SVMStruct = svmtrain(Training,Group,Name,Value)

Group = svmclassify(SVMStruct,Sample)

Training：训练样本集合，每行代表一个样本

Group：样本标签，每行代表一个样本标签

Name,Value ：参数配置

Sample：测试样本集合

ISBN 978-7-5603-4763-9

SVM中的常用核

- 多项式Polynomial核

$$\kappa(\mathbf{x},\mathbf{z}) = \left(<\mathbf{x},\mathbf{z}> + R\right)^d$$

特征空间维度：$\binom{n+d}{d}$

增加R会减小高次多项式的权重

SVM中的常用核

- RBF核 （径向基核、高斯核）

$$\kappa(\mathbf{x},\mathbf{z}) = \exp\left(-\|\mathbf{x}-\mathbf{z}\|^2 / 2\sigma^2\right)$$

特征空间具有无限维度

σ 越小，高次特征衰减慢，拟合能力越强

如何使用核函数？

n为特征数，m为训练样本数。

m<<n时: 线性SVM

n较小，而且m大小中等：例如n在 1-1000 之间，而m在10-10000之间，使用高斯核函数的支持向量机。

如果n较小，而m较大：例如n在1-1000之间，而m大于50000，则使用支持向量机会非常慢。

- 创造、增加更多的特征，然后使用逻辑回归或不带核函数的支持向量机。
- 神经网络

ISBN 978-7-5603-4763-9

SVM中的优化方法

- 于中等规模数据集
  - smo （Sequential Minimal Optimizer）

- 大规模数据集
  - *SVMlight ：http://svmlight.joachims.org/*
  - 也许该用深度学习了☺

根据待解决问题的规模及 软件包的说明，选择不同优化方法

SVM的使用、配置

matlab（2012a）自带函数svmtrain svmclassify

训练关键1）——核函数选择与参数配置

svmtrain(xdata,group, 'kernel_function' , 'polynomial', 'polyorder' ,3);

使用多项式核　　多项式次数为3

svmtrain(xdata,group,'kernel_function' , 'rbf', 'rbf_sigma',1);

使用rbf核　　σ 系数为1

训练关键2）——boxconstraint 软间隔惩罚系数C,默认为1

训练关键3）——method 优化算法，默认为’smo’

SVM多类别分类

- OAA (one against all)
- OAO (one against one) 投票
- *直接构造多类别SVM*

ISBN 978-7-5603-4763-9

多类别SVM

$$(\mathbf{W},\mathbf{w}_0,\boldsymbol{\xi}) = \arg\min_{\mathbf{W},\mathbf{b}} \frac{1}{2}\sum_{z\in Z}\left(\|\mathbf{w}_z\|^2 + b_z^2\right) + C\sum_{i\in I}\sum_{z\in Z\setminus\{z\}}(\xi)^p$$

满足:

$$<\mathbf{w}_{zi},\mathbf{x}_i> + b_{zi} - (<\mathbf{w}_z,\mathbf{x}_i> + b_z) \geq 1 - \xi_i^z, z \in Z\setminus\{z_i\}$$

$$\xi_i^z \geq 0, z \in Z\setminus\{z_i\}$$

训练样本$T_{\mathrm{XY}} = \{(y_i, z_i)\}$

权矩阵$\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_c]$,

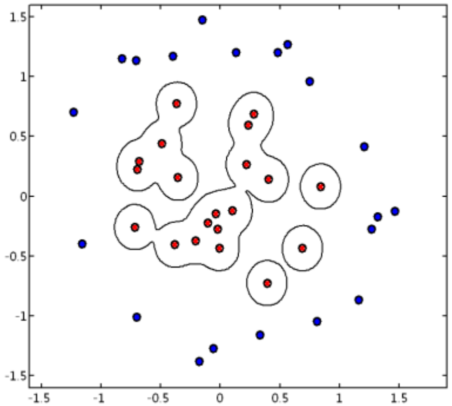
样本序数$i \in I = [1, 2, \cdots, l]$

偏置向量$\mathbf{b} = [w_{01}, w_{02}, \cdots, w_{0c}]^t$

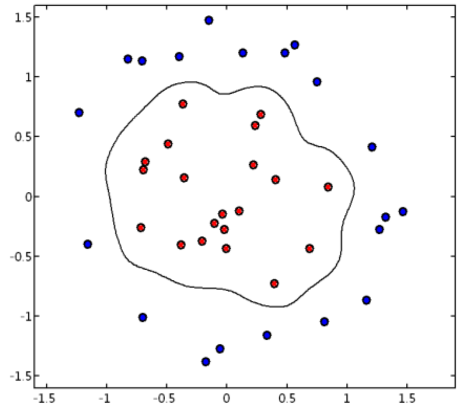
样本类别$z \in Z = [1, 2, \cdots, c]$

ISBN 978-7-5603-4763-9

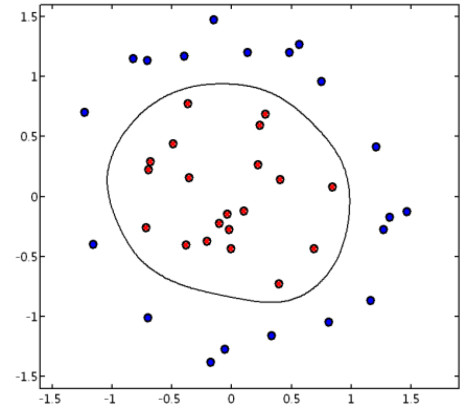
SVM的分类面



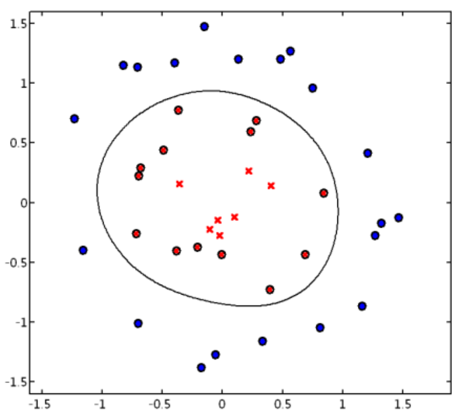
$\sigma = 0.1, C = 1$



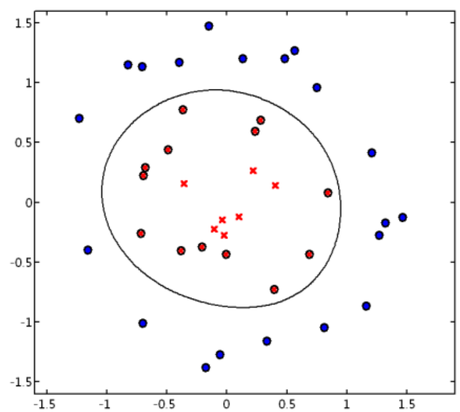
$\sigma = 0.25, C = 1$



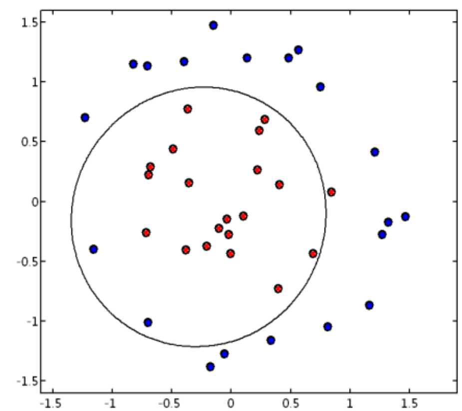
$\sigma = 0.5, C = 1$



$\sigma = 0.75, C = 1$



$\sigma = 1, C = 1$



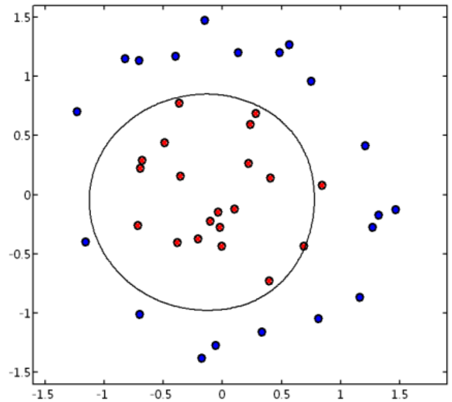
$\sigma = 3, C = 1$

正例：半径为0.9的圆内；　　反例：半径为1.2~1.5的圆环

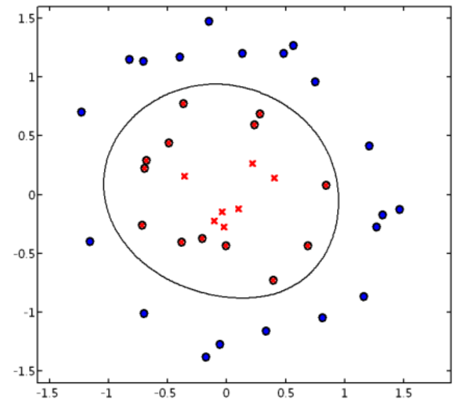
RBF核SVM分类，不同参数对分类面影响

ISBN 978-7-5603-4763-9

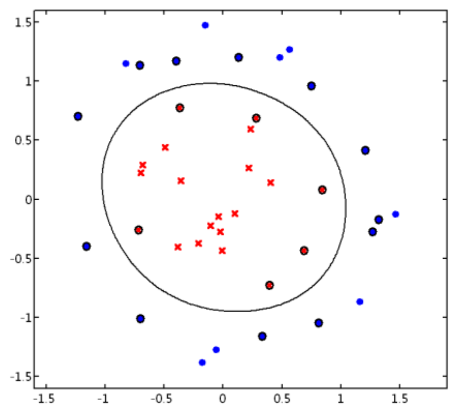
SVM的分类面



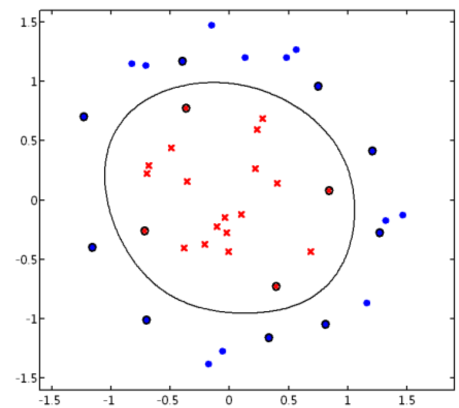
$\sigma = 1, C = 0.01$



$\sigma = 1, C = 1$



$\sigma = 1, C = 10$



$\sigma = 1, C = 100$

正例：半径为0.9的圆内；　反例：半径为1.2~1.5的圆环

RBF核SVM分类，不同参数对分类面影响

新手作业：手写数字识别+

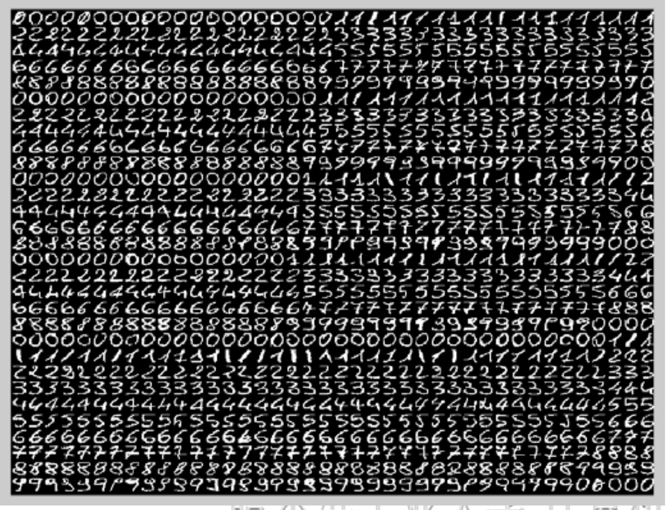
- 用Matlab或Python实现该字符集的识别
  - 代码+注释，规范命名
  - 采用SVM进行字符识别？
  - 如何选定核函数，如何设定参数？

UCI 数据集

http://archive.ics.uci.edu/ml/

Semeion Handwritten Digit Data Set

新手作业**自愿**完成，按组打包，下周1前发到群文件



ISBN 978-7-5603-4763-9

高手作业：

- 阅读《模式分析的核方法》7.1 节
  The smallest enclosing hypersphere
- 实现一个新颖性检测算法
  - 仅以某一类数字样本为训练样本
  - 输入一个待检测样本，判断是否与训练样本同类

  - 如何设计和测试实验结果？
  - 分析结果、如何改进？

ISBN 978-7-5603-4763-9