线性判别

Linear Discriminant

电子科技大学

信息与通信工程学院

主要内容

- 线性判别分类(Linear Discriminants for Classification)
- 分类举例
- 线性判别函数
  - 两类问题
  - 多类问题
- 支持矢量机(Support Vector Machine)
- Fisher线性判别
- 自适应增强AdaBoost(Adaptive Boosting)

主要内容

- 线性判别分类(Linear Discriminants for Classification)
- 分类举例
- 线性判别函数
  - 两类问题
  - 多类问题
- Fisher线性判别
- 支持矢量机(Support Vector Machine)
- 自适应增强AdaBoost(Adaptive Boosting)

线性判别函数

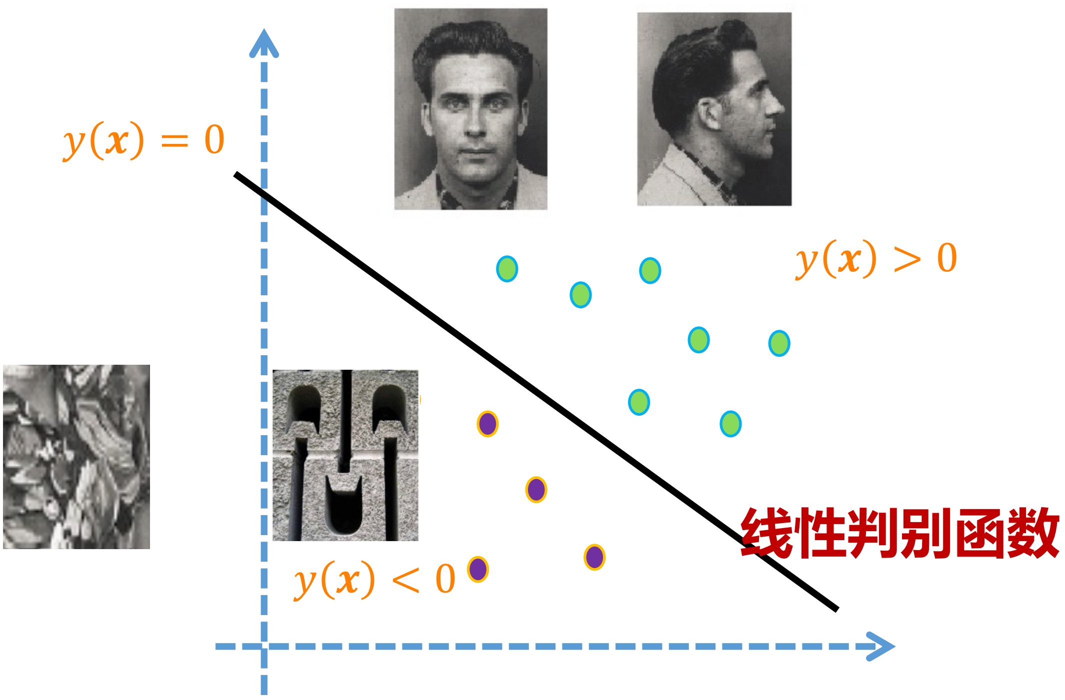
- 最简单的判别函数形式：输入变量x的线性组合，系数即为模型参数。linear combination

$$y = w_1 x_1 + ... + w_D x_D + b = \mathbf{w}^T \mathbf{x} + b$$

- 通过输入变量的固定非线性函数变换后进行线性组合，产生广义线性判别。generalized linear discriminants.

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

线性判别函数



$$y(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + w_0$$

主要内容

- 线性判别分类(Linear Discriminants for Classification)
- 分类举例
- 线性判别函数
  - 两类问题
  - 多类问题
- Fisher线性判别
- 支持矢量机(Support Vector Machine)
- 自适应增强AdaBoost(Adaptive Boosting)

线性判别函数

- 两类情况：

- 判别函数可以表示为x的线性函数：

$$y(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b$$

这里D-维向量 w 表示为权向量（weight vector)， 参数b为偏差 (or 有时也称 “−b" as a bias).

- 几何解释

  - 在D维$\boldsymbol{x}$空间中，判别边界 $y(\boldsymbol{x}) = 0$ 对应一个 $(D-1)$维超平面(hyperplane).

  - 权向量$\boldsymbol{w}$为超平面内任意向量的法线。

  - 偏差b确定了超平面在$\boldsymbol{x}$空间的位置。

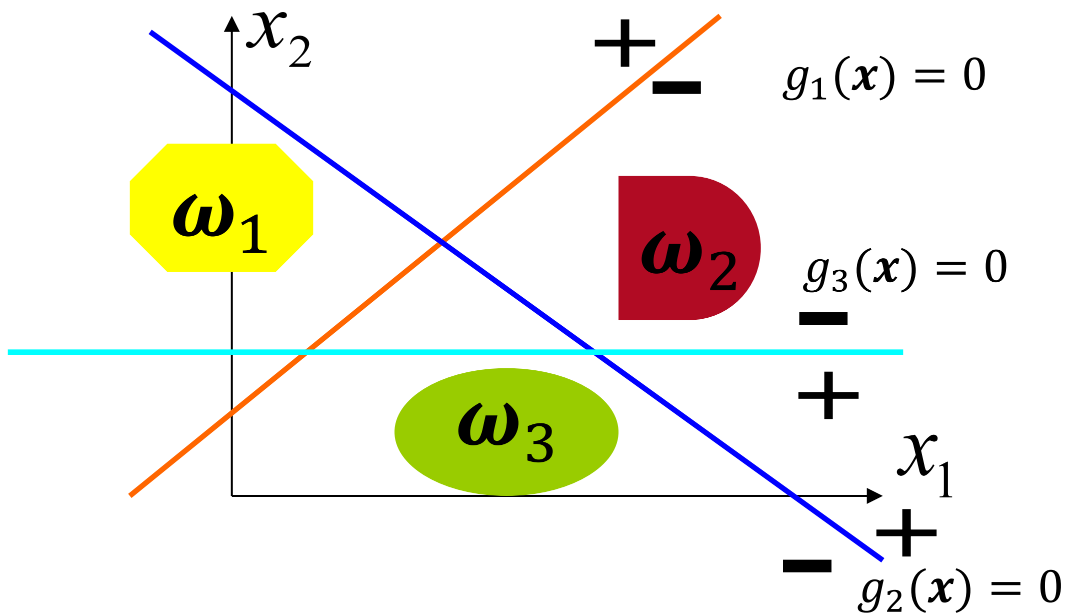
线性判别函数

- 多类情况：
  - K类: $C_1, C_2,\ldots, C_K$
  - 三种情形:
    - 对每一类都采用两类判别

      $\Rightarrow \boldsymbol{x}$ belongs to class C_1 or not?
    - 对每两类采用两类判别

      $\Rightarrow \boldsymbol{x}$ belongs to class C_i or C_j ?
    - Case 2的特例

- 第一种情况举例：
  - 右图所示，每一类别可用单个判别边界与其它类别相分开。
  - 如果一模式$\boldsymbol{x}$属于$\boldsymbol{C_1}$，则由图可清楚看出:这时$g_1(\boldsymbol{x}) > 0$而$g_2(\boldsymbol{x}) < 0$，$g_3(\boldsymbol{x}) < 0$。C_1类与其它类之间的边界由$g_1(\boldsymbol{x}) = 0$确定



线性判别函数

- 例：已知三类C_1, C_2, C_3的判别函数分别为：

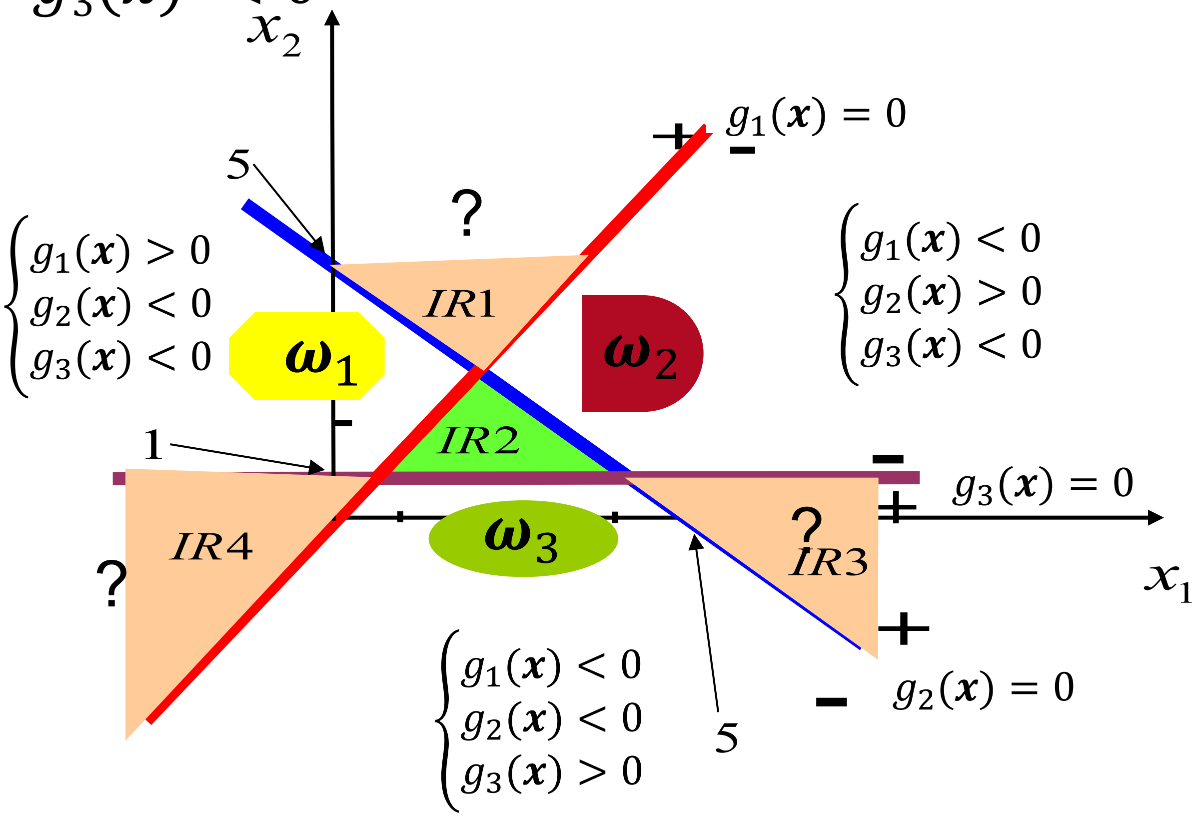
$$\begin{cases} g_1(\boldsymbol{x}) = -x_1 + x_2 \\ g_2(\boldsymbol{x}) = x_1 + x_2 - 5 \\ g_3(\boldsymbol{x}) = -x_2 + 1 \end{cases}$$

- 因此三个判别边界为：

$$\begin{cases} g_1(\boldsymbol{x}) = -x_1 + x_2 = 0 \\ g_2(\boldsymbol{x}) = x_1 + x_2 - 5 = 0 \\ g_3(\boldsymbol{x}) = -x_2 + 1 = 0 \end{cases}$$

线性判别函数

- 对于任一模式$\boldsymbol{x}$如果它的 $g_1(\boldsymbol{x})>0$， $g_2(\boldsymbol{x})<0$， $g_3(\boldsymbol{x})<0$



- 如果某个X使二个以上的判别函数 $g_i(\boldsymbol{x})>0$ 。则此模式X就无法作出确切的判决。如图中**IR1,IR3，IR4**区域。
- 另一种情况是**IR2**区域，判别函数都为负值。**IR1，IR2，IR3，IR4**。都为不确 定区域。

$$\begin{cases} g_1(\boldsymbol{x})=-x_1+x_2 \\ g_2(\boldsymbol{x})=x_1+x_2-5 \\ g_3(\boldsymbol{x})=-x_2+1 \end{cases}$$

线性判别函数

❖问： 当$\boldsymbol{x}=(x_1,x_2)^T=(6,5)^T$时属于哪一类？

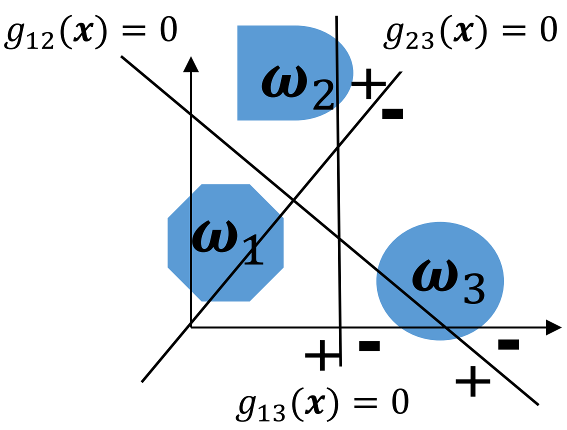
代入判别函数方程组：
$$\begin{cases} g_1(\boldsymbol{x})=-x_1+x_2 \\ g_2(\boldsymbol{x})=x_1+x_2-5 \\ g_3(\boldsymbol{x})=-x_2+1 \end{cases}$$

得： $g_1(\boldsymbol{x})=-1, g_2(\boldsymbol{x})=6, g_3(\boldsymbol{x})=-4.$

❖结论： $g_1(\boldsymbol{x})<0$， $g_2(\boldsymbol{x})>0$， $g_3(\boldsymbol{x})<0$所以它属于$\boldsymbol{C}_2$类

第二种情况举例

- 每个模式类和其它模式类间可分别用判别平面分开(M 种模式)。
  - 这样 有 $M(M-1)/2$个判别平面。
  - 对于两类问题，$M=2$，则有一个判别平面。
  - 同理，三类问题则有三个判别平面。
  - 判别函数： $g_{ij}(\boldsymbol{x})=\boldsymbol{\omega}_{ij}^{T}\boldsymbol{x}+b_{ij}$
  - 判别边界： $g_{ij}(\boldsymbol{x})=0$
  - 判别条件： $g_{ij}(\boldsymbol{x})\begin{cases}>0\rightarrow 当\boldsymbol{x}\in\boldsymbol{\omega}_{\mathrm{i}}\\<0\rightarrow 当\boldsymbol{x}\in\boldsymbol{\omega}_{\mathrm{j}}\end{cases}\quad i\neq j$



线性判别函数

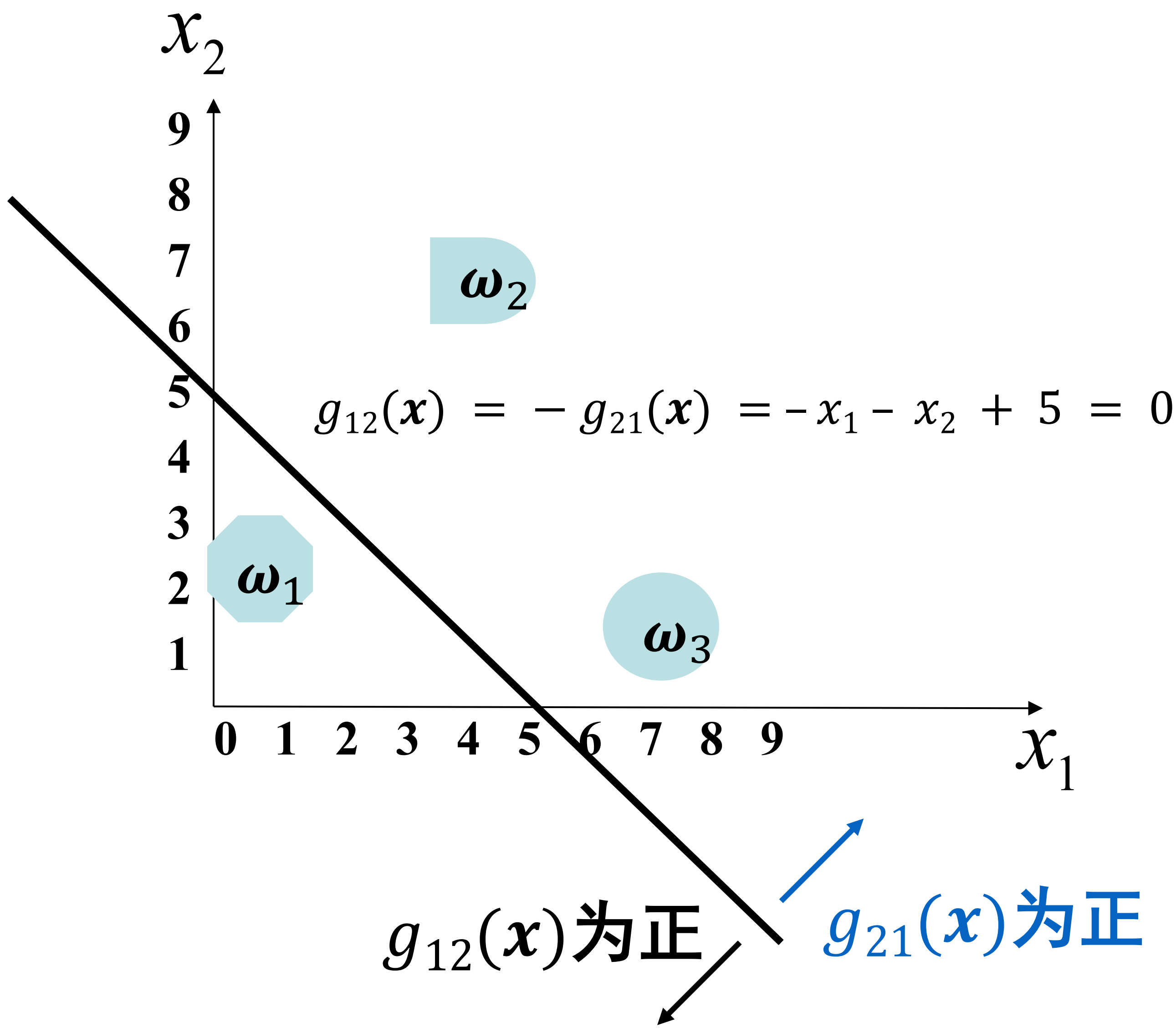
- 判别函数性质： $g_{ij}(\boldsymbol{x}) = -g_{ji}(\boldsymbol{x})$

- 假设判别函数为：

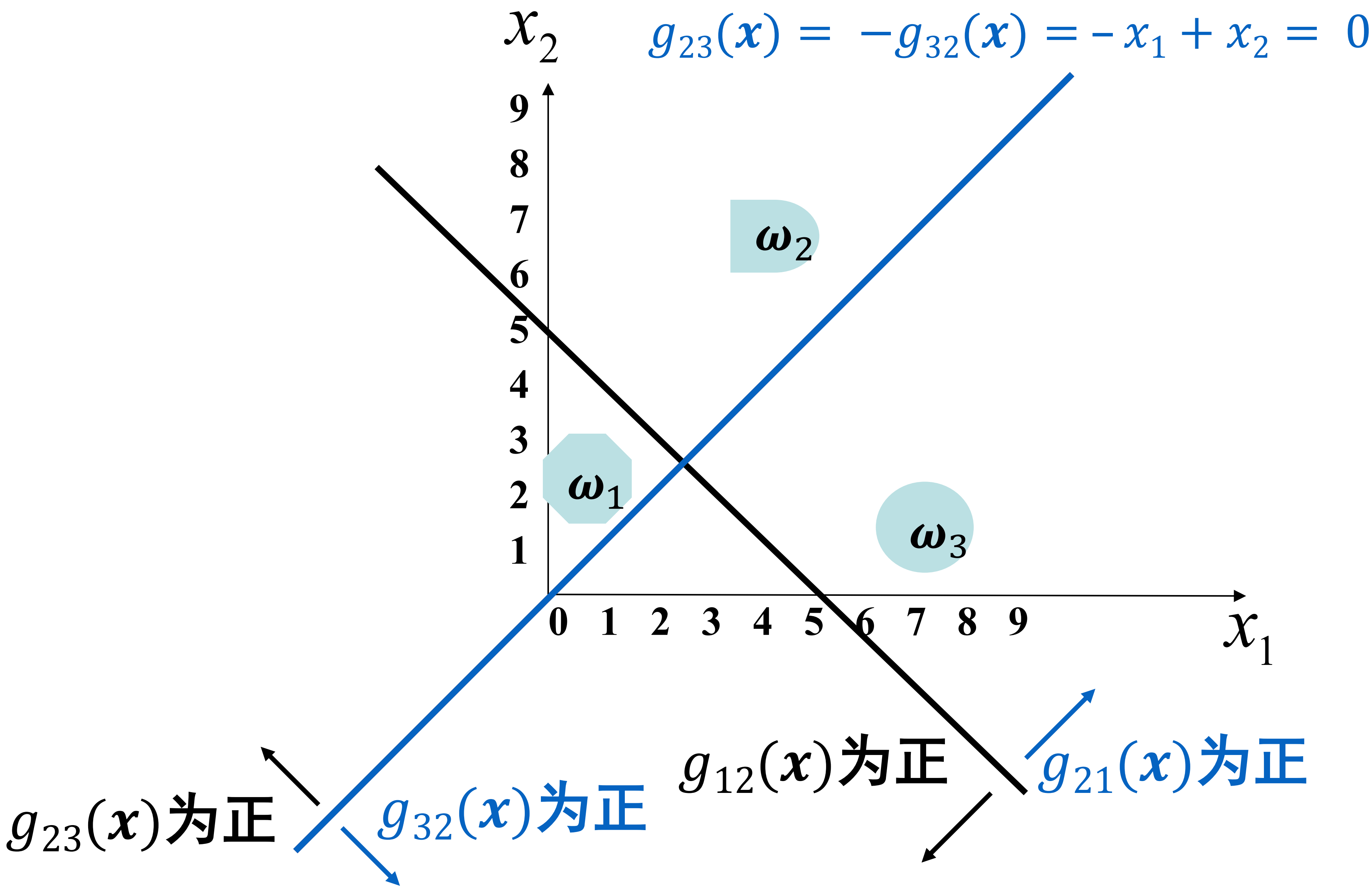
$$\begin{cases} g_{12}(\boldsymbol{x}) = -x_1 - x_2 + 5 \\ g_{13}(\boldsymbol{x}) = -x_1 + 3 \\ g_{23}(\boldsymbol{x}) = -x_1 + x_2 \end{cases}$$

- 判别边界为：

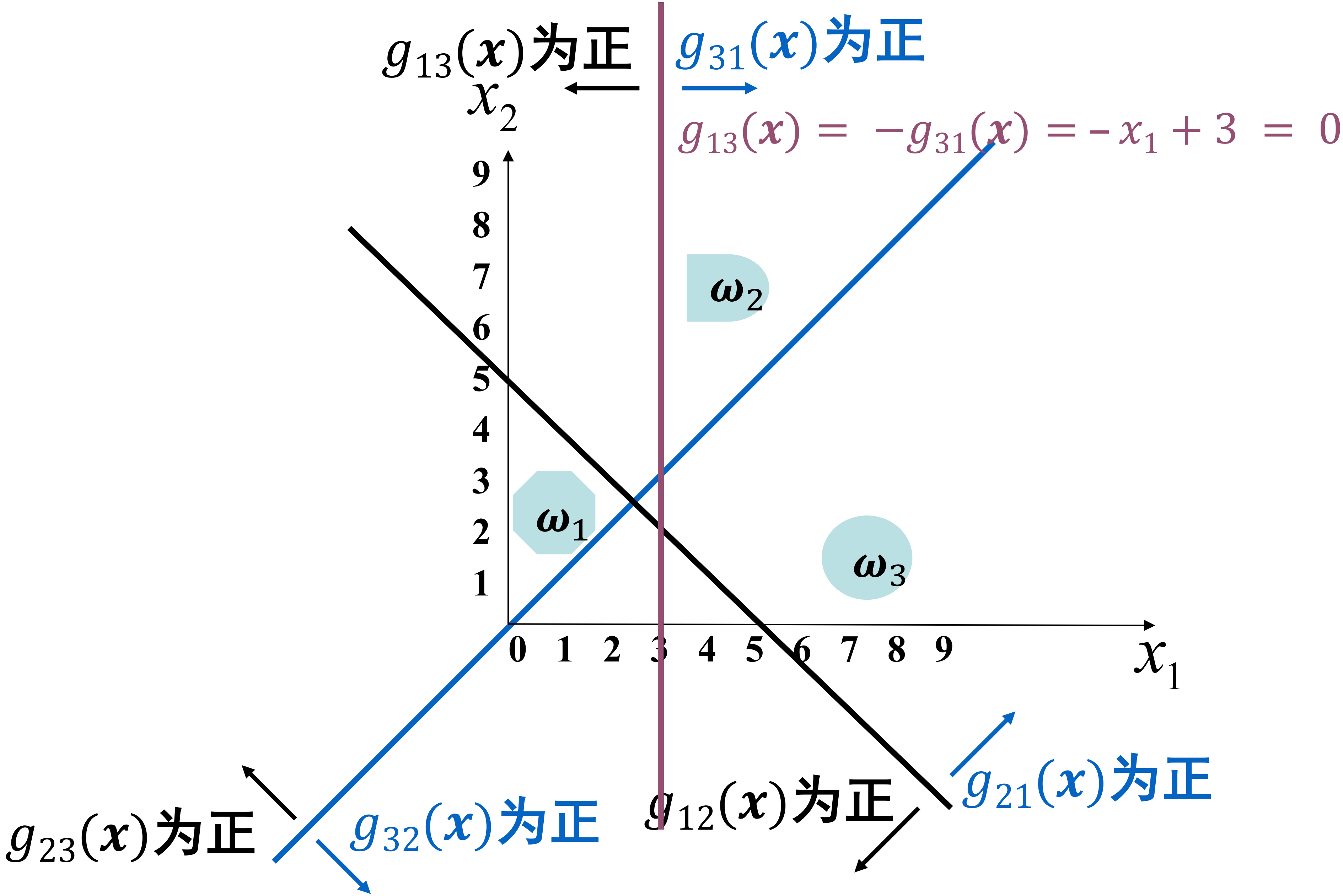
$$\begin{cases} g_{12}(\boldsymbol{x}) = -x_1 - x_2 + 5 = 0 \\ g_{13}(\boldsymbol{x}) = -x_1 + 3 = 0 \\ g_{23}(\boldsymbol{x}) = -x_1 + x_2 = 0 \end{cases}$$



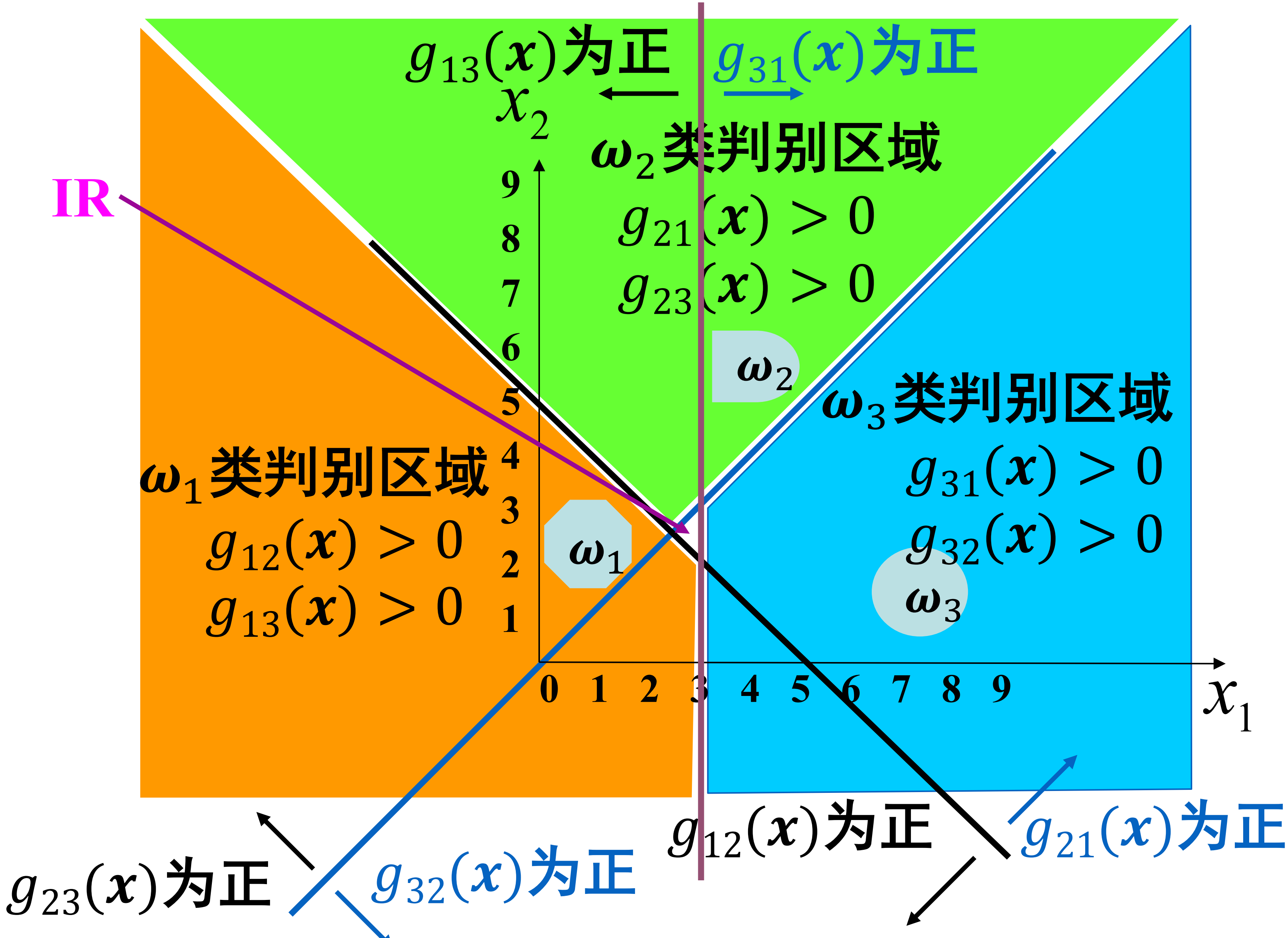
ω_i/ω_j两分法例题图示



ω_i/ω_j两分法例题图示



$\boldsymbol{\omega}_i/\boldsymbol{\omega}_j$ 两分法例题图示



ω_i/ω_j 两分法例题图示

线性判别函数

- 结论：判别区间增大，不确定区间减小，比第一种情况小的多.

❖问:未知模式$\boldsymbol{x}=(x_1,x_2)^T=(4,3)^T$属于哪一类?

$$\begin{cases} g_{12}(\boldsymbol{x})=-x_1-x_2+5=0 \\ g_{13}(\boldsymbol{x})=-x_1+3=0 \\ g_{23}(\boldsymbol{x})=-x_1+x_2=0 \end{cases}$$

❖代入判别函数可得:

$$g_{12}(\boldsymbol{x})=-2, g_{13}(\boldsymbol{x})=-1, g_{23}(\boldsymbol{x})=-1$$

❖把下标对换可得:

$$g_{21}(\boldsymbol{x})=2, g_{31}(\boldsymbol{x})=1, g_{32}(\boldsymbol{x})=1$$

❖因为 $g_{3j}(\boldsymbol{x})>0$

❖结论：所以$\boldsymbol{x}$属于$\boldsymbol{C}_3$类

线性判别函数

◆第三种情况举例

- 每类都有一个判别函数,存在M个判别函数
- 判别函数：$g_K(\boldsymbol{x}) = \boldsymbol{\omega}_k^T\boldsymbol{x} + b_K \qquad K = 1,2,\ldots,M$
- 判别规则：$g_i(\boldsymbol{x}) = \boldsymbol{\omega}_k^T\boldsymbol{x} + b_K \begin{cases} \boldsymbol{x} \in \boldsymbol{\omega}_i & g_i(\boldsymbol{x}) > g_j(\boldsymbol{x}), \forall j \,\&j \neq i \\ uncertain & otherwise \end{cases}$
- 判别边界：$g_i(\boldsymbol{x}) = g_j(\boldsymbol{x})$或$g_i(\boldsymbol{x}) - g_j(\boldsymbol{x}) = 0$
- 就是说，要判别模式$\boldsymbol{x}$属于哪一类，先把$\boldsymbol{x}$代入M个判别函数中，判别函数最大的那个类别就是$\boldsymbol{x}$所属类别。类与类之间的边界可由 $g_i(\boldsymbol{x}) = g_j(\boldsymbol{x})$或$g_i(\boldsymbol{x}) - g_j(\boldsymbol{x}) = 0$来确定。
- **无不确定区间**

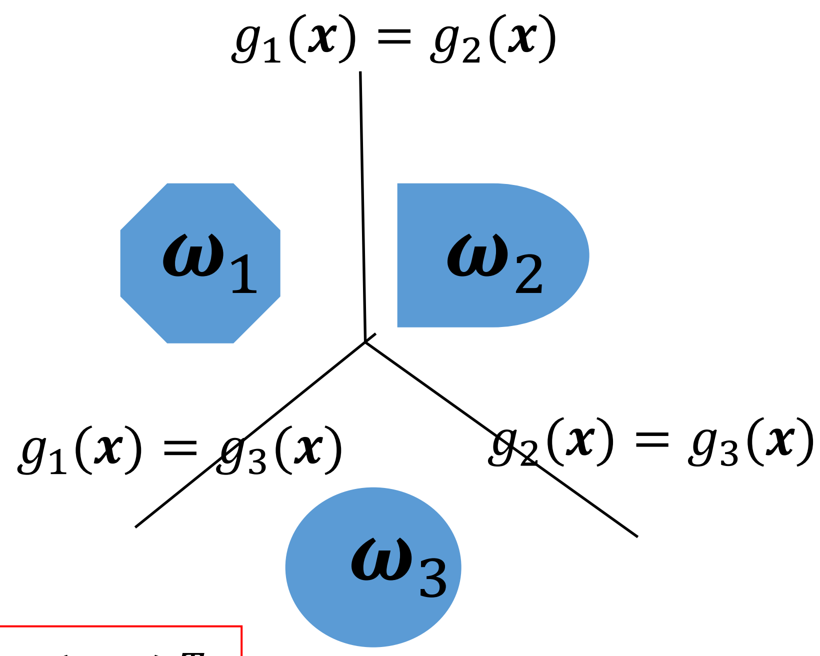
线性判别函数

- 右图所示是K=3 的例子。对于$\boldsymbol{C}_1$类模式，
- 必然满足$g_1(\boldsymbol{x}) > g_2(\boldsymbol{x})$ 和 $g_1(\boldsymbol{x}) > g_3(\boldsymbol{x})$ 。
- 假设判别函数为：

$$\begin{cases} g_1(\boldsymbol{x}) = -x_1 + x_2 \\ g_2(\boldsymbol{x}) = x_1 + x_2 - 1 \\ g_3(\boldsymbol{x}) = -x_2 \end{cases}$$

- 则判别边界为：

$$\begin{cases} g_1(\boldsymbol{x}) - g_2(\boldsymbol{x}) = -2x_1 + 1 = 0 \\ g_1(\boldsymbol{x}) - g_3(\boldsymbol{x}) = -x_1 + 2x_2 = 0 \\ g_2(\boldsymbol{x}) - g_3(\boldsymbol{x}) = x_1 + 2x_2 - 1 = 0 \end{cases}$$



问：假设未知模式$\boldsymbol{x} = (x_1, x_2)^T = (1,1)^T$，
则$\boldsymbol{x}$属于那一类？

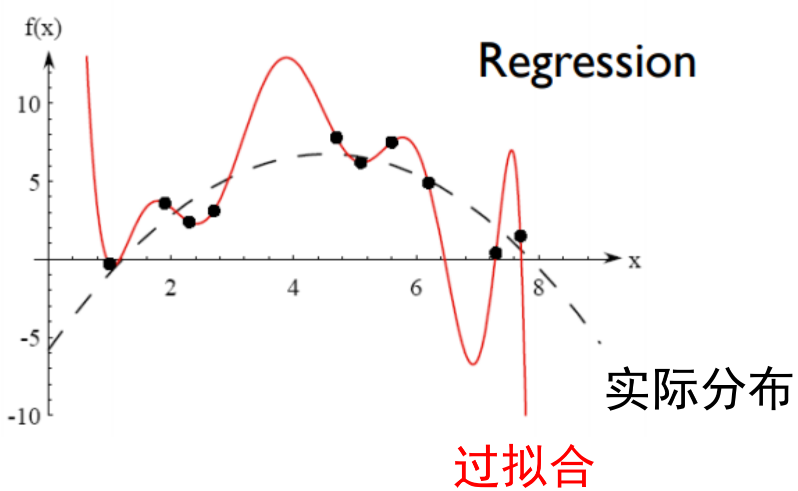
主要内容

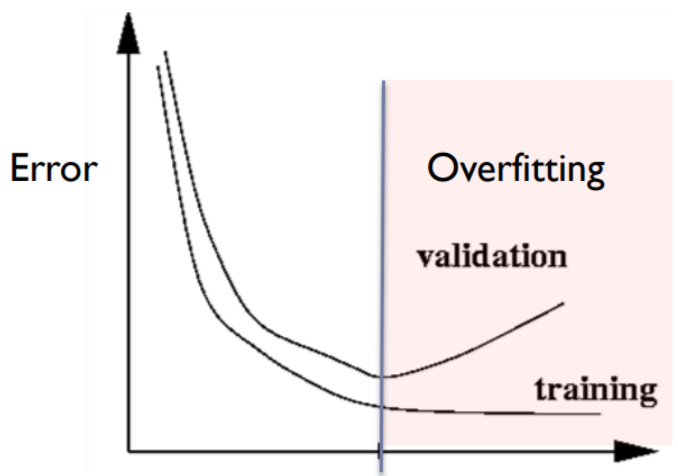
- 线性判别分类(Linear Discriminants for Classification)
- 分类举例
- 线性判别函数
  - 两类问题
  - 多类问题
- Fisher线性判别
- 支持矢量机(Support Vector Machine)
- 自适应增强AdaBoost(Adaptive Boosting)

Fisher线性判别

挑战： 高维空间上的分类界面计算复杂，样本稀疏，过拟合风险高

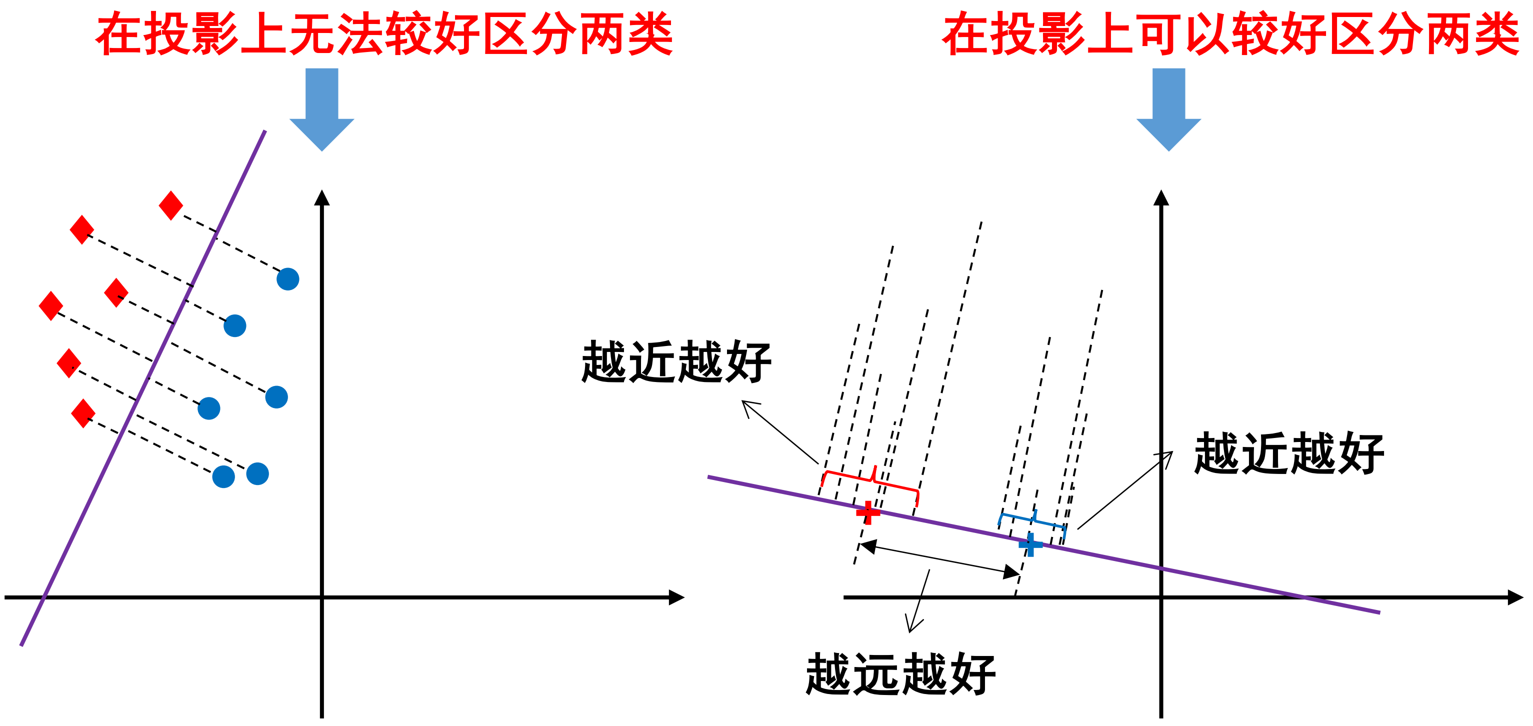
思路： 高维空间**降维**到低维空间实现简单高效判别





Fisher线性判别

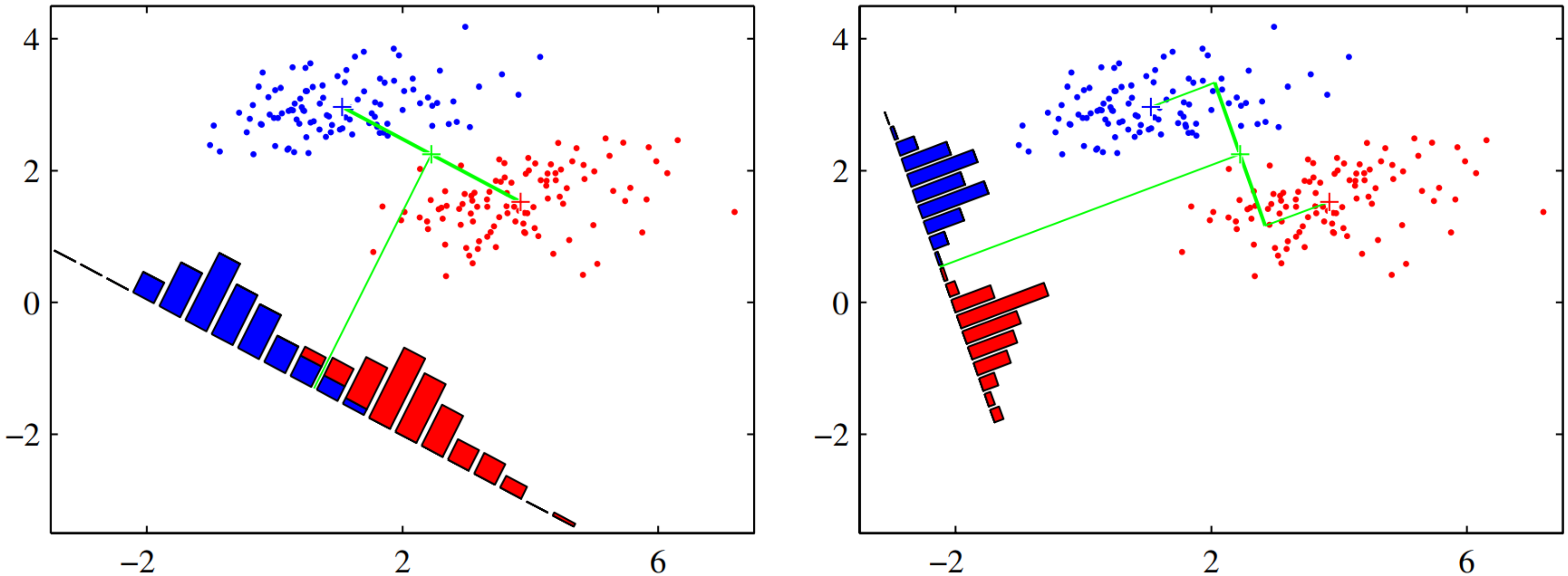
方法：求取线性投影$y = \boldsymbol{w}^T\boldsymbol{x}$，将高维表示$x$降维到低维表示$y$，保证$y$**利于分类**。



二维模式向一维空间投影示意图

Fisher线性判别

- Fisher线性判别



Fisher判别法是1936年提出来的，该方法的主要思想是通过将多维数据投影到某个方向上，投影的原则是将总体与总体之间尽可能的分开，然后再选择合适的判别规则，将新的样本进行分类判别。

(1)求解Fisher准则函数

设给定n维训练样本$x_1 \dots x_N$，其中有N_1和$N_2 = N - N_1$个样本分属C_1类和C_2类。

待求w

$$y = \mathbf{w}^{\mathrm{T}}\mathbf{x}.$$

$$J(\mathbf{w}) = \frac{(m_2 - m_1)^2}{s_{W_1} + s_{W_2}}$$

s_{W_k}记为类内离差度。

Fisher线性判别

类别中心：

$$\boldsymbol{m}_1 = \frac{1}{\boldsymbol{N}_1}\sum_{\boldsymbol{x}_n \in \boldsymbol{C}_1} \boldsymbol{x}_n$$

$$\boldsymbol{m}_2 = \frac{1}{\boldsymbol{N}_2}\sum_{\boldsymbol{x}_n \in \boldsymbol{C}_2} \boldsymbol{x}_n$$

变换后各个模式在投影方向$\boldsymbol{w}$上的均值可以表示为：

$$m_1 = \frac{1}{\boldsymbol{N}_1}\sum_{\boldsymbol{x}_n \in \boldsymbol{C}_1} \boldsymbol{w}^T \boldsymbol{x}_n = \boldsymbol{w}^T \boldsymbol{m}_1$$

$$m_2 = \frac{1}{\boldsymbol{N}_1}\sum_{\boldsymbol{x}_n \in \boldsymbol{C}_2} \boldsymbol{w}^T \boldsymbol{x}_n = \boldsymbol{w}^T \boldsymbol{m}_2$$

类内散度：

$$s_W = \sum_{\boldsymbol{x}_n \in \boldsymbol{C}_1} (\boldsymbol{w}^T \boldsymbol{x}_n - \boldsymbol{w}^T \boldsymbol{m}_1)^2 + \sum_{\boldsymbol{x}_n \in \boldsymbol{C}_2} (\boldsymbol{w}^T \boldsymbol{x}_n - \boldsymbol{w}^T \boldsymbol{m}_2)^2 = \boldsymbol{w}^T \boldsymbol{S}_W \boldsymbol{w}$$

s_{W_1} s_{W_2}

类内散度矩阵：

$$\boldsymbol{S}_W = \sum_{\boldsymbol{x}_n \in \boldsymbol{C}_1} (\boldsymbol{x}_n - \boldsymbol{m}_1)(\boldsymbol{x}_n - \boldsymbol{m}_1)^T + \sum_{\boldsymbol{x}_n \in \boldsymbol{C}_2} (\boldsymbol{x}_n - \boldsymbol{m}_2)(\boldsymbol{x}_n - \boldsymbol{m}_2)^T$$

Fisher线性判别

类间散度：

$$s_B = (\boldsymbol{m}_2 - \boldsymbol{m}_1)^2$$
$$= (\boldsymbol{w}^T\boldsymbol{m}_2 - \boldsymbol{w}^T\boldsymbol{m}_1)^2 = \boldsymbol{w}^T\boldsymbol{S_B}\boldsymbol{w}$$

类间散度矩阵：

$$\boldsymbol{S_B} = (\boldsymbol{m}_2 - \boldsymbol{m}_1)(\boldsymbol{m}_2 - \boldsymbol{m}_1)^T$$

根据Fisher 准则的定义，期望类内散度s_W越小，类间散度s_B越大越好，根据以上准则定义目标函数：

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T\boldsymbol{S_B}\boldsymbol{w}}{\boldsymbol{w}^T\boldsymbol{S_W}\boldsymbol{w}}$$

Fisher线性判别

求解$J(\boldsymbol{w})=\frac{\boldsymbol{w}^T\boldsymbol{S}_B\boldsymbol{w}}{\boldsymbol{w}^T\boldsymbol{S}_W\boldsymbol{w}}$转化为优化问题：

$$max_{\boldsymbol{w}}\ \ \boldsymbol{w}^T\boldsymbol{S}_B\boldsymbol{w}$$

$$s.t.\ \boldsymbol{w}^T\boldsymbol{S}_W\boldsymbol{w}=c\neq 0$$

s.t.表示优化问题中需要满足约束的条件 （subject to的缩写）

拉格朗日乘子法：

假设需要求极值的目标函数（objective function）为$f(x,y)$，限制条件为 $\varphi(x,y)=M$

设$g(x,y)=M-\varphi(x,y)$

定义一个新函数

$$F(x,y,\lambda)=f(x,y)+\lambda g(x,y)$$

则用偏导数方法列出方程：

$$\partial F/\partial x=0$$
$$\partial F/\partial y=0$$
$$\partial F/\partial \lambda=0$$

求出x,y,λ的值，代入即可得到目标函数的极值

Fisher线性判别

扩展为多个变量的式子为：

$$F(x_1, x_2, \ldots \lambda) = f(x_1, x_2, \ldots) + \lambda g(x_1, x_2 \ldots)$$

则求极值点的方程为：

$\partial F/\partial x_i = 0$（$x_i$即为$x_1, x_2$……等自变量）

$\partial F/\partial \lambda = g(x_1, x_2 \ldots) = 0$

通过引入拉格朗日乘子转化为以下拉格朗日无约束极值问题：

$$L(\boldsymbol{w},\lambda) = \boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w} - \lambda(\boldsymbol{w}^T \boldsymbol{S}_W \boldsymbol{w} - c)$$

令导数

$$\frac{\partial L(\boldsymbol{w},\lambda)}{\partial \boldsymbol{w}} = 0$$

可得：

$$\boldsymbol{S}_B \boldsymbol{w} - \lambda \boldsymbol{S}_W \boldsymbol{w} = \boldsymbol{0}$$

可得：

$$\boldsymbol{S}_B\boldsymbol{w} = \lambda\boldsymbol{S}_W\boldsymbol{w}$$

假定$\boldsymbol{S}_W$为非奇异，则：

$$\boldsymbol{S}_W{}^{-1}\boldsymbol{S}_B\boldsymbol{w} = \lambda\boldsymbol{w}$$

即最佳投影方向为$\boldsymbol{S}_W{}^{-1}\boldsymbol{S}_B$的特征向量。

（3）求解Fisher判别函数

- 由于变换后的模式是一维的，因此判别界面实际上是各类模式所在轴上的一个点，所以可以根据训练模式确定一个阈值w_0，于是**Fisher判别规则**为：

$$\boldsymbol{w}^T\boldsymbol{x} \gtrless w_0 \quad \boldsymbol{x} \in \begin{cases} \boldsymbol{C}_1 \\ \boldsymbol{C}_2 \end{cases}$$

- 判别阈值可取两个类心在w方向上轴的投影连线的中点作为阈值，即：

$$w_0 = \frac{m_1 + m_2}{2}$$

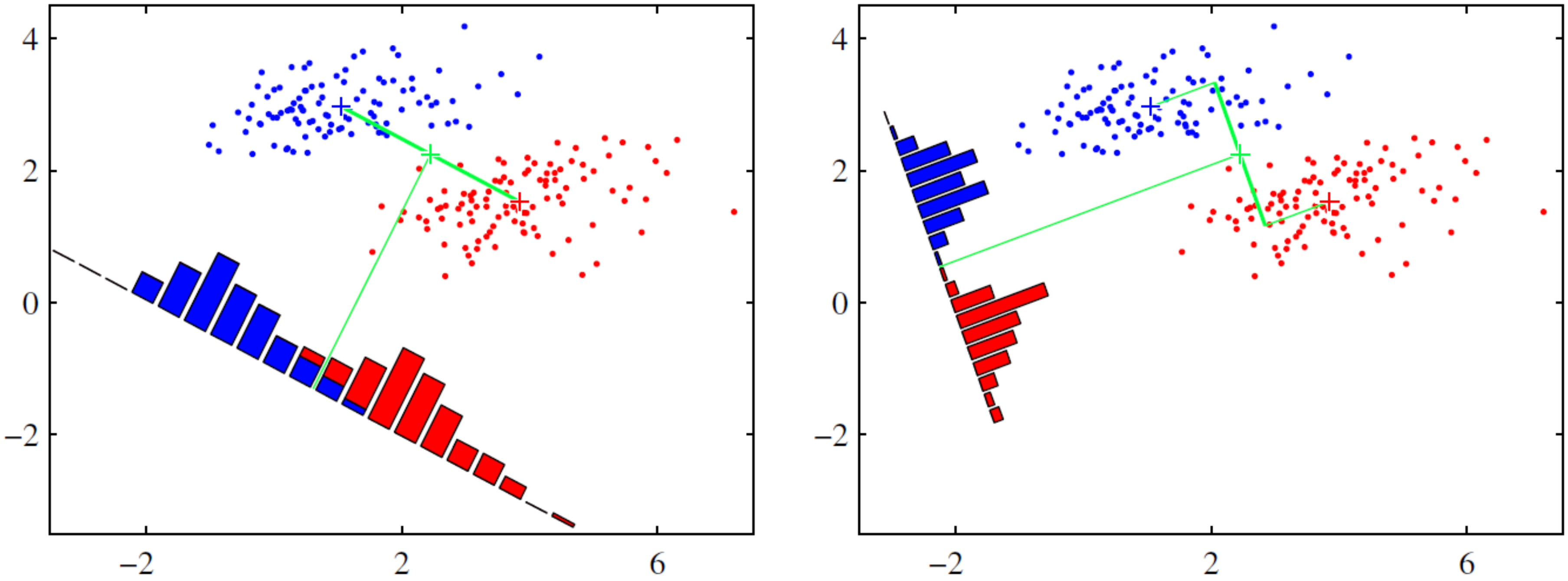
Fisher线性判别

判别规则为：

$$\boldsymbol{w}^T\boldsymbol{x}_n \underset{<}{\overset{>}{}} w_0，\quad \boldsymbol{x}_n \in \begin{cases} \boldsymbol{C}_1 \\ \boldsymbol{C}_2 \end{cases}$$

等价于：

$$g(\boldsymbol{x}_n) = \boldsymbol{w}^T\boldsymbol{x}_n - w_0 \underset{<}{\overset{>}{}} 0，\quad \boldsymbol{x}_n \in \begin{cases} \boldsymbol{C}_1 \\ \boldsymbol{C}_2 \end{cases}$$



Fisher线性判别

1. 计算类内散度矩阵：

$$\boldsymbol{S_W} = \sum_{\boldsymbol{x}_n \in \boldsymbol{C}_1} (\boldsymbol{x}_n - \boldsymbol{m}_1)(\boldsymbol{x}_n - \boldsymbol{m}_1)^T + \sum_{\boldsymbol{x}_i \in \boldsymbol{C}_2} (\boldsymbol{x}_i - \boldsymbol{m}_2)(\boldsymbol{x}_i - \boldsymbol{m}_2)^T$$

2. 计算类间散度矩阵：

$$\boldsymbol{S_B} = (\boldsymbol{m}_2 - \boldsymbol{m}_1)(\boldsymbol{m}_2 - \boldsymbol{m}_1)^T$$

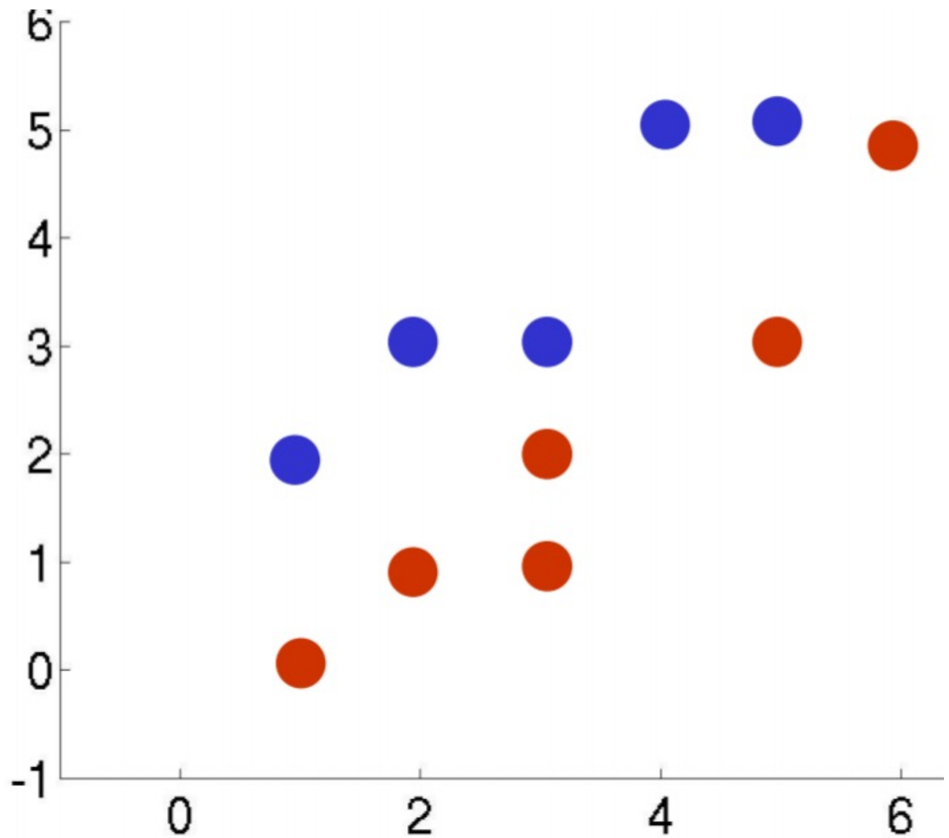
3. 求解特征值：

$$\boldsymbol{S_W}^{-1}\boldsymbol{S_B}\boldsymbol{w} = \lambda\boldsymbol{w}$$

Fisher线性判别示例：

给定如下两类样本，

$$\boldsymbol{X}_1 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 3 \\ 4 & 5 \\ 5 & 5 \end{bmatrix}^T \quad \boldsymbol{X}_2 = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 1 \\ 3 & 2 \\ 5 & 3 \\ 6 & 5 \end{bmatrix}^T$$



Fisher线性判别

(1) 首先计算两类样本均值：

$$\boldsymbol{m_1} = [3 \quad 3.6]^T \qquad \boldsymbol{m_2} = [3.3 \quad 2]^T$$

(2) 计算两类各自的类内离差阵：

$$\boldsymbol{S_{W_1}} = \sum_n (\boldsymbol{x_n} - \boldsymbol{m_1})(\boldsymbol{x_n} - \boldsymbol{m_1})^T = \begin{bmatrix} 10 & 8 \\ 8 & 7.2 \end{bmatrix}$$

$$\boldsymbol{S_{W_2}} = \sum_n (\boldsymbol{x_n} - \boldsymbol{m_2})(\boldsymbol{x_n} - \boldsymbol{m_2})^T = \begin{bmatrix} 17.3 & 16 \\ 16 & 16 \end{bmatrix}$$

(3) 计算类内总离差阵：

$$\boldsymbol{S_W} = \boldsymbol{S_{W_1}} + \boldsymbol{S_{W_2}} = \begin{bmatrix} 27.3 & 24 \\ 24 & 23.2 \end{bmatrix}$$

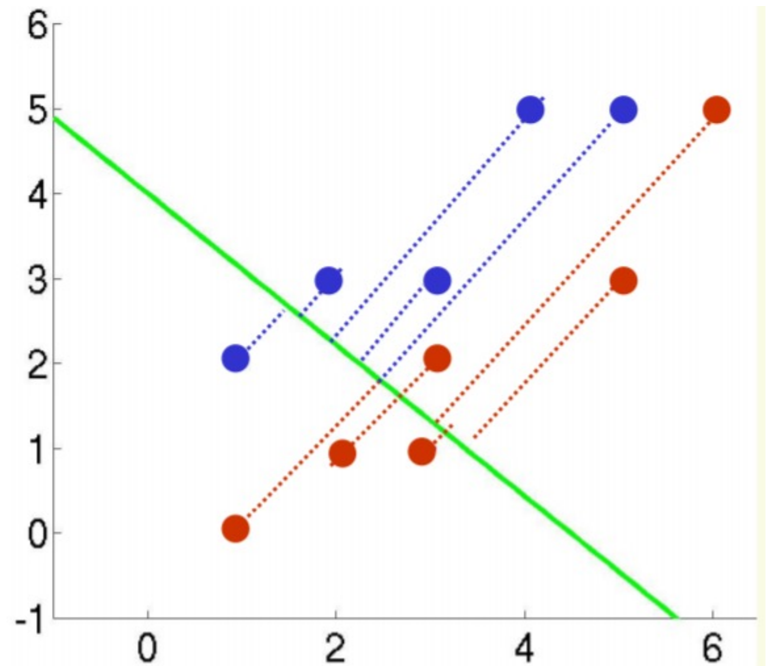
(4) 计算总离差阵的逆矩阵和投影向量：

$$S_W^{-1}=\begin{bmatrix}0.40 & -0.42\\ -0.42 & 0.48\end{bmatrix}, S_B=\begin{bmatrix}0.09 & -0.48\\ -0.48 & 2.56\end{bmatrix}, w=\begin{bmatrix}-0.79\\ 0.89\end{bmatrix}$$

(5) 计算每类样本的投影：

$$Y_1=w\cdot X_1=\begin{bmatrix}0.99\\ 1.09\\ 0.30\\ 1.29\\ 0.50\end{bmatrix}^T$$

$$Y_2=w\cdot X_2=\begin{bmatrix}-0.79\\ -0.69\\ -1.48\\ -0.59\\ -1.28\\ -0.29\end{bmatrix}^T$$



Fisher线性判别

(6) 计算两类判别门限：

$$w_0 = \frac{\boldsymbol{w}^T\boldsymbol{m}_1 + \boldsymbol{w}^T\boldsymbol{m}_2}{2} = 0.004$$

$$\boldsymbol{Y}_1 = \boldsymbol{w}^T\boldsymbol{X}_1 = \begin{bmatrix} 0.99 \\ 1.09 \\ 0.30 \\ 1.29 \\ 0.5 \end{bmatrix}^T > w_0 \qquad \boldsymbol{Y}_2 = \boldsymbol{w}^T\boldsymbol{X}_1 = \begin{bmatrix} -0.79 \\ -0.69 \\ -1.48 \\ -0.59 \\ -1.28 \\ -0.29 \end{bmatrix}^T < w_0$$

两类样本全部正确划分

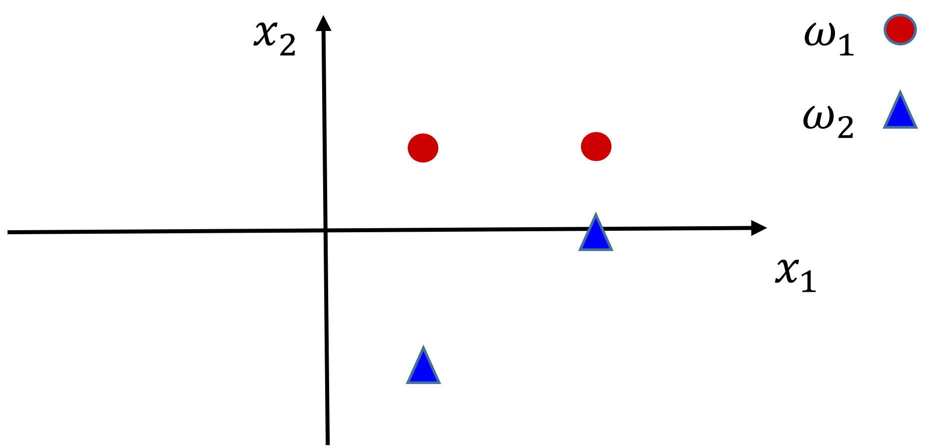
课堂练习：

1. 有两类样本：

$$C_1: \{(1,1)^T, (3,1)^T\}$$

$$C_2: \{(1,-2)^T, (3,0)^T\}$$

用Fisher准则进行分类，并分析其分类性能。



答案：(1) 首先计算两类样本均值：

$$\boldsymbol{m}_1 = [2 \quad 1]^T \qquad \boldsymbol{m}_2 = [2 \quad -1]^T$$

(2) 计算两类各自的类内离差阵：

$$\boldsymbol{S}_{W_1} = \sum_n (\boldsymbol{x}_n - \boldsymbol{m}_1)(\boldsymbol{x}_n - \boldsymbol{m}_1)^T = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\boldsymbol{S}_{W_2} = \sum_n (\boldsymbol{x}_n - \boldsymbol{m}_2)(\boldsymbol{x}_n - \boldsymbol{m}_2)^T = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

(3) 计算类内总离差阵：

$$\boldsymbol{S}_W = \boldsymbol{S}_{W_1} + \boldsymbol{S}_{W_2} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$$

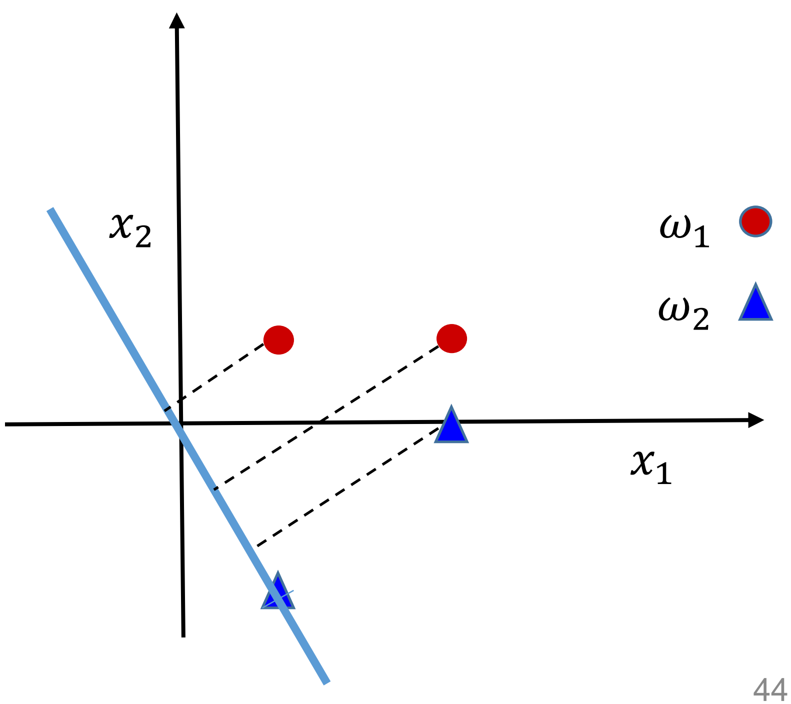
(4) 计算总离差阵的逆矩阵和投影向量：

$$\boldsymbol{S_W}^{-1} = \frac{1}{4}\begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}, \boldsymbol{S_B} = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}, \boldsymbol{w} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

(5) 计算每类样本的投影：

$$\boldsymbol{Y}_1 = \boldsymbol{w} \cdot \boldsymbol{X}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}^T$$

$$\boldsymbol{Y}_2 = \boldsymbol{w} \cdot \boldsymbol{X}_2 = \begin{bmatrix} -5 \\ -3 \end{bmatrix}^T$$



(6) 计算两类判别门限：

$$w_0 = \frac{\boldsymbol{w}^T\boldsymbol{m}_1 + \boldsymbol{w}^T\boldsymbol{m}_2}{2} = -2$$

$$\boldsymbol{Y}_1 = \boldsymbol{w}^T\boldsymbol{\omega}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}^T > w_0 \qquad \boldsymbol{Y}_2 = \boldsymbol{w}^T\boldsymbol{\omega}_2 = \begin{bmatrix} -5 \\ -3 \end{bmatrix}^T < w_0$$

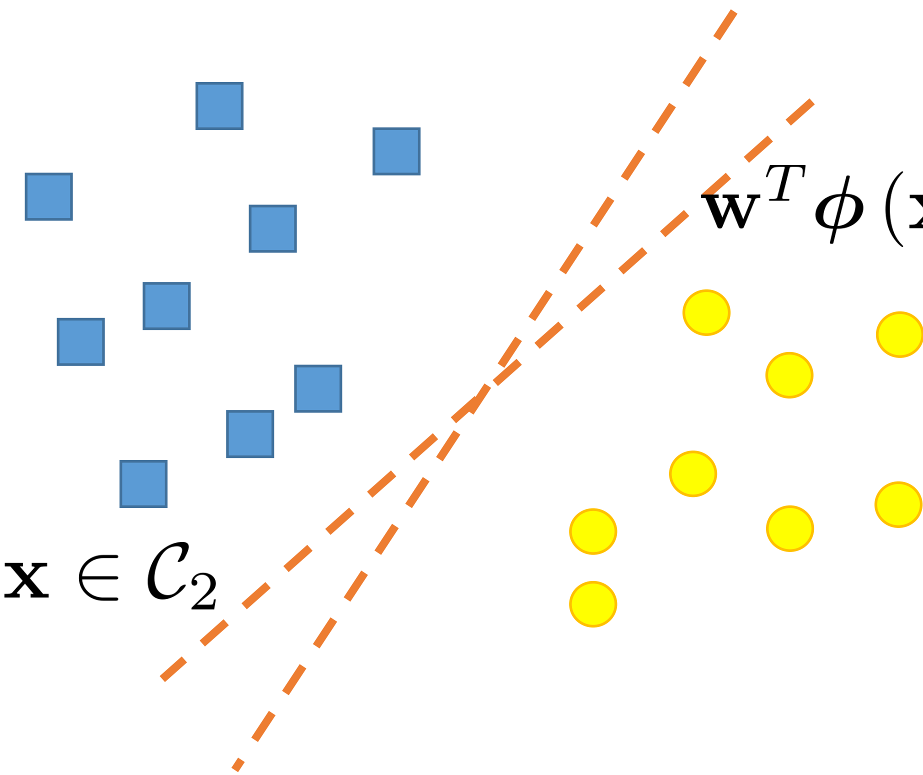
两类样本全部正确划分

主要内容

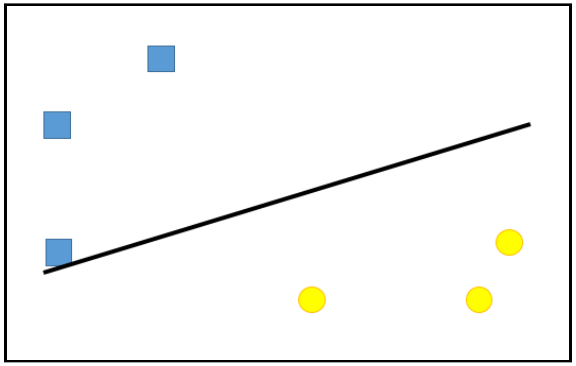
- 线性判别分类(Linear Discriminants for Classification)
- 分类举例
- 线性判别函数
  - 两类问题
  - 多类问题
- Fisher线性判别
- 支持矢量机(Support Vector Machine)
- 自适应增强AdaBoost(Adaptive Boosting)

线性分类器

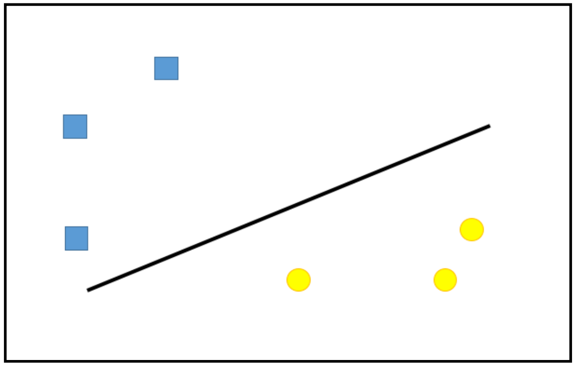
- 哪个超平面更好？



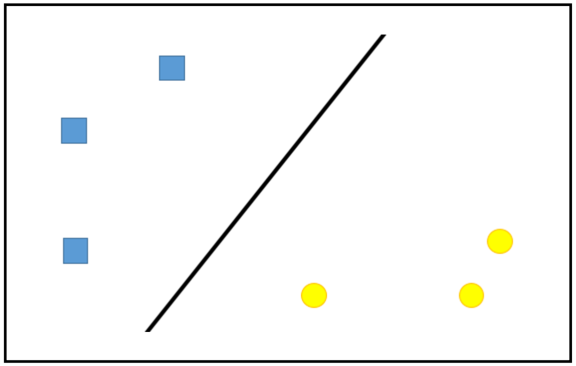
- 如下分类界面的优劣？



A



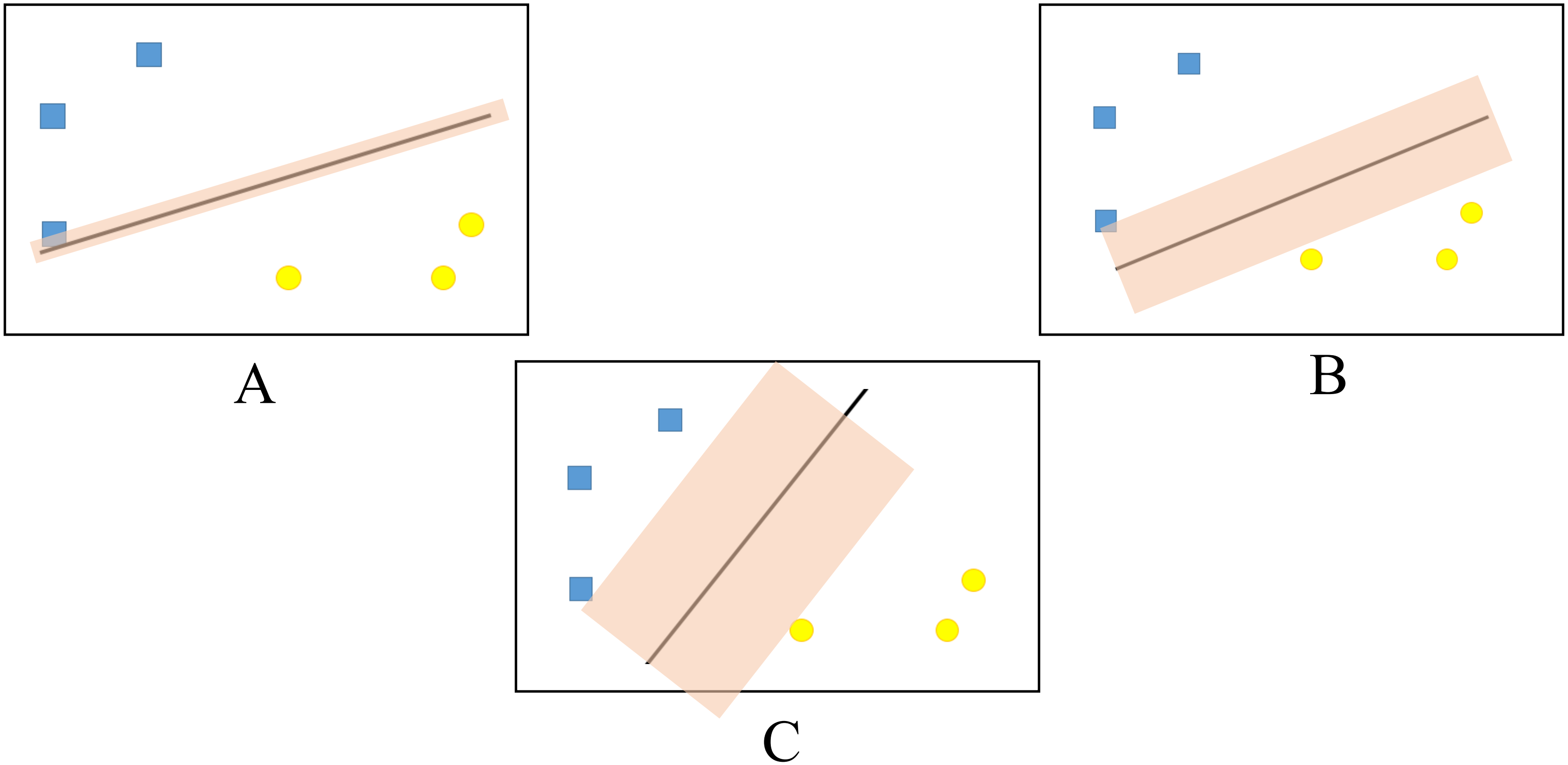
B



C

支持向量机

- 如何提高分类界面对噪声的鲁棒性？

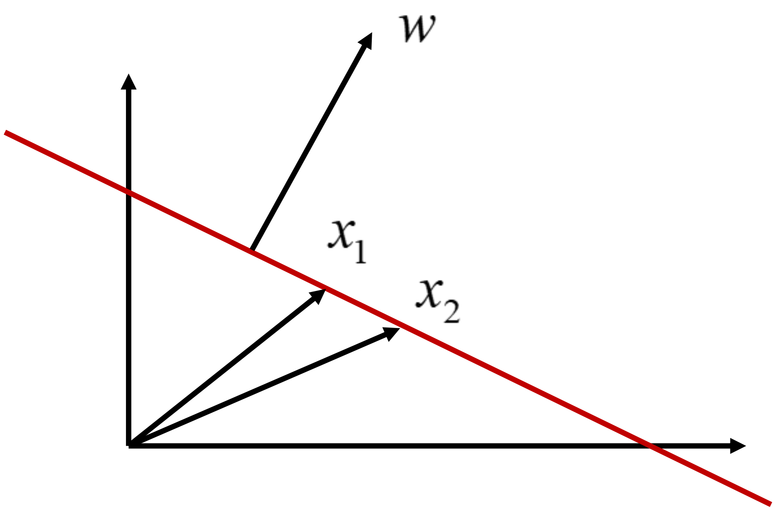


增大分类界面的缓冲空间 (margin)，是提升其鲁棒性的关键

支持向量机

- 如何度量margin?

➢从回顾线性代数的几何性质开始



1. 超平面的权重向量，与平面上所有向量正交：

证明：设$\boldsymbol{x}_1, \boldsymbol{x}_2$是平面$\boldsymbol{w}^T\boldsymbol{x}+b$上的点，且满足

$$\boldsymbol{w}^T\boldsymbol{x}_1 + b = \boldsymbol{w}^T\boldsymbol{x}_2 + b = 0$$

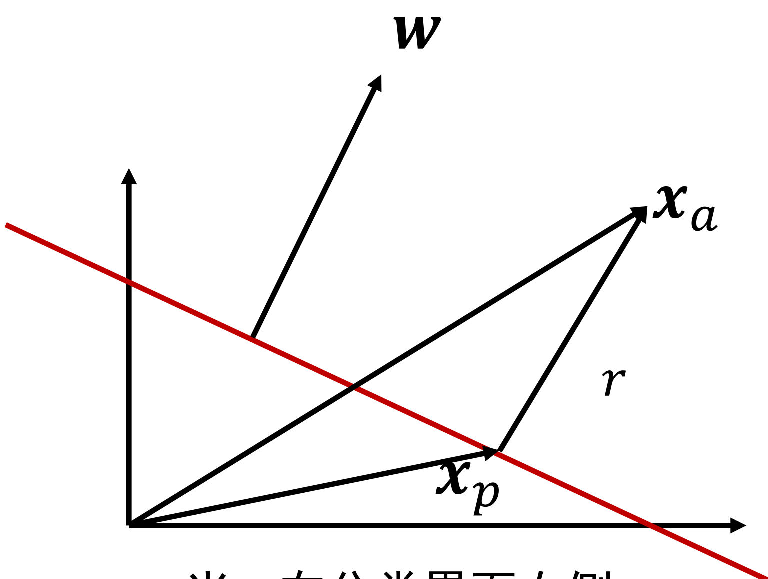
$$\boldsymbol{w}^T(\boldsymbol{x}_1 - x_2) + b = 0$$

即，向量$\boldsymbol{w}$垂直于该超平面

支持向量机

- 如何度量margin?

➢从回顾线性代数的几何性质开始



当$\boldsymbol{x}_a$在分类界面右侧

2. 超平面外任意一点$\boldsymbol{x}_a$, 到该平面的距离为:

$$\frac{|\boldsymbol{w}^T\boldsymbol{x}_a + b|}{\|\boldsymbol{w}\|}$$

证明：设r表示$\boldsymbol{x}$到平面$\boldsymbol{w}^T\boldsymbol{x} + b$的垂直距离，其中$\boldsymbol{x}_p$是$\boldsymbol{x}_a$到平面$\boldsymbol{w}^T\boldsymbol{x} + b$的正交投影向量，则

$$\boldsymbol{x}_a = \boldsymbol{x}_p + r \cdot \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|} \quad 且 \quad \boldsymbol{w}^T\boldsymbol{x}_p + b = 0$$

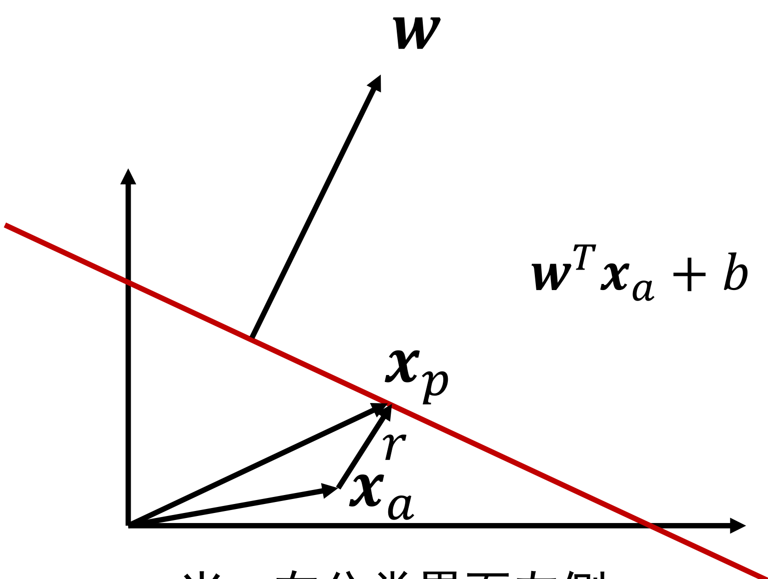
$$\boldsymbol{w}^T\boldsymbol{x}_a + b = \boldsymbol{w}^T\left(\boldsymbol{x}_p + r \cdot \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\right) + b = r \cdot \frac{\boldsymbol{w}^T\boldsymbol{w}}{\|\boldsymbol{w}\|} = r \cdot \|\boldsymbol{w}\|$$

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

支持向量机

- 如何度量margin?

➢从回顾线性代数的几何性质开始



当$\boldsymbol{x}_a$在分类界面左侧

$$\boldsymbol{x}_a = \boldsymbol{x}_p - r \cdot \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|} \quad 且 \quad \boldsymbol{w}^T \boldsymbol{x}_p + b = 0$$

$$\boldsymbol{w}^T \boldsymbol{x}_a + b = \boldsymbol{w}^T \left(\boldsymbol{x}_p - r \cdot \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|} \right) + b = -r \cdot \frac{\boldsymbol{w}^T \boldsymbol{w}}{\|\boldsymbol{w}\|} = -r \cdot \|\boldsymbol{w}\|$$

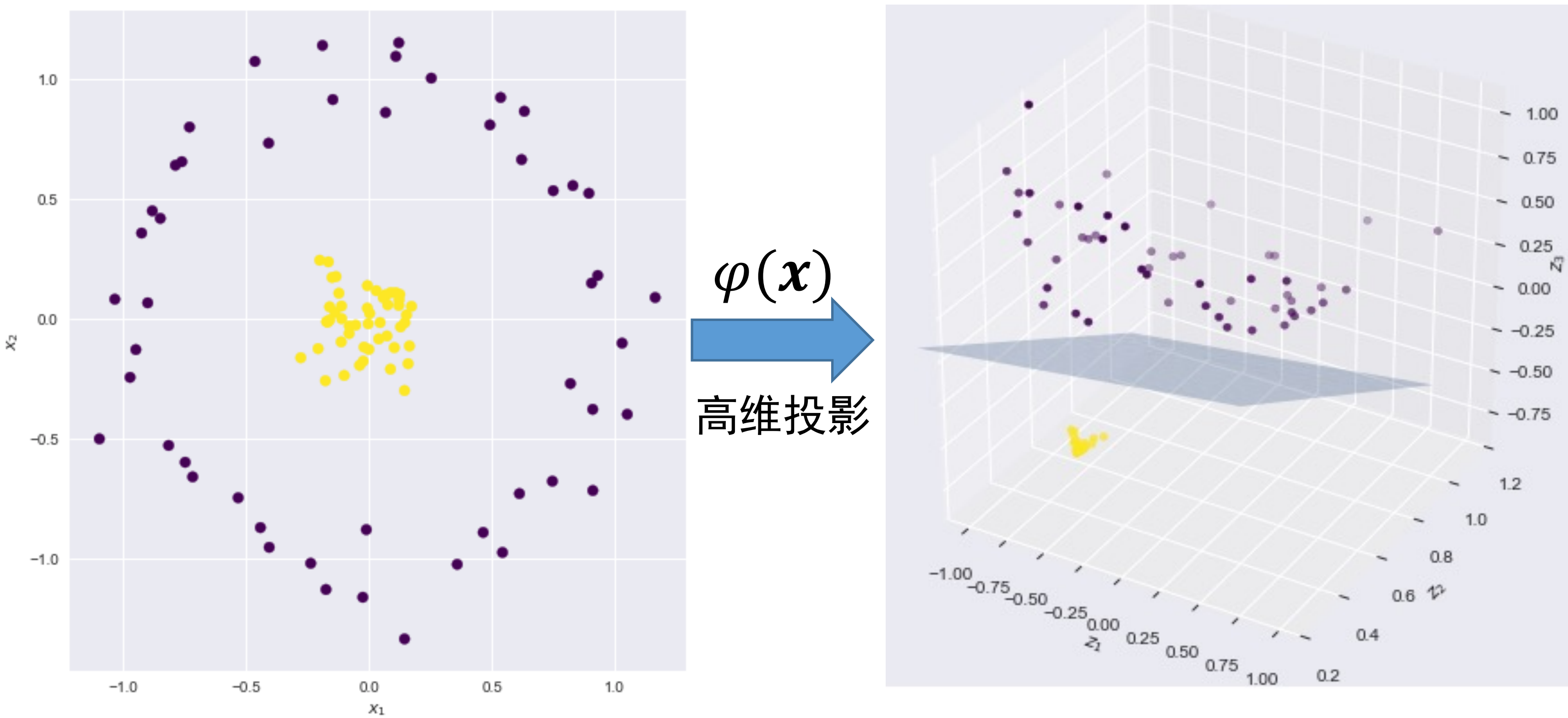
$$r = -\frac{\boldsymbol{w}^T \boldsymbol{x}_a + b}{\|\boldsymbol{w}\|}$$

综上，无论$\boldsymbol{x}_a$在任何位置，其到超平面的垂直距离为 $\dfrac{|\boldsymbol{w}^T \boldsymbol{x}_a + b|}{\|\boldsymbol{w}\|}$

支持向量机

- 特征投影示例

➢假设我们采用如下投影函数 $\boldsymbol{x}=(x_1,x_2)^T, \varphi(\boldsymbol{x})=(\sqrt{2}x_1x_2, {x_1}^2, {x_2}^2)$

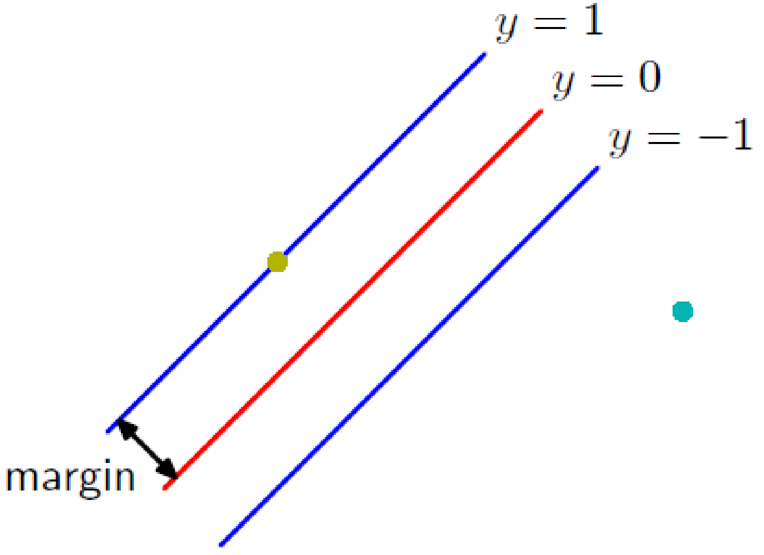


投影后特征线性可分，我们可在投影空间上求取线性判别函数

边界(Margin)

Margin：决策边界和任意一个样本之间的最小距离。点$\boldsymbol{x}_n$到决策面的距离为：

$$\frac{t_n y(\mathbf{x}_n)}{\|\mathbf{w}\|} = \frac{t_n(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)+b)}{\|\mathbf{w}\|}$$



最大化边界

我们希望优化参数w和b，以使Margin最大化。

$$\arg\max_{\mathbf{w},b}\left\{\frac{1}{\|\mathbf{w}\|}\min_{n}\left[t_n\left(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)+b\right)\right]\right\}$$

如果进行重新缩放，使得$\mathbf{w}\rightarrow k\mathbf{w}, b\rightarrow kb$；则对于任意一点$\boldsymbol{x}_n$到决策面的距离$t_n y(\boldsymbol{x}_n)/\mathbf{w}$是保持不变的。

我们可以利用这一点来设置距离该决策面最近的点的取值：

$$t_n\left(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)+b\right)=1$$

支持向量(Support vector)

对于支持向量:

$$t_n\left(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)+b\right)=1$$

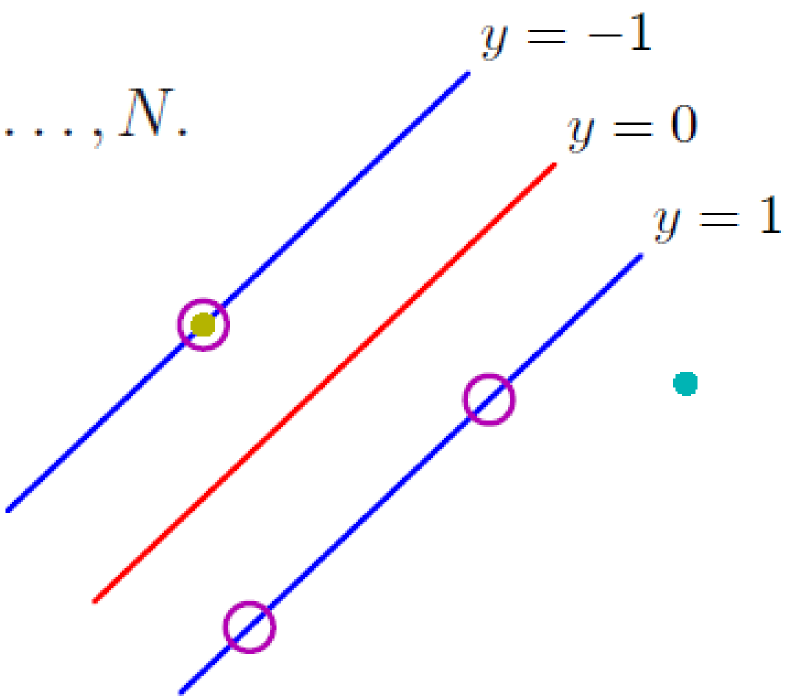
对于所有点:

$$t_n\left(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)+b\right)\geqslant 1,\qquad n=1,\ldots,N.$$

我们希望最小化边界:

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

$$s.t.\quad t_n\left(\mathbf{w}^T\boldsymbol{\phi}\left(\mathbf{x_n}\right)+b\right)\geq 1$$



目标函数

等效于最小化$\|\mathbf{w}\|^2$ ：

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

$$s.t. \quad t_n \left(\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_n) + b\right) \geq 1$$

以上是一个具有凸二次目标函数、仅有线性约束的优化问题。
这个优化问题可以用商业QP软件来解决。

拉格朗日对偶(Lagrange duality)

引入拉格朗日乘子$a_n \geq 0$，每个约束条件都对应一个乘子a_n。

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n \left\{ t_n(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n) + b) - 1 \right\}$$

令$L(\mathbf{w}, b, \boldsymbol{a})$关于$\mathbf{w}$和$b$的偏导数等于零，可获得以下两个条件：

$$\begin{aligned} \mathbf{w} &= \sum_{n=1}^{N} a_n t_n \phi(\mathbf{x}_n) \\ 0 &= \sum_{n=1}^{N} a_n t_n. \end{aligned}$$

拉格朗日对偶(Lagrange duality)

引入拉格朗日乘子$a_n \geq 0$，每个约束条件都对应一个乘子a_n。

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2}\underbrace{\|\mathbf{w}\|^2}_{\mathbf{w} = \sum_{n=1}^{N} a_n t_n \phi(\mathbf{x}_n)} - \sum_{n=1}^{N} \underbrace{a_n \{t_n}_{=0}(\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n) + b) - 1\}$$

使用这些条件从$L(\mathbf{w}, b, \boldsymbol{a})$中消除$\mathbf{w}$和$b$，可得到对偶表示$\tilde{L}(\mathbf{a})$：

$$\sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m t_n t_m \underbrace{\boldsymbol{\phi}^T(\mathbf{x}_n)\boldsymbol{\phi}(\mathbf{x}_m)}_{k(\boldsymbol{x}_n, \boldsymbol{x}_m)}$$

拉格朗日对偶(Lagrange duality)

使用核函数$k(\boldsymbol{x},\boldsymbol{x}) = \phi(\boldsymbol{x})^T\phi(\boldsymbol{x})$，可得到对偶问题：

$$\widetilde{L}(\mathbf{a}) = \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m t_n t_m k(\mathbf{x}_n, \mathbf{x}_m)$$

$$\begin{aligned} a_n &\geqslant 0, \qquad n = 1, \ldots, N, \\ \sum_{n=1}^{N} a_n t_n &= 0. \end{aligned}$$

对偶表示是一个二次规划问题。

具有N个参数：$\{a_1, \ldots, a_N\}$，从而能够对新的数据点进行分类。

$$y = \mathbf{w}^T\phi(\mathbf{x}) + b = \sum_{n=1}^{N} a_n t_n \underbrace{\phi(\mathbf{x}_n)^T\phi(\mathbf{x})}_{k(\boldsymbol{x},\boldsymbol{x}_n)} + b$$

KKT条件

这种形式的约束性优化满足*Karush-Kuhn-Tucker*(KKT)条件，在这种情况下，它要求以下三个性质成立：

$$\begin{aligned} a_n &\geqslant 0 \\ t_n y(\mathbf{x}_n) - 1 &\geqslant 0 \\ a_n \{t_n y(\mathbf{x}_n) - 1\} &= 0. \end{aligned}$$

根据 $a_n \{t_n y(\mathbf{x}_n) - 1\} = 0$，可以得到对于任意数据点，有$a_n = 0$或$t_n y(\boldsymbol{x}_n) = 1$。$a_n = 0$的任何数据点不会出现在目标函数中，其余满足$t_n y(\boldsymbol{x}_n) = 1$的数据点称为支持向量。

$$\widetilde{L}(\mathbf{a}) = \sum_{n=1}^{N} a_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} a_n a_m t_n t_m k(\mathbf{x}_n, \mathbf{x}_m)$$

求解a和b

求解二次规划问题，并找到$\boldsymbol{a}$的值。然后根据$t_n y(\boldsymbol{x}_n) = 1$，确定任意支持向量的$b$值。

$$t_n \left(\sum_{m \in \mathcal{S}} a_m t_m k(\mathbf{x}_n, \mathbf{x}_m) + b \right) = 1$$

乘t_n（利用$t_n^2 = 1$）：

$$t_n \left(\sum_{m \in \mathcal{S}} a_m t_m k(\mathbf{x}_n, \mathbf{x}_m) + b \right) = 1$$

然后可以得到：

$$b = \frac{1}{N_{\mathcal{S}}} \sum_{n \in \mathcal{S}} \left(t_n - \sum_{m \in \mathcal{S}} a_m t_m k(\mathbf{x}_n, \mathbf{x}_m) \right)$$

拉格朗日对偶(Lagrange duality)

使用核函数$k(\boldsymbol{x},\boldsymbol{x}) = \phi(\boldsymbol{x})^T\phi(\boldsymbol{x})$，可得到对偶问题：

$$\widetilde{L}(\mathbf{a}) = \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m t_n t_m k(\mathbf{x}_n, \mathbf{x}_m)$$

$$\begin{aligned} a_n &\geqslant 0, \qquad n = 1, \ldots, N, \\ \sum_{n=1}^{N} a_n t_n &= 0. \end{aligned}$$

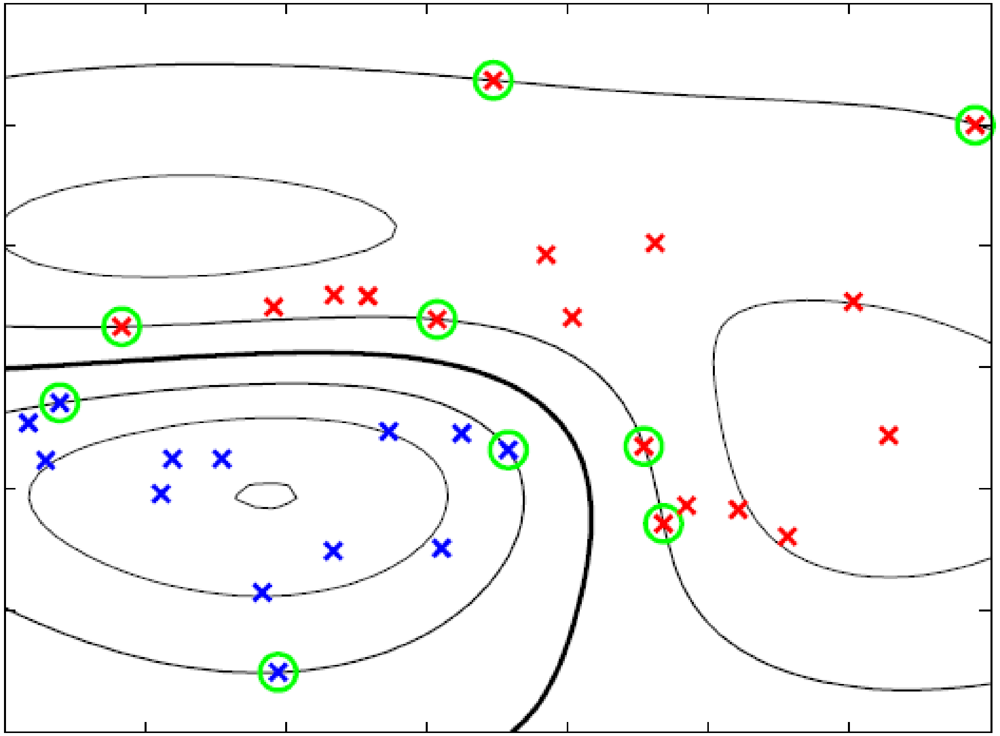
乘t_n（利用$t_n^2 = 1$）：
$$t_n\left(\sum_{m\in\mathcal{S}} a_m t_m k(\mathbf{x}_n, \mathbf{x}_m) + b\right) = 1$$

然后可以得到：

$$b = \frac{1}{N_{\mathcal{S}}}\sum_{n\in\mathcal{S}}\left(t_n - \sum_{m\in\mathcal{S}} a_m t_m k(\mathbf{x}_n, \mathbf{x}_m)\right)$$

支持向量(Support vectors)

- 决策边界（the decision boundary）
- 边界界限（The margin boundaries）
- 支持向量（the support vectors）



核函数(Kernels)

假设一个核函数：

$$k\left(\mathbf{x},\mathbf{x}'\right)=\left(\mathbf{x}^T\mathbf{x}'\right)^2$$

可将其改写为：

$$k\left(\mathbf{x},\mathbf{x}'\right)=\left(\sum_{i=1}^{D}x_ix_i'\right)\left(\sum_{j=1}^{D}x_jx_j'\right)$$

$$=\sum_{i=1}^{D}\sum_{j=1}^{D}x_ix_jx_i'x_j'$$

$$=\sum_{i,j=1}^{D}\left(x_ix_j\right)\left(x_i'x_j'\right)$$

$$k(\boldsymbol{x}, \boldsymbol{x}') = \phi(\boldsymbol{x})^T \phi(\boldsymbol{x})$$

其中特征映射（对于D=3）为：

$$\phi(\boldsymbol{x}) = \begin{pmatrix} x_1 x_1 \\ x_1 x_2 \\ x_1 x_3 \\ x_2 x_1 \\ x_2 x_2 \\ x_2 x_3 \\ x_3 x_1 \\ x_3 x_2 \\ x_3 x_3 \end{pmatrix}$$

核函数(Kernels)

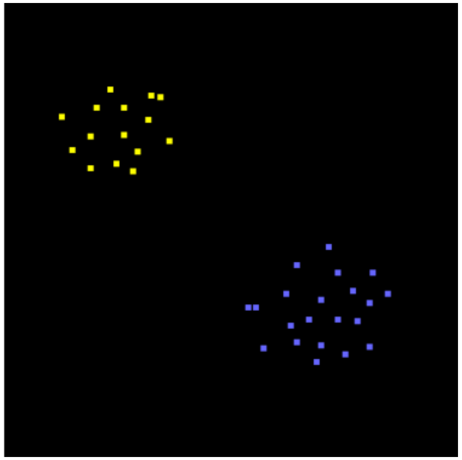
- 多项式核函数：$$k(\mathbf{x}, \mathbf{x}') = \left(\mathbf{x}^{\mathrm{T}}\mathbf{x}' + c\right)^{M}$$

- 高斯核函数：$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\|\mathbf{x} - \mathbf{x}'\|^2 / 2\sigma^2\right)$$

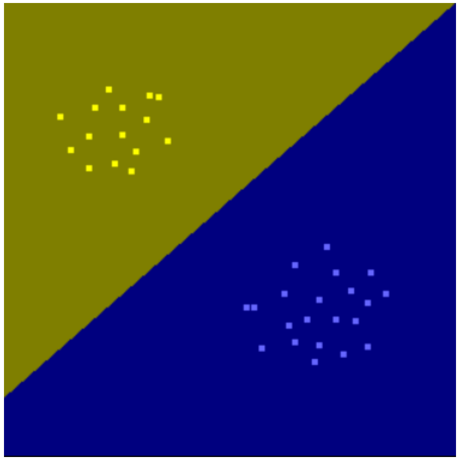
- Sigmoid核函数：$$k(\mathbf{x}, \mathbf{x}') = \tanh\left(a\mathbf{x}^{\mathrm{T}}\mathbf{x}' + b\right)$$

核函数(Kernels)

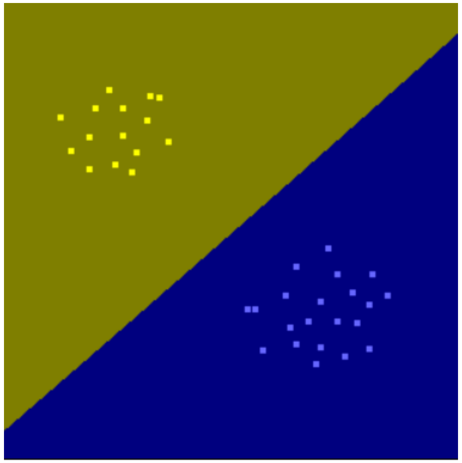
- 不同核函数对不同数据分布表现各异



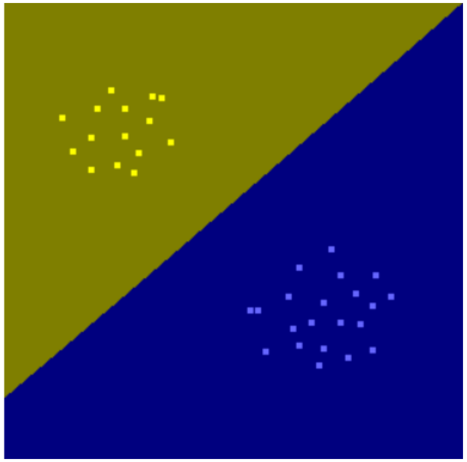
原始数据分布



线性核分类结果



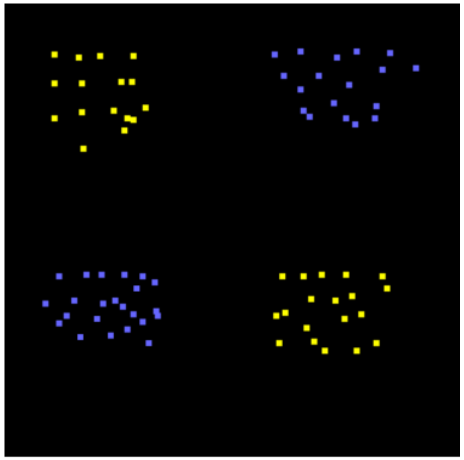
多项式核分类结果



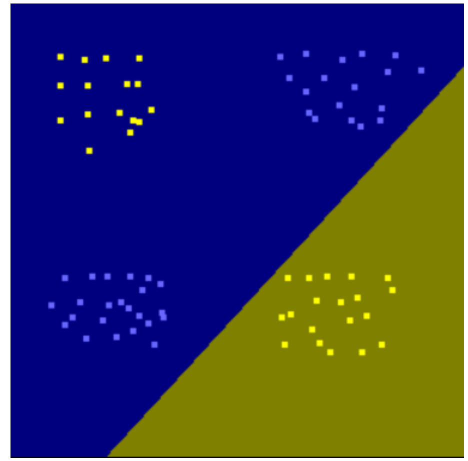
径向基核分类结果

核函数(Kernels)

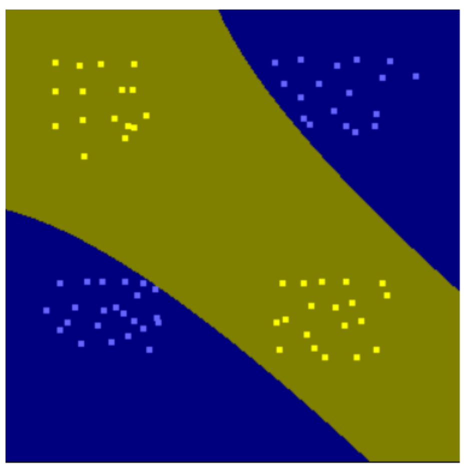
- 不同核函数对不同数据分布表现各异



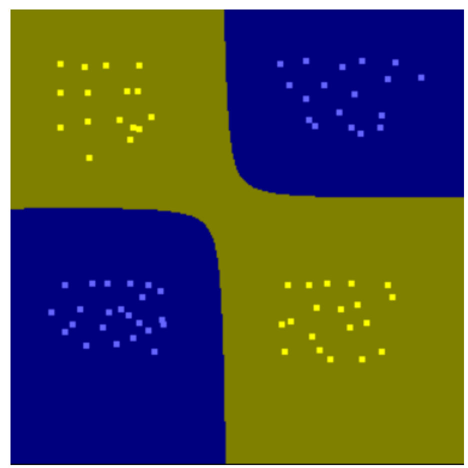
原始数据分布



线性核分类结果



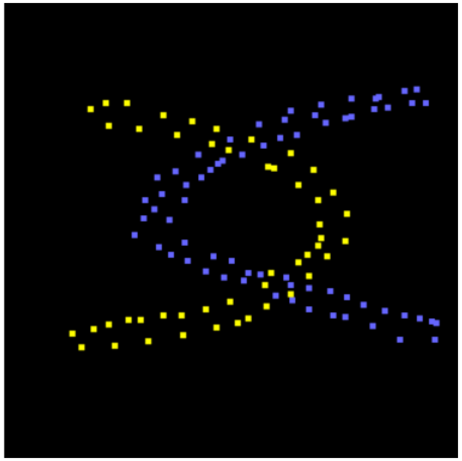
多项式核分类结果



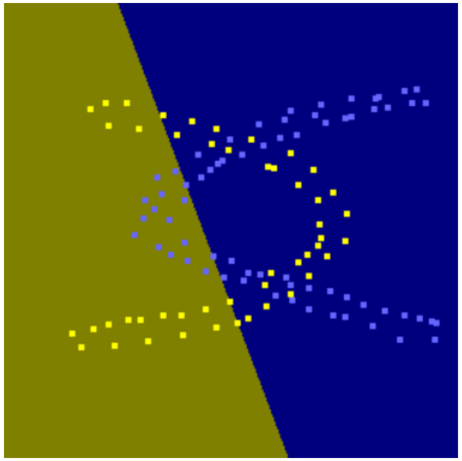
径向基核分类结果

核函数(Kernels)

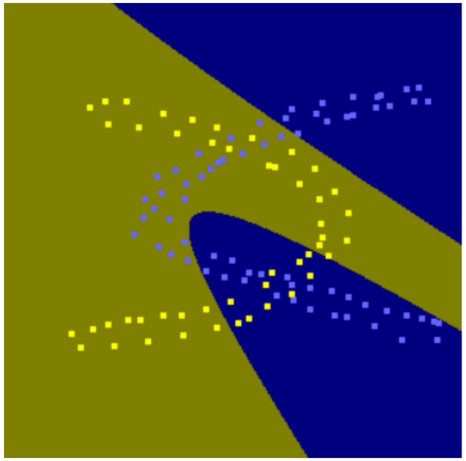
- 不同核函数对不同数据分布表现各异



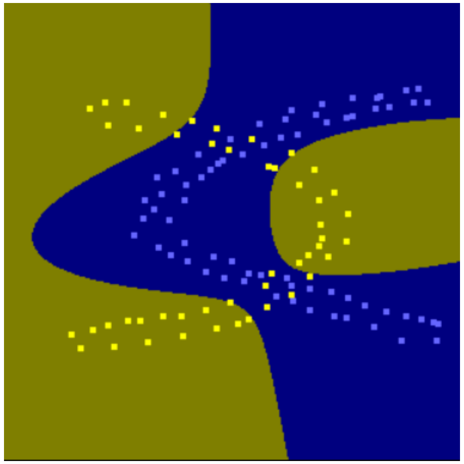
原始数据分布



线性核分类结果



多项式核分类结果



径向基核分类结果

支持向量机

- 利用LIBSVM工具包实现分类器训练和测试

➢ 第一步：下载LIBSVM工具包(见如下链接)

工具包链接：http://ivipc.uestc.edu.cn/

➢ 第二步：将解压的工具包放入matlab目录下，如

C: MATLAB\R2014a\toolbox

➢ 第三步：添加路径至当前工作区，如

p=‘C: MATLAB\R2014a\toolbox\libsvm\matlab’;

addpath(genpath(p));

➢ 第四步：编译工具包

mex –setup；

make;

支持向量机

- 利用LIBSVM工具包实现分类器训练和测试
- 第五步：训练和测试

[heart_scale_label,heart_scale_inst]=libsvmread('heart_scale');%读取数据

标签 特征

model = svmtrain(heart_scale_label,heart_scale_inst, '-c 1 -g 0.07');训练模型

分类模型 参数设置

[predict_label, accuracy, dec_values] = svmpredict(heart_scale_label, heart_scale_inst, model); 测试

预测标签 分类精度 决策值

支持向量机

- 作业

➢利用LIBSVM工具包，编写MATLAB程序，对下列a1a性别分类数据集进行SVM分类，对比线性核，多项式核以及径向基核分类的差异。

➢数据集链接：

http://ivipc.uestc.edu.cn/

主要内容

- 线性判别分类(Linear Discriminants for Classification)
- 分类举例
- 线性判别函数
    - 两类问题
    - 多类问题
- Fisher线性判别
- 支持矢量机(Support Vector Machine)
- 自适应增强AdaBoost(Adaptive Boosting)

AdaBoost算法

- 挑战：

➢对任意机器学习算法或策略，出错在所难免；

➢一种机器学习算法或策略难以始终实现最优；

- 思路：

➢集成多个机器学习算法或策略的结果，彼此互补，整体提升。

AdaBoost算法

- 如何有效集成多个分类器？

➤用许多弱分类器自适应加权构成一个强分类器

$$y_1(\boldsymbol{x}) \in \{-1,+1\}$$
$$\vdots$$
$$y_M(\boldsymbol{x}) \in \{-1,+1\}$$

自适应权重

$$Y_M(\boldsymbol{x}) = \mathrm{sign}\left(\sum_{m=1}^{M} \alpha_m y_m(\boldsymbol{x})\right)$$

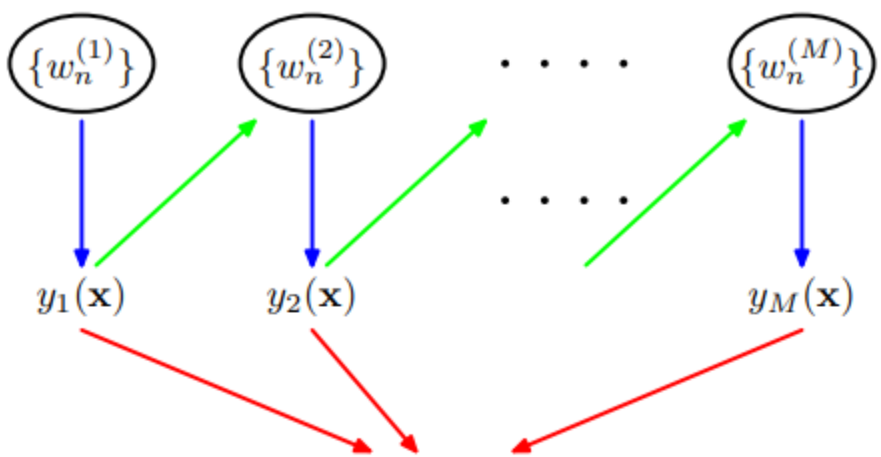
弱分类器　　　　强分类器

单个弱分类器可能只稍微好于随机分类效果

AdaBoost算法

$$y_1(\boldsymbol{x}) \in \{-1, +1\}$$
$$\vdots$$
$$y_M(\boldsymbol{x}) \in \{-1, +1\}$$

- 需要解决的问题：
  - 如何选择弱分类器？
  - 如何为训练样本分配权重？
  - 如何为每个弱分类器分配权重？



$$Y_M(\mathbf{x}) = \mathrm{sign}\left(\sum_m^M \alpha_m y_m(\mathbf{x})\right)$$

AdaBoost算法

给定：$(x_1, t_1), \dots (x_N, t_N)$，

均匀初始化样本权重：$w_n^{(1)} = \frac{1}{N}, n = 1, \dots, N$

For $m = 1, \dots, M$：

- 学习弱分类器 $y_m(\boldsymbol{x})$，最小化误差 $J_m = \sum_{n=1}^{N} w_n^{(m)} I(y_m(\boldsymbol{x}_n) \neq t_n)$，

其中 $I(y_m(\boldsymbol{x}_n) \neq t_n)$为指示函数，当$y_m(\boldsymbol{x}_n) \neq t_n$时等于1，否则为0

误差越小，权重越大，重视可靠分类器

- 分类器权重： $\epsilon_m = \frac{\sum_{n=1}^{N} w_n^{(m)} I(y_m(\boldsymbol{x}_n) \neq t_n)}{\sum_{n=1}^{N} w_n^{(m)}}$， $\alpha_m = \ln\left\{\frac{1-\epsilon_m}{\epsilon_m}\right\}$

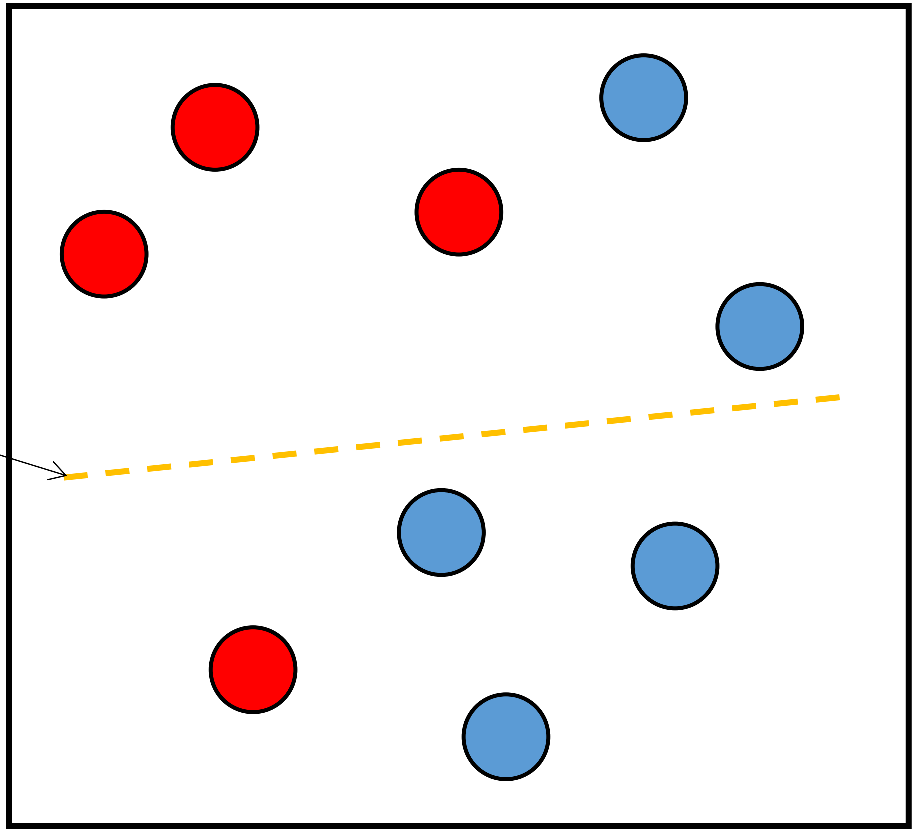
误差越大，权重越大，重视困难样本

- 更新样本权重：$w_n^{(m+1)} = w_n^{(m)} \exp\{\alpha_m I(y_m(\boldsymbol{x}_n) \neq t_n)\}$

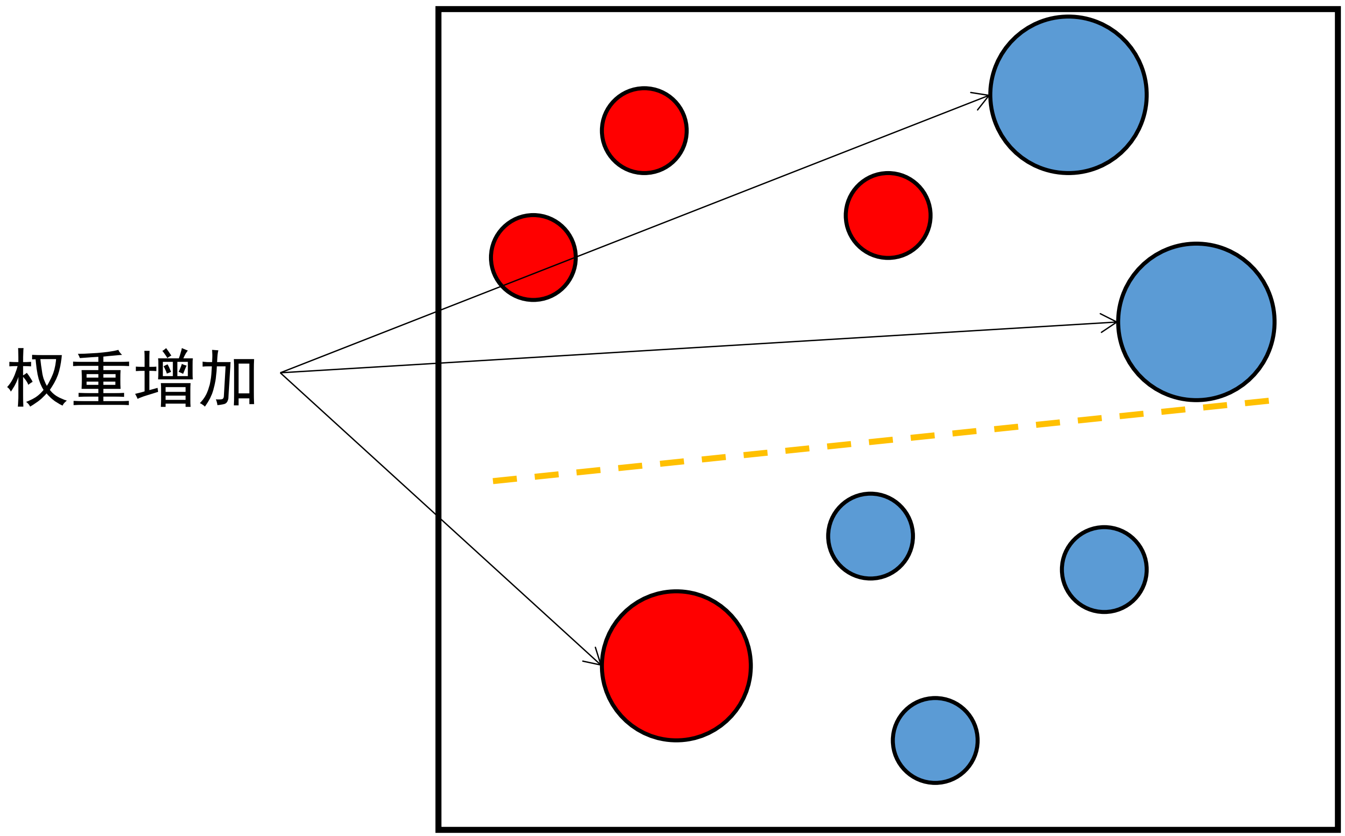
输出最终分类器：$Y_M(\boldsymbol{x}) = sign(\sum_{m=1}^{M} \alpha_m y_m(\boldsymbol{x}))$

AdaBoost算法

弱分类器1

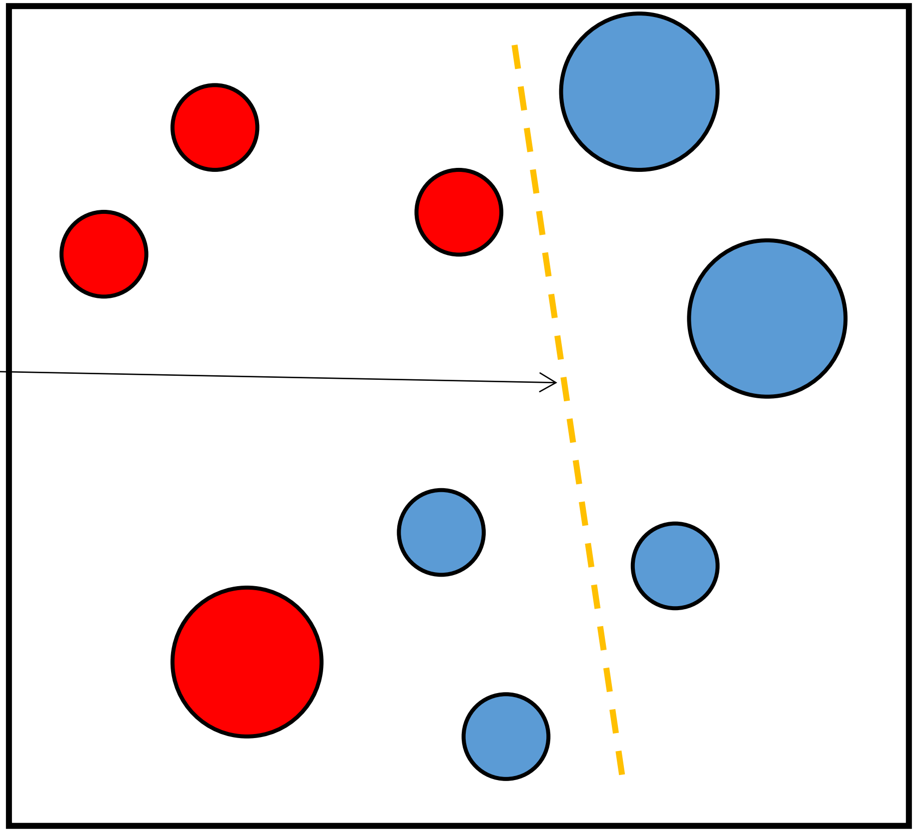


AdaBoost算法

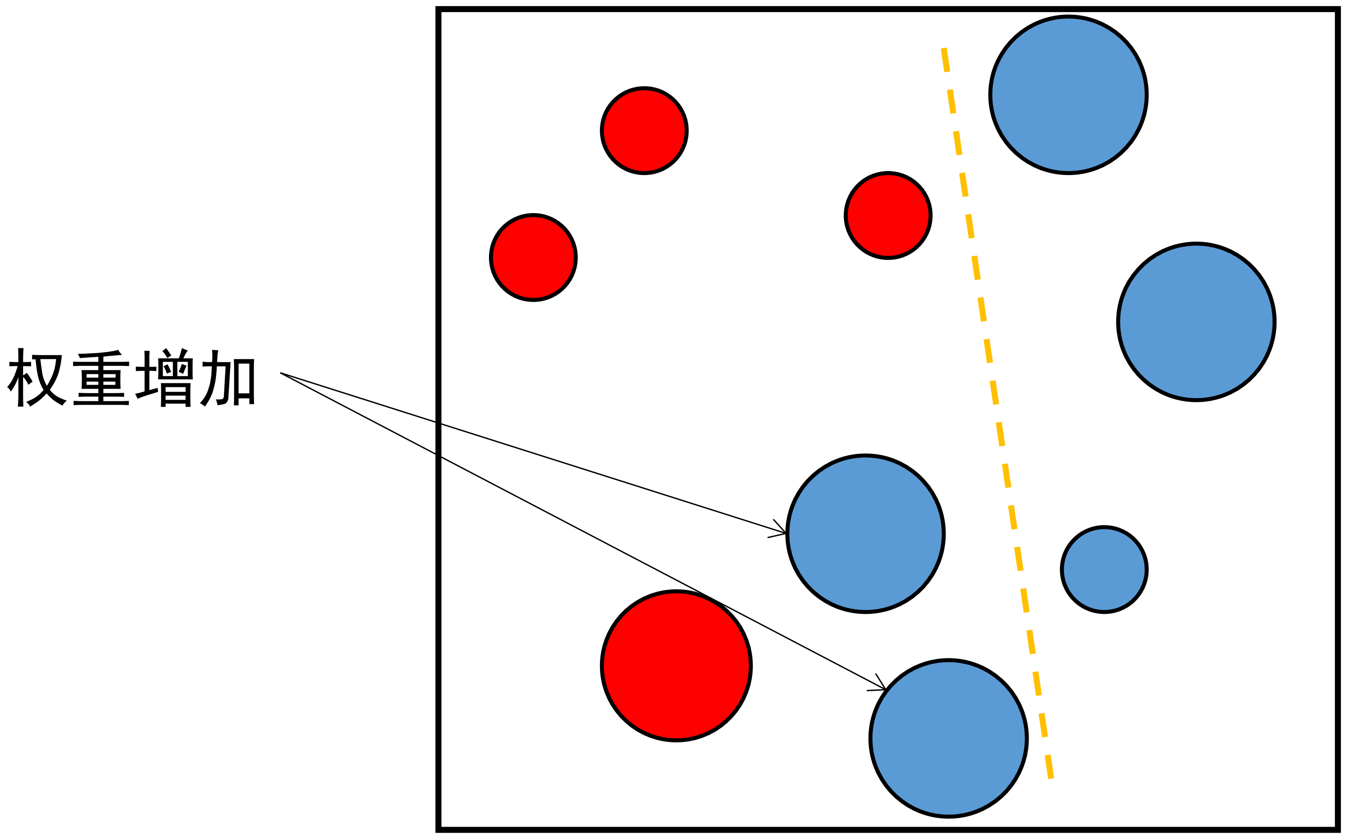


AdaBoost算法

弱分类器2

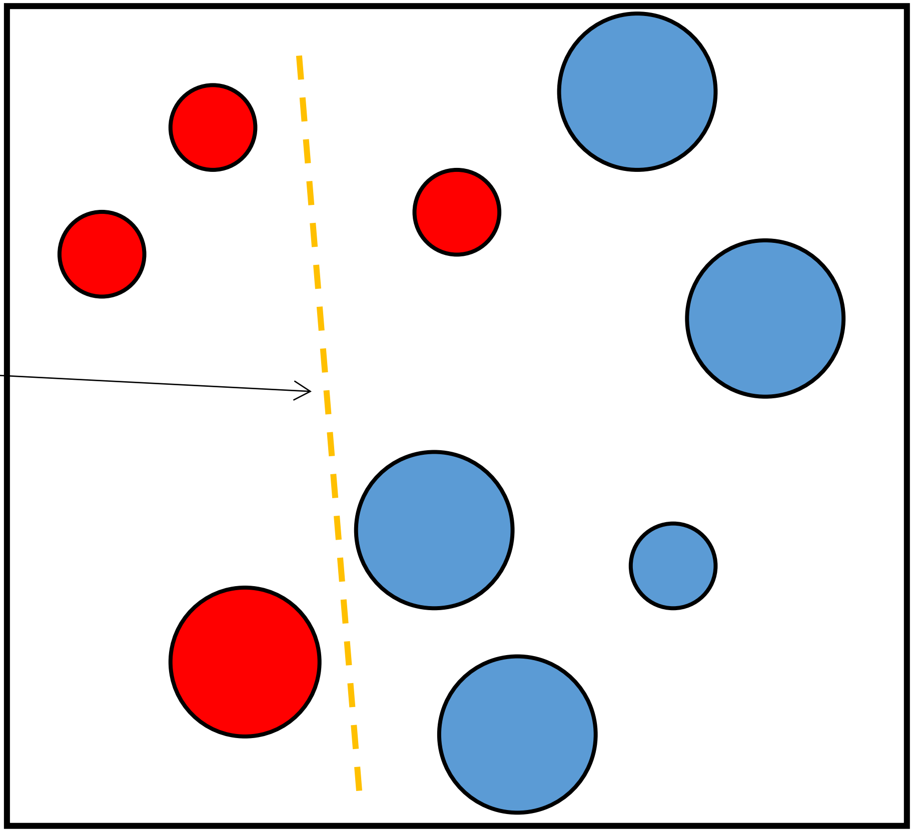


AdaBoost算法



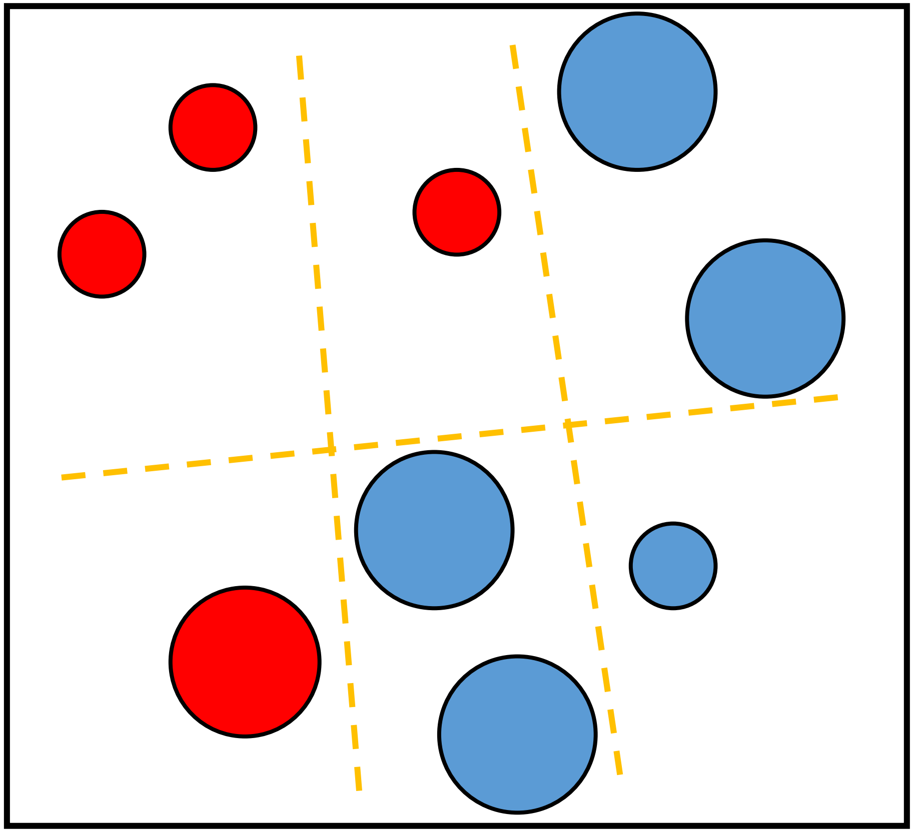
AdaBoost算法

弱分类器3



AdaBoost算法

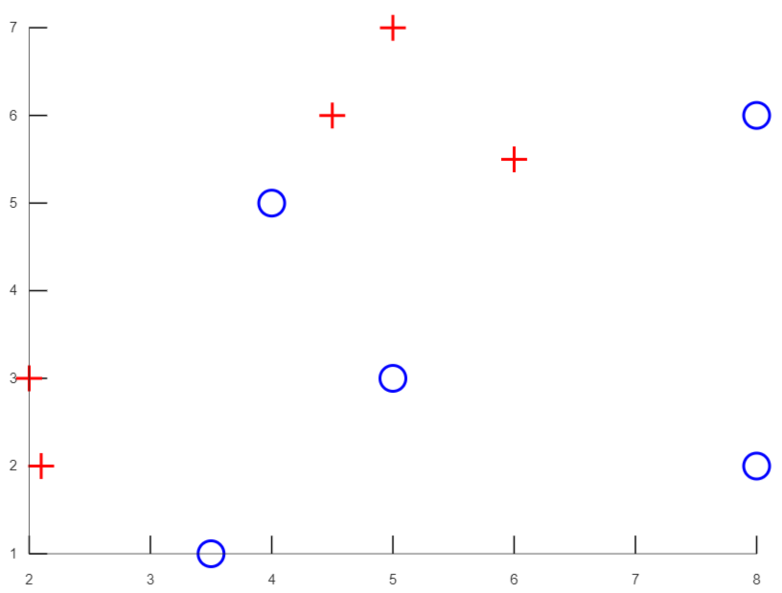
最终分类器
是弱分类器的组合



➢AdaBoost应用实例

给定训练样本：

X1	2	2.1	4.5	4	3.5	5	5	6	8	8
X2	3	2	6	5	1	7	3	5.5	6	2
t	1	1	1	-1	-1	1	-1	1	-1	-1



假设备选的弱分类器如下：

$$y_1(x_1,x_2)=\begin{cases}1, x_1<3\\-1, x_1>3\end{cases}$$

$$y_2(x_1,x_2)=\begin{cases}1, x_1<7\\-1, x_1>7\end{cases}$$

$$y_3(x_1,x_2)=\begin{cases}-1, x_2<5.1\\1, x_2>5.1\end{cases}$$

AdaBoost算法

➢弱分类器 $t=1$

X1	2	2.1	4.5	4	3.5	5	5	6	8	8
X2	3	2	6	5	1	7	3	5.5	6	2
t	1	1	1	-1	-1	1	-1	1	-1	-1
w	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

初始化训练数据权重分布：

$$w_n = \frac{1}{N} = 0.1, (N = 10, n = 1, \dots, N)$$

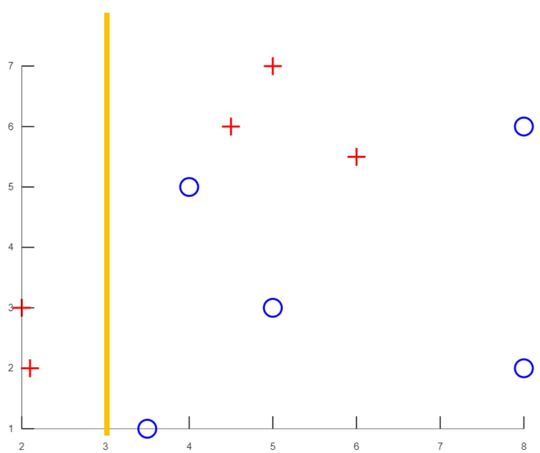
误差最小的阈值取在X1方向2. 1与3. 5之间，则第一个弱分类器应选：

$$y_1(x_1, x_2) = \begin{cases} 1, x_1 < 3 \\ -1, x_1 > 3 \end{cases}$$

➤弱分类器 $t=1$

给定训练样本（10个数据初始化权重皆为0.1）：

X1	2	2.1	4.5	4	3.5	5	5	6	8	8
X2	3	2	6	5	1	7	3	5.5	6	2
t	1	1	1	-1	-1	1	-1	1	-1	-1
w	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
y_1	1	1	-1	-1	-1	-1	-1	-1	-1	-1



则样本3、6、8为错分样本，错误率为：

$$\epsilon_1 = \frac{\sum_{n=1}^{N} w_n^{(m)} I(y_m(\boldsymbol{x}_n) \neq t_n)}{\sum_{n=1}^{N} w_n} = 0.3$$

由错误率计算得弱分类器1的权重：

$$\alpha_1 = \ln\frac{1-\epsilon_1}{\epsilon_1} = 0.847$$

AdaBoost算法

➢弱分类器t=2

更新训练样本权重： $w_n^{(2)} = w_n^{(1)} \exp\{\alpha_1 I(y_1(\boldsymbol{x}_n) \neq t_n)\}$

X1	2	2.1	4.5	4	3.5	5	5	6	8	8
X2	3	2	6	5	1	7	3	5.5	6	2
t	1	1	1	-1	-1	1	-1	1	-1	-1
w	0.1	0.1	0.233	0.1	0.1	0.233	0.1	0.233	0.1	0.1

误差最小的阈值取在x_1方向6与8之间，则第二个

弱分类器应选：

$$y_2(x_1, x_2) = \begin{cases} 1, x_1 < 7 \\ -1, x_1 > 7 \end{cases}$$

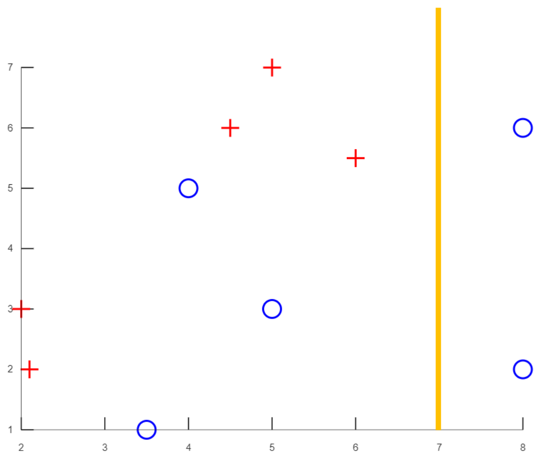
AdaBoost算法

➢弱分类器t=2

分类结果：

$$y_2(x_1,x_2)=\begin{cases}1, x_1<7\\-1, x_1>7\end{cases}$$

X1	2	2.1	4.5	4	3.5	5	5	6	8	8
X2	3	2	6	5	1	7	3	5.5	6	2
t	1	1	1	-1	-1	1	-1	1	-1	-1
w	0.1	0.1	0.233	0.1	0.1	0.233	0.1	0.233	0.1	0.1
y_2	1	1	1	1	1	1	1	1	-1	-1



则样本4、5、7为错分样本，错误率为：

$$\epsilon_2=\frac{\sum_{n=1}^{N} w_n^{(m)} I(y_m(\boldsymbol{x}_n)\neq t_n)}{\sum_{n=1}^{N} w_n}=0.214$$

由错误率计算得弱分类器2的权重：

$$\alpha_2=\ln\frac{1-\epsilon_2}{\epsilon_2}=1.301$$

AdaBoost算法

➢弱分类器t=3

更新训练样本权重： $w_n^{(3)} = w_n^{(2)}\exp\{\alpha_2 I(y_2(\boldsymbol{x}_n) \neq t_n\}$

X1	2	2.1	4.5	4	3.5	5	5	6	8	8
X2	3	2	6	5	1	7	3	5.5	6	2
t	1	1	1	-1	-1	1	-1	1	-1	-1
w	0.1	0.1	0.233	0.367	0.367	0.233	0.367	0.233	0.1	0.1

误差最小的阈值取在X2方向5与5.5之间，则第三个弱分类器应选：

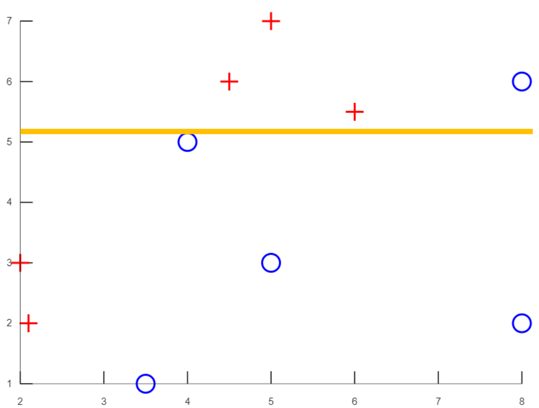
$$y_3(x_1, x_2) = \begin{cases} -1, x_2 < 5.1 \\ 1, x_2 > 5.1 \end{cases}$$

➤弱分类器t=3

$$y_3(x_1,x_2)=\begin{cases}-1, x_2<5.1\\ 1, x_2>5.1\end{cases}$$

分类结果：

X1	2	2.1	4.5	4	3.5	5	5	6	8	8
X2	3	2	6	5	1	7	3	5.5	6	2
t	1	1	1	-1	-1	1	-1	1	-1	-1
w	0.1	0.1	0.233	0.367	0.367	0.233	0.367	0.233	0.1	0.1
y_3	-1	-1	1	-1	-1	1	-1	1	1	-1



则样本1、2、9为错分样本，错误率为：

$$\epsilon_3=\frac{\sum_{n=1}^{N}w_n^{(m)}I(y_m(\boldsymbol{x}_n)\neq t_n)}{\sum_{n=1}^{N}w_n}=0.136$$

由错误率计算得弱分类器3的权重：

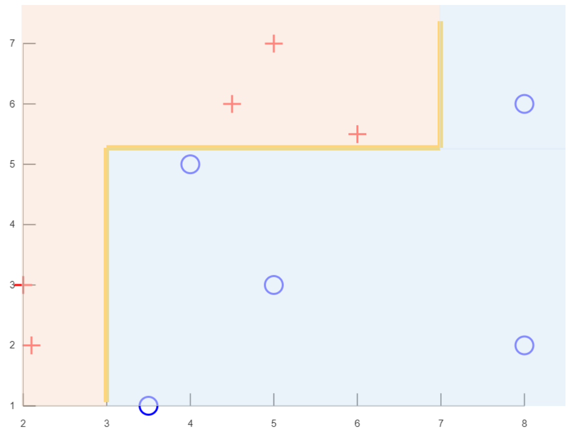
$$\alpha_3=\ln\frac{1-\epsilon_3}{\epsilon_3}=1.849$$

AdaBoost算法

➢3个弱分类器合并成一个强分类器：

$$sign(Y(x_1,x_2)) = sign(0.847y_1(x_1,x_2) + 1.301y_2(x_1,x_2) + 1.849y_3(x_1,x_2))$$

X1	2	2.1	4.5	4	3.5	5	5	6	8	8
X2	3	2	6	5	1	7	3	5.5	6	2
t	1	1	1	-1	-1	1	-1	1	-1	-1
$sign(Y)$	1	1	1	-1	-1	1	-1	1	-1	-1



在训练集上有0个误差点，训练结束

多个简单的线性分类器实现二维空间的非线性分类功能。

- AdaBoost习题

给定训练样本：

X	0	1	2	3	4	5	6	7	8	9
t	1	1	1	-1	-1	-1	1	1	1	-1

可选的弱分类器如下：

$$y(\boldsymbol{x}) = \begin{cases} 1, x < 2.5 \\ -1, x > 2.5 \end{cases} \qquad y(\boldsymbol{x}) = \begin{cases} 1, x < 8.5 \\ -1, x > 8.5 \end{cases} \qquad y(\boldsymbol{x}) = \begin{cases} -1, x < 5.5 \\ 1, x > 5.5 \end{cases}$$

试利用AdaBoost算法求取强分类器判别函数。

- 弱分类器 $t=1$

给定训练样本（10个数据初始化权重皆为0.1）：

X	0	1	2	3	4	5	6	7	8	9
t	1	1	1	-1	-1	-1	1	1	1	-1
w	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

误差最小的阈值取在2与3或者8与9之间，任取其一，

则第一个弱分类器为：

$$y_1(\boldsymbol{x}) = \begin{cases} 1, x < 2.5 \\ -1, x > 2.5 \end{cases}$$

AdaBoost算法

➢弱分类器 $t=1$

分类结果：

$$y_1(\boldsymbol{x}) = \begin{cases} 1, x < 2.5 \\ -1, x > 2.5 \end{cases}$$

X	0	1	2	3	4	5	6	7	8	9
t	1	1	1	-1	-1	-1	1	1	1	-1
w	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
y_1	1	1	1	-1	-1	-1	-1	-1	-1	-1

则样本6、7、8为错分样本，错误率为：

$$\epsilon_1 = \frac{P(y_1(\boldsymbol{x}_i) \neq t_i)}{\sum_{n=1}^{N} w_n} = 0.3$$

由错误率计算得弱分类器1的权重： $\alpha_1 = \ln\frac{1-\epsilon_1}{\epsilon_1} = 0.8473$

AdaBoost算法

➢弱分类器 $t=2$

更新训练样本权重： $w_n^{(2)}=w_n^{(1)}\exp\{\alpha_1 I(y_1(\boldsymbol{x}_n)\neq t_n\}$

X	0	1	2	3	4	5	6	7	8	9
t	1	1	1	-1	-1	-1	1	1	1	-1
w	0.1	0.1	0.1	0.1	0.1	0.1	0.233	0.233	0.233	0.1

误差最小的阈值取在8与9之间，则第二个弱分类器为：

$$y_2(\boldsymbol{x})=\begin{cases}1, x<8.5\\-1, x>8.5\end{cases}$$

AdaBoost算法

- 弱分类器 $t = 2$

分类结果：
$$y_2(\boldsymbol{x}) = \begin{cases} 1, x < 8.5 \\ -1, x > 8.5 \end{cases}$$

X	0	1	2	3	4	5	6	7	8	9
$\boldsymbol{t}$	1	1	1	-1	-1	-1	1	1	1	-1
$\boldsymbol{w}$	0.1	0.1	0.1	0.1	0.1	0.1	0.233	0.233	0.233	0.1
y_2	1	1	1	1	1	1	1	1	1	-1

则样本3、4、5为错分样本，错误率为：

$$\epsilon_2 = \frac{P(y_2(\boldsymbol{x}_i) \neq t_i)}{\sum_{n=1}^{N} w_n} = 0.2144$$

由错误率计算得弱分类器1的权重：$\alpha_2 = \ln\frac{1-\epsilon_2}{\epsilon_2} = 1.2986$

➢弱分类器 $t=3$

更新训练样本权重： $w_n^{(3)} = w_n^{(2)} \exp\{\alpha_2 I(y_2(\boldsymbol{x}_n) \neq t_n)\}$

X	0	1	2	3	4	5	6	7	8	9
t	1	1	1	-1	-1	-1	1	1	1	-1
w	0.1	0.1	0.1	0.366	0.366	0.366	0.233	0.233	0.233	0.1

误差最小的阈值取在5与6之间，则第三个弱分类器为：

$$y_3(\boldsymbol{x}) = \begin{cases} -1, x < 5.5 \\ 1, x > 5.5 \end{cases}$$

- 弱分类器 $t = 3$

分类结果： $y_3(\boldsymbol{x}) = \begin{cases} -1, x < 5.5 \\ 1, x > 5.5 \end{cases}$

X	0	1	2	3	4	5	6	7	8	9
t	1	1	1	-1	-1	-1	1	1	1	-1
w	0.1	0.1	0.1	0.366	0.366	0.366	0.233	0.233	0.233	0.1
y_3	-1	-1	-1	-1	-1	-1	1	1	1	1

则样本0、1、2、9为错分样本，错误率为：

$$\epsilon_3 = \frac{P(y_3(\boldsymbol{x}_i) \neq t_i)}{\sum_{n=1}^{N} w_n} = 0.1821$$

由错误率计算得弱分类器1的权重： $\alpha_3 = \ln \frac{1-\epsilon_3}{\epsilon_3} = 1.5022$

➢3个弱分类器合并成一个强分类器：

$$sign(Y(x)) = sign(0.8473y_1(x) + 1.2986y_2(x) + 1.5022y_3(x))$$

X	0	1	2	3	4	5	6	7	8	9
t	1	1	1	-1	-1	-1	1	1	1	-1
$sign(Y)$	1	1	1	-1	-1	-1	1	1	1	-1

在训练集上有0个误差点，训练结束