模式识别

课程团队：谢凤英 崔林艳 张浩鹏
邹征夏 李洪珏 李家军
单位：宇航学院

第二章 线性分类器

CONTENTS PAGE

2.1 基本概念和数学基础

2.1.1 决策面和判别函数

2.1.2 分类问题描述

2.1.3 模式识别的数学基础

2.1 基本概念和数学基础

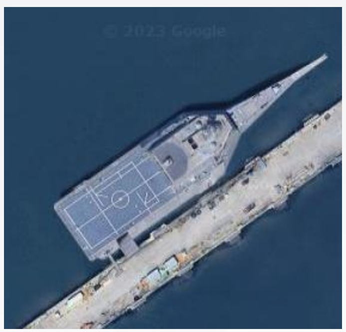
2.1.1 决策面和判别函数

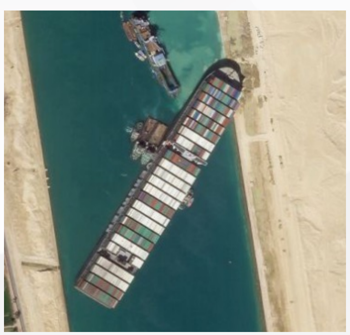
- 分类是模式识别的本质
- 基于样本集设计分类器的过程就是寻找分类决策边界的过程
- 在特征空间中，将不同种类样本区分开的决策边界，称之为“决策面”或“分类面”
- 用数学形式描述决策面的函数就是判别函数

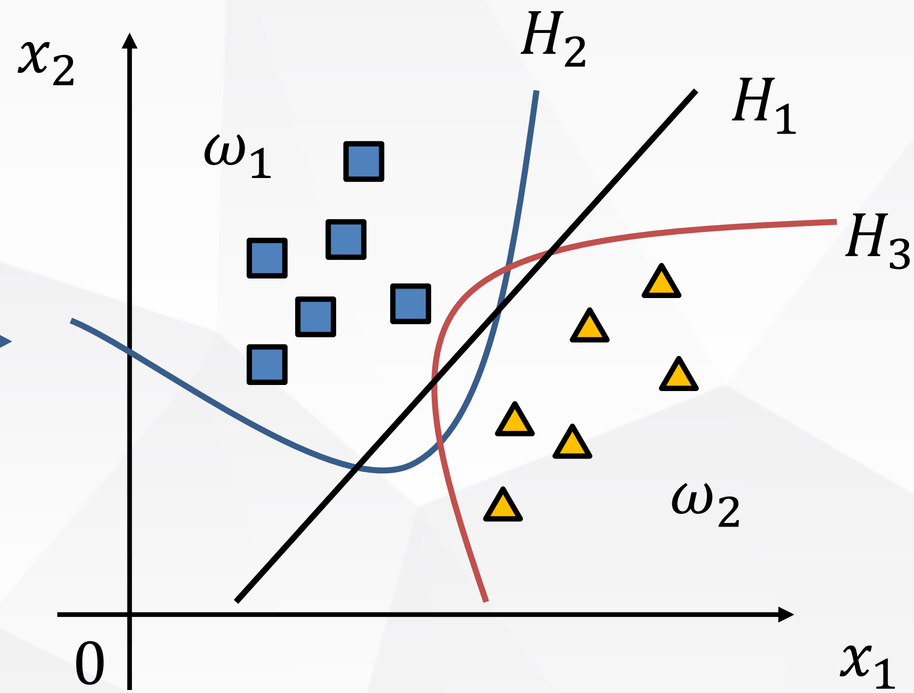
2.1 基本概念和数学基础

例：在二维特征空间中，二分类任务可以有多种决策边界（决策面、分类面）

遥感图像舰船目标二分类（军船 VS 民船）

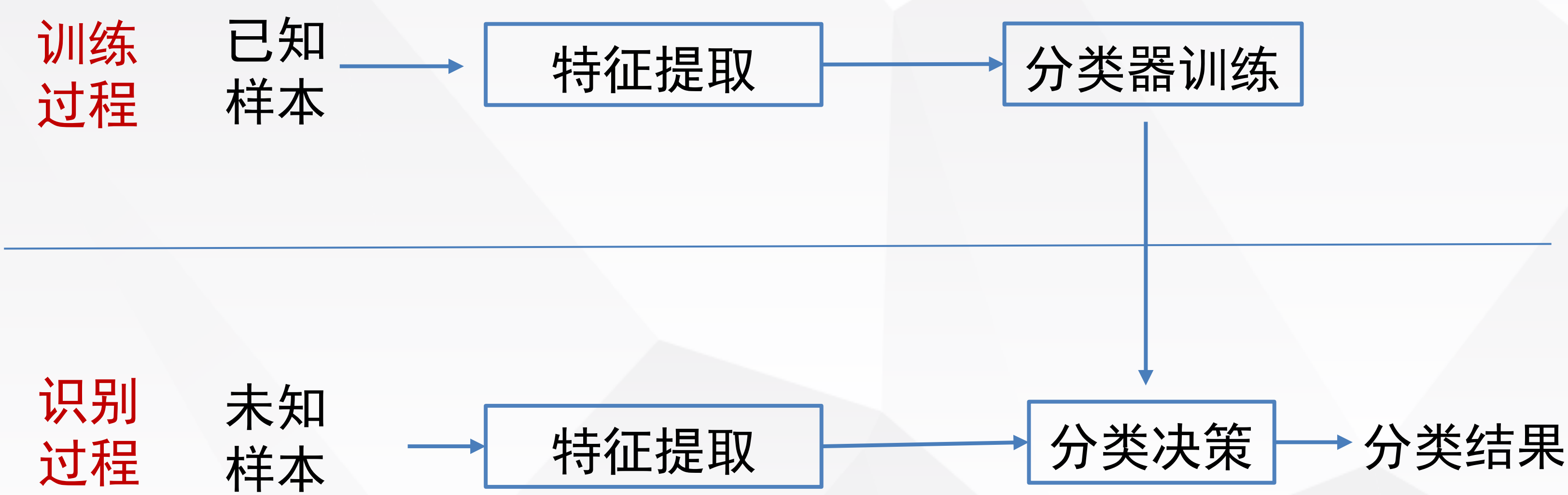






2.1 基本概念和数学基础

训练过程与识别过程



2.1 基本概念和数学基础

2.1.2 分类问题描述

- C——类别数
- D——特征维度
- N——样本数

例：

- 两类二维问题：在二维特征空间中的二分类，C=2，D=2
- 三类多维问题：在多维特征空间中的三分类，C=3，D＞2

2.1 基本概念和数学基础

2.1.3 模式识别的数学基础

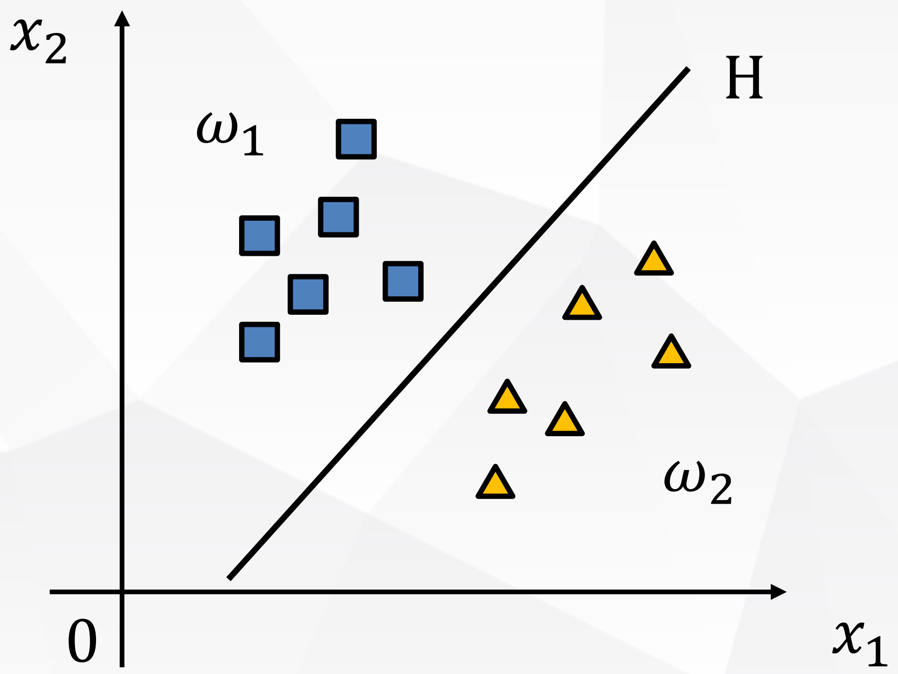
- 线性代数
  - 矩阵、向量、转置、向量运算、矩阵运算……
- 概率统计
  - 概率、条件概率、分布、概率密度函数、期望方差……
- 最优化
  - 最小二乘、梯度下降……

2.2 线性分类器基础

2.2 线性分类器基础

2.2.1 线性分类器概念

- **线性分类器**：对于两类的分类问题，采用线性判别函数划分特征空间（即采用直线或平面等将两类样本在特征空间中的区域划分开），这样的分类器是线性分类器。



2.2.2 线性判别函数

- 两类二维问题（C=2，D=2）
- 直线方程（决策面）
  - 代数形式 $$w_1x_1 + w_2x_2 + w_0 = 0$$
  - 向量形式 $$\mathbf{w}^T\mathbf{x} + w_0 = 0$$
- 定义线性判别函数

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

2.2 线性分类器基础

2.2.2 线性判别函数

- 两类多维问题（C=2，D＞2）
- 定义线性判别函数

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

$\mathbf{w}$：权重向量　　w_0：偏置系数

- 决策面H

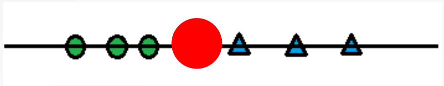
$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = 0$$

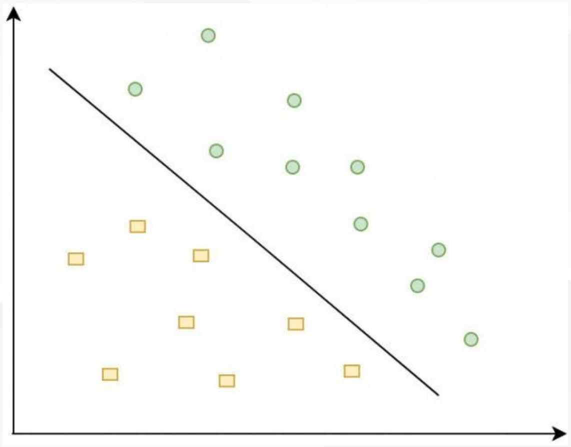
2.2 线性分类器基础

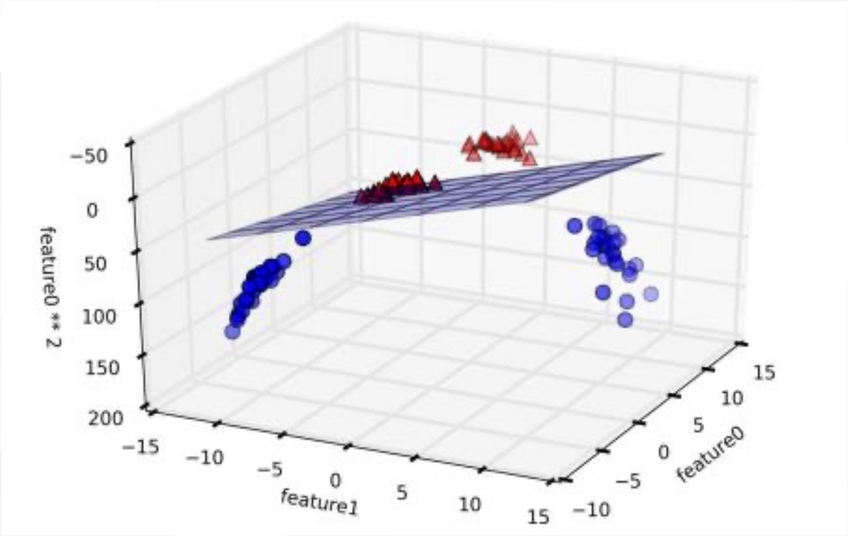
2. 2. 2 线性判别函数

➢ 线性分类器决策面H在不同维度特征空间的表现形式

特征维度	一维（D＝1）	二维（D＝2）	三维（D＝3）	多维（D＞3）
决策面情况	点	直线	平面	超平面

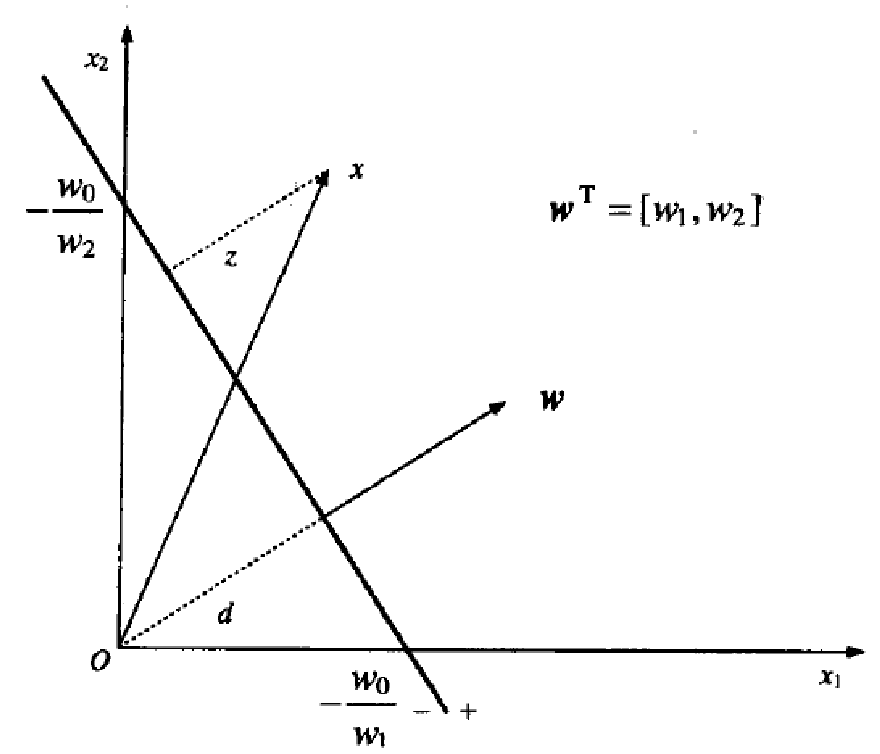






2.2.2 线性判别函数

- 线性判别函数的几何性质 $g(\mathbf{x}) = \mathbf{w}^{T}\mathbf{x} + w_0$
  - 法向量方向 $\mathbf{w}$
  - 原点距离 $|w_0| / \|\mathbf{w}\|$



2. 2. 3 增广变换

- 线性判别函数 $g(\mathbf{x})=\mathbf{w}^{T}\mathbf{x}+w_0$ 的增广变换

  - 对原特征x定义增广变换

$$\mathbf{y}=\begin{bmatrix}1\\ x_1\\ x_2\\ \vdots\\ x_d\end{bmatrix}=\begin{bmatrix}1\\ \mathbf{x}\end{bmatrix}$$

  - 对应权向量 $\mathbf{a}=\begin{bmatrix}w_0\\ w_1\\ w_2\\ \vdots\\ w_d\end{bmatrix}=\begin{bmatrix}w_0\\ \mathbf{w}\end{bmatrix}$

$$g(\mathbf{x})=\mathbf{a}^{T}\mathbf{y}$$

2.2.3 增广变换

- 线性判别函数 $g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0$ 的增广变换为

$$g(\mathbf{x}) = \mathbf{a}^T\mathbf{y}$$

$\mathbf{a}$：增广后权重向量　　　y：增广后偏置系数

- 增广变换的特点
  - 特征维数增加了一维：$D_G=D+1$
  - 样本之间的欧氏距离保持不变
  - 增广变换后的决策面 $\mathbf{a}^T\mathbf{y} = 0$ 是过原点的超平面H_G

2.2.3 增广变换

- 决策规则
  - 已知判别函数 $g(\mathbf{x}) = \mathbf{w}^{T}\mathbf{x} + w_0$ 或 $g(\mathbf{x}) = \mathbf{a}^{T}\mathbf{y}$
  - 决策规则为：

对于未知样本x，

若g(x) ＞ 0，则x决策为ω_1类（正类，特征空间正侧）

若g(x) ＜ 0，则x决策为ω_2类（负类，特征空间负侧）

2. 2. 4 线性分类器设计

- 线性分类器设计常规步骤
  - 给定类别已知的样本——训练样本集
  - 选择一个准则函数J，其值反映分类器性能（分类结果优劣）
  - 采用求最优解的数学方法求准则函数J的极值解，从而求得权重向量**w**和偏置w_0，或增广权重向量**a**

2.2.4 归纳小结

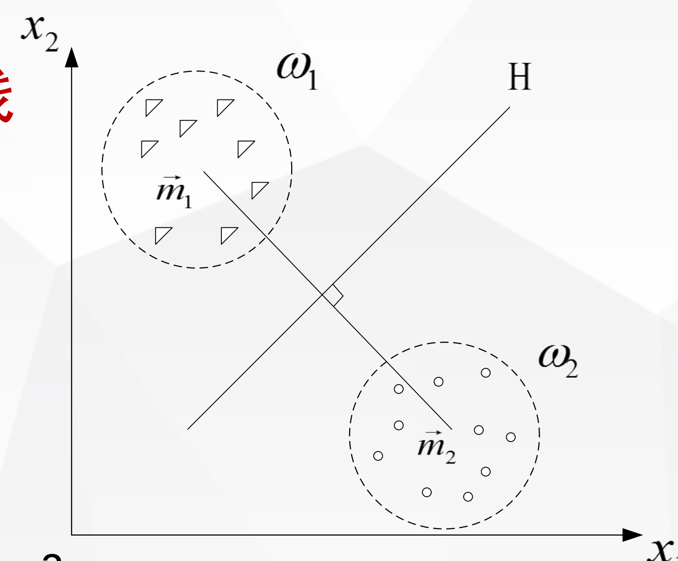
- 线性分类器特点：简单高效，适合解决线性可分问题
- 采用增广表达可使问题表达形式更加简洁
- 线性分类器的核心在于求解权重向量和偏置

2.3 垂直平分分类器

2.3 垂直平分分类器

2.3.1 问题与思路

- 二分类问题（C=2），已知训练样本集X有N个样本，其中
  - ω$_1$类样本有N_1个，样本集用X_1表示；
  - ω$_2$类样本有N_2个，样本集用X_2表示
- 垂直平分分类器设计思路：
  - 基于两类样本均值点作垂直平分线



2

2.3 垂直平分分类器

2.3.2 垂直平分分类器求解

- 判别函数与决策面方程
  - 垂直平分线性判别函数

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

  - 决策面方程（垂直平分直线方程）

$$g(\mathbf{x}) = 0 \quad \text{即} \quad \mathbf{w}^{T}\mathbf{x} + w_0 = 0$$

2.3.2 垂直平分分类器求解

- 求解权向量与阈值权
  - 先求均值向量$\mathbf{m}_1$和$\mathbf{m}_2$
  - 利用垂直几何关系，设权向量

$$\mathbf{w} = \mathbf{m}_1 - \mathbf{m}_2$$

  - 则直线方程为

$$(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{x} + w_{\mathbf{0}} = 0$$

正侧在$\mathbf{m}_1$侧

2.3.2 垂直平分分类器求解

- 求解权向量与阈值权
  - 再利用平分几何关系，中点$\mathbf{x}_0$在直线上

$$\mathbf{x}_0 = (\mathbf{m}_1 + \mathbf{m}_2)/2$$

  - 代入方程求得偏置量

$$w_0 = -(\mathbf{m}_1 - \mathbf{m}_2)^{\mathrm{T}}(\mathbf{m}_1 + \mathbf{m}_2)/2$$

2.3.2 垂直平分分类器求解

- 最终结果

  - 线性判别函数

$$g(\mathbf{x}) = (\mathbf{m}_1 - \mathbf{m}_2)^{\mathrm{T}} \mathbf{x} - (\mathbf{m}_1 - \mathbf{m}_2)^{\mathrm{T}} (\mathbf{m}_1 + \mathbf{m}_2) / 2$$
$$= (\mathbf{m}_1 - \mathbf{m}_2)^{\mathrm{T}} (\mathbf{x} - (\mathbf{m}_1 + \mathbf{m}_2) / 2)$$

  - 决策面方程

$$(\mathbf{m}_1 - \mathbf{m}_2)^{\mathrm{T}} (\mathbf{x} - (\mathbf{m}_1 + \mathbf{m}_2) / 2) = 0$$

2.3 垂直平分分类器

2.3.2 垂直平分分类器求解

- 决策规则
  - 已知垂直平分判别函数

$$g(\mathbf{x}) = (\mathbf{m}_1 - \mathbf{m}_2)^T(\mathbf{x} - (\mathbf{m}_1 + \mathbf{m}_2)/2)$$

  - 垂直平分决策规则为

对于未知样本$\mathbf{x}$，若$g(\mathbf{x}) > 0$，则$\mathbf{x}$决策为ω_1类

若$g(\mathbf{x}) < 0$，则$\mathbf{x}$决策为ω_2类

2.3.2 垂直平分分类器求解

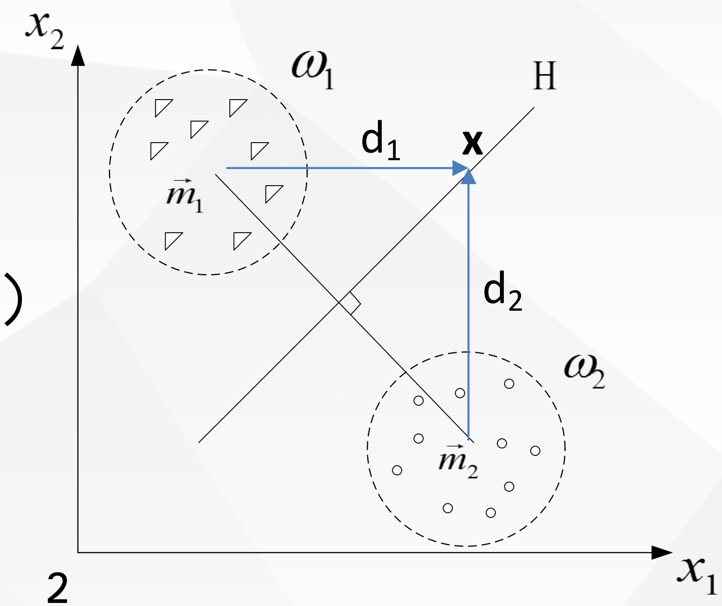
- 推广到两类多维问题（C=2，D＞2）判别函数与决策面方程形式不变
  - 垂直平分线性判别函数

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

  - 垂直平分决策面方程

$$g(\mathbf{x})=0 \qquad 即 \qquad \mathbf{w}^T\mathbf{x} + w_0 = 0$$

2.3.3 最小距离等价决策



- 定义判别函数为欧氏距离（非线性）

$$G_1(\mathbf{x}) = d_1(\mathbf{x}) = \|\mathbf{x} - \mathbf{m}_1\|$$

$$G_2(\mathbf{x}) = d_2(\mathbf{x}) = \|\mathbf{x} - \mathbf{m}_2\|$$

2

- 垂直平分分类器等价的最小距离决策规则为

  对于未知样本x，若$d_1(\mathbf{x}) < d_2(\mathbf{x})$，则x决策为ω$_1$类

  若$d_1(\mathbf{x}) > d_2(\mathbf{x})$，则x决策为ω$_2$类

- 垂直平分分类器又叫最小距离分类器

2.3 垂直平分分类器

2.3.4 归纳小结

- 垂直平分分类器的主要特点
  - 解决二分类问题的线性分类器，简单直观
  - 未采用准则函数求极值解（非最佳决策）
  - 垂直平分分类器可等价为最小距离分类器

2.4 Fisher投影准则

2.4.1 问题与思路

2.4.2 Fisher准则函数

2.4.3 准则函数化简

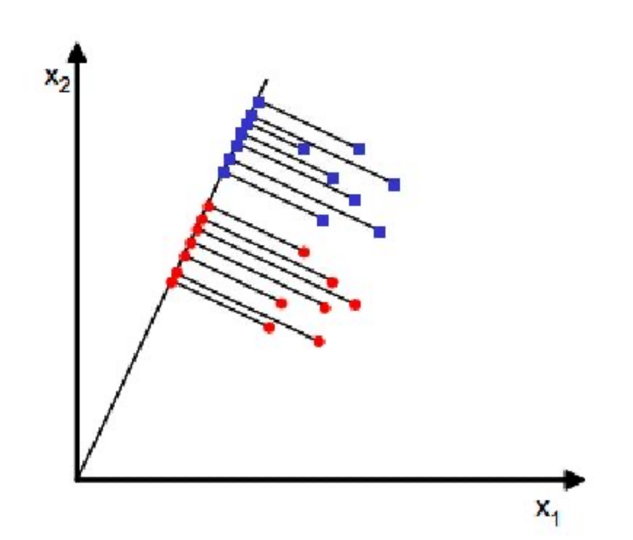
2.4.4 求极值解

2.4.5 归纳小结

2.4 Fisher投影准则

2.4.1 问题与思路

- 高维问题：特征个数多、特征空间维数高、计算量大
  - 用经典理论设计分类器困难



- 设计思路
  - 通过投影对高维分类问题降维
  - Fisher将高维特征空间的样本投影到一维直线上

2.4 Fisher投影准则

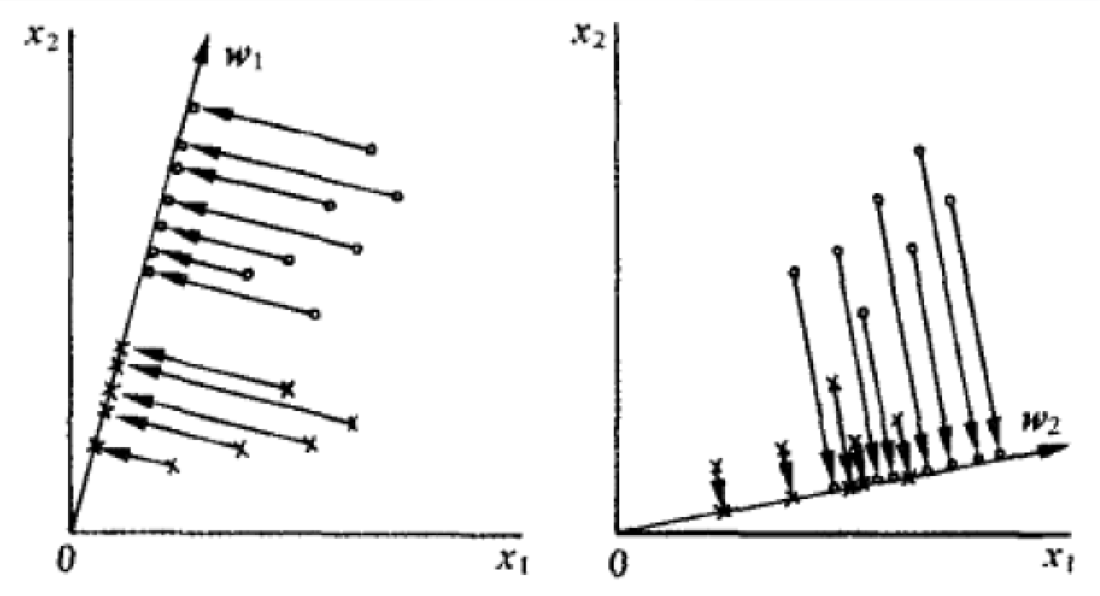
2. 4. 1 问题与思路

- 问题：
  - 已知C=2，D维分类问题的样本集
  - 设投影向量为**w**
  - 则一维投影方程为 $y = \mathbf{w}^{T}\mathbf{x}$
  - 求最佳投影向量**w**

2.4 Fisher投影准则

2.4.2 Fisher准则函数

- Fisher投影准则的物理含义
  - 投影后异类样本尽量远离
  - 投影后同类样本尽量靠近



2.4 Fisher投影准则

2.4.2 Fisher准则函数

- 在原特征空间X中
  - 类均值向量 $$\mathbf{m}_i = \frac{1}{N_i}\sum_{\mathbf{x}_j \in X_i} \mathbf{x}_j, i = 1,2$$
  - 类内离散度矩阵 $$\mathbf{S}_i = \sum_{\mathbf{x}_j \in X_i} (\mathbf{x}_j - \mathbf{m}_i)(\mathbf{x}_j - \mathbf{m}_i)^T, i = 1,2$$
  - 总类内离散度矩阵 $$\mathbf{S}_w = \mathbf{S}_1 + \mathbf{S}_2$$
  - 类间离散度矩阵 $$\mathbf{S}_b = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$$

2.4 Fisher投影准则

2.4.2 Fisher准则函数

- 在投影后的一维特征空间Y中
  - 类均值 $$\tilde{m}_i = \frac{1}{N_i}\sum_{y_j \in Y_i} y_j = \frac{1}{N_i}\sum_{\mathbf{x}_j \in X_i} \mathbf{w}^T\mathbf{x}_j = \mathbf{w}^T\mathbf{m}_i, i = 1,2$$
  - 类内离散度 $$\tilde{S}_i^2 = \sum_{y_j \in Y_i} (y_j - \tilde{m}_i)^2, i = 1,2$$
  - 总类内离散度 $$\tilde{S}_w^2 = \tilde{S}_1^2 + \tilde{S}_2^2$$
  - 类间离散度 $$\tilde{S}_b^2 = (\tilde{m}_1 - \tilde{m}_2)^2$$

2.4.2 Fisher准则函数

- Fisher准则函数定义

$$J_F(\mathbf{w})=\frac{\tilde{S}_b^2}{\tilde{S}_w^2}=\frac{(\tilde{m}_1-\tilde{m}_2)^2}{\tilde{S}_1^2+\tilde{S}_2^2}$$

- Fisher准则函数值越大，投影结果越符合预期
  - 投影后异类样本尽量远离
  - 投影后同类样本尽量靠近

2. 4. 3 准则函数化简

$$J_F(\mathbf{w})=\frac{\tilde{S}_b^2}{\tilde{S}_w^2}=\frac{(\tilde{m}_1-\tilde{m}_2)^2}{\tilde{S}_1^2+\tilde{S}_2^2}$$

➢ 分子化简

$$\begin{aligned}\tilde{S}_b^2&=(\tilde{m}_1-\tilde{m}_2)^2\\&=(\mathbf{w}^T\mathbf{m}_1-\mathbf{w}^T\mathbf{m}_2)^2\\&=\mathbf{w}^T(\mathbf{m}_1-\mathbf{m}_2)(\mathbf{m}_1-\mathbf{m}_2)^T\mathbf{w}\\&=\mathbf{w}^T\mathbf{S}_b\mathbf{w}\end{aligned}$$

2.4 Fisher投影准则

2.4.3 准则函数化简

$$J_F(\mathbf{w}) = \frac{\tilde{S}_b^2}{\tilde{S}_w^2} = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{\tilde{S}_1^2 + \tilde{S}_2^2}$$

➢ 分母化简

$$\begin{aligned}
\tilde{S}_w^2 &= \tilde{S}_1^2 + \tilde{S}_2^2 \\
&= \sum_{\mathbf{x}_j \in X_1} (\mathbf{w}^T \mathbf{x}_j - \mathbf{w}^T \mathbf{m}_1)^2 + \sum_{\mathbf{x}_j \in X_2} (\mathbf{w}^T \mathrm{x}_j - \mathbf{w}^T \mathbf{m}_2)^2 \\
&= \sum_{\mathbf{x}_j \in X_1} \mathbf{w}^T (\mathbf{x}_j - \mathbf{m}_1)(\mathbf{x}_j - \mathbf{m}_1)^T \mathbf{w} + \sum_{\mathrm{x}_j \in X_2} \mathbf{w}^T (\mathbf{x}_j - \mathbf{m}_2)(\mathbf{x}_j - \mathbf{m}_2)^T \mathbf{w} \\
&= \mathbf{w}^T \mathbf{S}_1 \mathbf{w} + \mathbf{w}^T \mathbf{S}_2 \mathbf{w} \\
&= \mathbf{w}^T \mathbf{S}_w \mathbf{w}
\end{aligned}$$

2.4 Fisher投影准则

2.4.3 准则函数化简

$$J_F(\mathbf{w}) = \frac{\tilde{S}_b^2}{\tilde{S}_w^2} = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{\tilde{S}_1^2 + \tilde{S}_2^2}$$

- 化简结果

$$\Downarrow$$

$$J_F(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}}$$

- 这一表达式被称为广义Rayleigh商
- 问题的求解可等价为一个“极大化广义Rayleigh商 ”的最优化问题

2. 4. 4 求极值解

- 令分母为非零常数，最大化分子部分，转化为等式约束条件下求极值问题

$$\max \ \mathbf{w}^{\mathrm{T}}\mathbf{S}_{\mathrm{b}}\mathbf{w}$$

$$\text{s.t.} \quad \mathbf{w}^{\mathrm{T}}\mathbf{S}_{\mathrm{w}}\mathbf{w} = \mathrm{c} \neq 0$$

- 引入Lagrange乘子

$$\mathrm{L}(\mathbf{w},\lambda) = \mathbf{w}^{\mathrm{T}}\mathbf{S}_{\mathrm{b}}\mathbf{w} - \lambda(\mathbf{w}^{\mathrm{T}}\mathbf{S}_{\mathrm{w}}\mathbf{w} - \mathrm{c})$$

2.4.4 求极值解

- 采用Lagrange乘子法求Fisher准则函数的极值
  - 求偏导数
  - 令偏导数为零

$$L(\mathbf{w}, \lambda) = \mathbf{w}^T \mathbf{S}_b \mathbf{w} - \lambda(\mathbf{w}^T \mathbf{S}_w \mathbf{w} - c)$$

$$\frac{\partial L(\mathbf{w}, \lambda)}{\partial \mathbf{w}} = \mathbf{S}_b \mathbf{w} - \lambda \mathbf{S}_w \mathbf{w}$$

$$\mathbf{S}_b \mathbf{w}^* - \lambda \mathbf{S}_w \mathbf{w}^* = 0$$

$$\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{w}^* = \lambda \mathbf{w}^*$$

2. 4. 4 求极值解

- 采用Lagrange乘子法求Fisher准则函数的极值
  - 重构$\mathrm{S_b}$

$$\mathbf{S}_{\mathrm{b}}\mathbf{w}^{*}=(\mathbf{m}_1-\mathbf{m}_2)(\mathbf{m}_1-\mathbf{m}_2)^{\mathrm{T}}\mathbf{w}^{*}=(\mathbf{m}_1-\mathbf{m}_2)\mathrm{R}$$

标量常数 $\mathrm{R}=(\mathbf{m}_1-\mathbf{m}_2)^{\mathrm{T}}\mathbf{w}^{*}$

$$\mathbf{w}^{*}=\frac{\mathrm{R}}{\lambda}\mathbf{S}_{\mathrm{w}}^{-1}(\mathbf{m}_1-\mathbf{m}_2)$$

- Fisher准则函数的极值解（极大值）

$$\mathbf{w}^{*}=\mathbf{S}_{\mathrm{w}}^{-1}(\mathbf{m}_1-\mathbf{m}_2)$$

2.4 Fisher投影准则

2.4.5 归纳小结

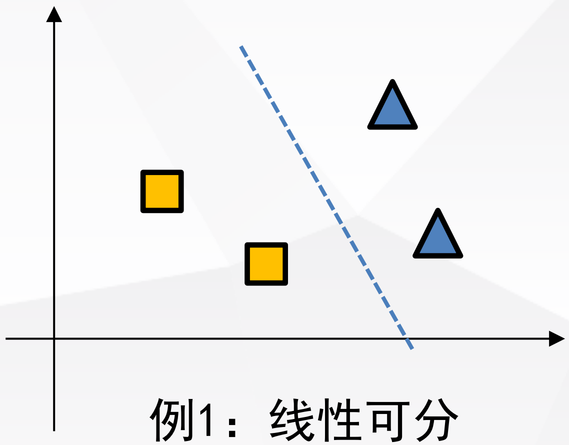
- Fisher投影的特点
  - 解决两类问题的线性投影，分类器设计较容易
  - 原则上对样本集无特殊要求（S_w矩阵可逆）
  - 采用Fisher投影准则函数求极值解（最佳决策）

2.5 感知准则

2.5 感知准则

2.5.1 线性可分问题

- 样本集的线性可分性
  - 若训练样本集可以被某个线性分类器完全正确分类，则该样本集是线性可分的。
  - 否则样本集就是线性不可分的



例1：线性可分

例2：线性不可分（异或问题）

2.5 感知准则

2.5.2 解向量和解区

➢ 解向量

能将线性可分样本集中的每个样本都正确分类的权向量。

➢ 解区

解向量往往不是一个，而是由无穷多个解向量组成的区域，称为解区。

2.5.3 感知准则函数

- 线性可分性样本集的规范化

  对ω_2类（负类）样本的增广向量全部乘以-1

- 规范化之后的分类结果

  $\mathbf{a}^T\mathbf{y} > 0$，正确分类

  $\mathbf{a}^T\mathbf{y} < 0$，错误分类

2.5 感知准则

2.5.3 感知准则函数

- Rosenblatt定义感知准则函数（感知机）
  - 对于规范化的增广样本集， $\mathbf{a}^T\mathbf{y} < 0$ 表示错误分类
  - 定义感知准则函数作为优化目标函数

$$J_p(\mathbf{a}) = \sum_{a^T y_k \le 0} (-\mathbf{a}^T \mathbf{y}_k)$$

  - 求解向量（或解区）
  - 梯度下降法求解

2.5.4 求极值解

- 用梯度下降法求解感知准则函数极值（极小值）
  - 先求梯度方向

$$\nabla J_p(\mathbf{a}) = \frac{\partial J_p(\mathbf{a})}{\partial \mathbf{a}} = \sum_{\mathbf{a}^T\mathbf{y}_k \le 0} (-\mathbf{y}_k)$$

  - 得到迭代公式

$$\mathbf{a}(t+1) = \mathbf{a}(t) - \rho_t \nabla J_p(\mathbf{a}) = \mathbf{a}(t) + \rho_t \sum_{\mathbf{a}^T\mathbf{y}_k \le 0} \mathbf{y}_k$$

2.5.5 归纳小结

- 感知准则（分类器）的特点
  - 解决两类分类的线性分类器构造问题
  - 样本集必须是线性可分的
  - 采用感知准则函数求极值解（最优决策）
- 感知机的后续研究
  - 人工神经网络
  - 支持向量机SVM

2.6 最小错分样本数准则

2.6 最小错分样本数准则

2.6.1 问题与思路

- 问题：
  - 感知准则只适用线性可分样本集——无错分
  - 实际情况未必线性可分——有错分
  - 另外线性可分的判断也很困难
  - 既然存在错分样本——求错分样本数最少

2.6 最小错分样本数准则

2.6.1 问题与思路

- 数学描述：
  - 规范化的增广样本集，$\mathbf{a}^T\mathbf{y} < 0$ 表示错误分类
  - 设样本数为N， N个不等式联立

$$\mathbf{a}^T\mathbf{y}_i > 0, \quad i = 1, \ldots, N$$

  - 求满足不等式最多的解（权向量）

2.6 最小错分样本数准则

2.6.1 问题与思路

➢ 数学描述：

➢ 写成矩阵形式 $\mathbf{Ya} > 0$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^T \\ \mathbf{y}_2^T \\ \vdots \\ \mathbf{y}_N^T \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1d} \\ y_{21} & y_{22} & \cdots & y_{2d} \\ \cdots & \cdots & \cdots & \cdots \\ y_{N1} & y_{N2} & \cdots & y_{Nd} \end{bmatrix}$$

引入余量 $\mathbf{b} > 0$

$$\mathbf{Ya} \geq \mathbf{b} > 0$$

2.6 最小错分样本数准则

2.6.2 最小错分准则

➢ 准则函数一

$$J_{q2}(\mathbf{a})=\sum_{i=1}^{N}\frac{1+\operatorname{sgn}(\mathbf{y}_i^T\mathbf{a})}{2}\qquad \operatorname{sgn}(\mathbf{y}_i^T\mathbf{a})=\begin{cases}+1, & \text{对于}\mathbf{y}_i^T\mathbf{a}\ge 0\\ -1, & \text{对于}\mathbf{y}_i^T\mathbf{a}\le 0\end{cases}$$

➢ 准则函数二

$$J_{q1}(\mathbf{a})=\left\|(\mathbf{Y}\mathbf{a}-\mathbf{b})-\left|\mathbf{Y}\mathbf{a}-\mathbf{b}\right|\right\|^2$$

➢ 求极值解：搜索算法、梯度下降法

2.6 最小错分样本数准则

2. 6. 3 归纳小结

- 最小错分样本数准则（分类器）的特点
  - 解决两类问题的线性分类器
  - 样本集不限，可以是线性不可分的
  - 求满足不等式个数最多的权向量（最优）

2.7 最小平方误差准则

2.7 最小平方误差准则

2.7.1 问题与思路

- 问题：
  - 对于线性不可分问题
  - 最小错分样本数准则——求错分样本数最少
  - 最小平方误差准则——实际应用中往往是求误差平方和最小

2.7 最小平方误差准则

2.7.1 问题与思路

- 数学描述：
  - 规范化的增广样本集，$\mathbf{a}^T\mathbf{y} < 0$ 表示错误分类
  - 引入余量$\mathbf{b}_i$，将不等式组改造为等式组

$$\mathbf{a}^T\mathbf{y}_i = \mathbf{b}_i > 0,\ i = 1,\ \ldots,\ N$$

  - 求满足等式组的最小平方误差解（权向量）

2.7 最小平方误差准则

2.7.1 问题与思路

- 数学描述：
  - 写成矩阵形式 $\mathbf{Y}\mathbf{a}=\mathbf{b}$

$$\mathbf{Y}=\begin{bmatrix}\mathbf{y}_1^{\mathrm{T}}\\ \mathbf{y}_2^{\mathrm{T}}\\ \vdots\\ \mathbf{y}_N^{\mathrm{T}}\end{bmatrix}=\begin{bmatrix}y_{11} & \cdots & y_{1d}\\ \vdots & \ddots & \vdots\\ y_{N1} & \cdots & y_{Nd}\end{bmatrix}$$

$$\mathbf{b}=[b_1\,b_2\cdots b_N]^{\mathrm{T}}$$

2.7.2 最小平方误差准则

➢ 定义决策误差

$$\mathbf{e} = \mathbf{Y}\mathbf{a} - \mathbf{b}$$

➢ 定义最小平方误差准则函数

$$J_s(\mathbf{a}) = \left\|\mathbf{Y}\mathbf{a} - \mathbf{b}\right\|^2 = \sum_{i=1}^{N} (\mathbf{a}^T \mathbf{y}_i - b_i)^2$$

2.7.2 最小平方误差准则

➢ 直接求极值解

$$\nabla J_s(\mathbf{a}) = 2\mathbf{Y}^{\mathrm{T}}(\mathbf{Y}\mathbf{a} - \mathbf{b}) = 0$$

$$\mathbf{a}^* = (\mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{b} = \mathbf{Y}^{+}\mathbf{b}$$

$$\mathbf{Y}^{+} = (\mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{-1}\mathbf{Y}^{\mathrm{T}}$$

2.7.3 余量的选择

- 选择不同余量的结果

例1

$$\mathbf{b} = \begin{bmatrix} N/N_1 \\ \vdots \\ N/N_1 \\ N/N_2 \\ \vdots \\ N/N_2 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} N \\ \vdots \\ N \end{matrix}}\right\} N_1\text{个} \\ \left.\vphantom{\begin{matrix} N \\ \vdots \\ N \end{matrix}}\right\} N_2\text{个} \end{matrix}$$

- 等价于Fishier解

2.7 最小平方误差准则

2.7.3 余量的选择

- 选择不同余量的结果

例2

$$\mathbf{b}_i = 1, \quad i = 1, \quad \ldots, \quad N$$

- 当N趋于无穷时，逼近Bayes解

2.7 最小平方误差准则

2. 7. 4 归纳小结

- 最小平方误差准则（分类器）的特点
  - 解决两类问题的线性分类器
  - 样本集不限，可以是线性不可分的
  - 求最小平方误差的权向量，存在闭式解

2.8 多分类算法设计

2.8.1 问题与思路

2.8.2 一对一法

2.8.3 一对多法

2.8.4 有向无环图法

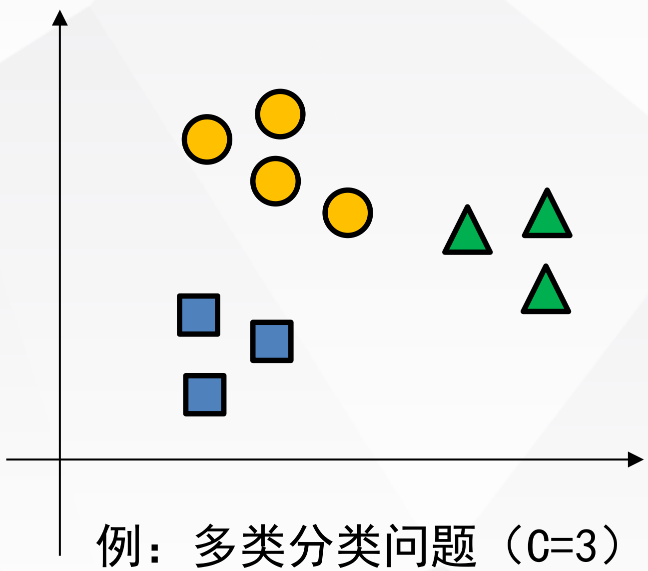
2.8.5 归纳小结

2.8 多分类算法设计

2.8.1 问题与思路

- 问题：现实世界中要处理的分类问题往往是多类别的

- 思路：
  - 基于二类分类算法构造多类分类器



例：多类分类问题（C=3）

2.8.2 一对一法（One-Versus-One，OVO）

- 原理：类别两两组合，一对一调用模型进行训练和分类

  第一个分类器只回答“是第1类还是第2类”

  第二个分类器只回答“是第1类还是第3类”

  ……

  c个类别，则需构造c(c-1)/2个分类器，投票多的为胜者

- 缺点：类别数多时，分类器数量太多，假如有1000个类别，则需约要500,000个分类器

2.8.3 一对多法（One-Versus-Rest，OVR）

- 原理：某个类别作为正类，剩余类作为反类

  c个类别的样本可以构造出c个二分类器

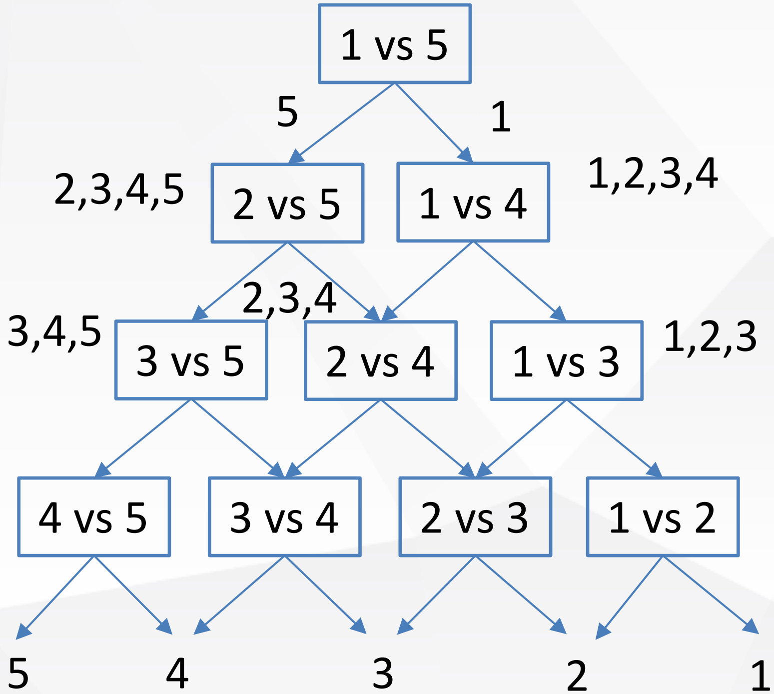
  分类时将未知样本分类为具有最大分类函数值的那类

- 优点：所需分类器个数较少，分类速度相对较快

- 缺点：若“其余”一类数据大，会导致“类别不平衡”问题

2.8.4 有向无环图法（Directed Acyclic Graph，DAG）

- 原理：属一对一策略中的一种方法，采用一对一方法训练，但按有向无环图组织分类器进行分类



有向无环图法示意图

- 优点：分类速度快
- 缺点：如最一开始就分类错误，则后面的分类器无法纠正错误

2.8.4 有向无环图法（Directed Acyclic Graph，DAG）

一对一法&有向无环图法对比

	一对一法（OVO）	一对多法（OVR）	有向无环图法（DAG）
训练分类器	c(c-1)/2个	c个	c(c-1)/2个
调用分类器	c(c-1)/2个	c个	c-1个
优点	设计思想简单	训练和识别速度快	识别阶段速度快
缺点	不适合类别较多场景	可能出现类别不平衡	如最一开始就分类错误，则后面的分类器无法纠正错误

- 假如有1000个类别:
- OVO法训练约50万个分类器，分类时调用这50万个分类器，按投票多者统计
- DAG法训练50万个分类器，分类时调用999个分类器

2.7.5 归纳小结

- 组合多个二类分类器是解决多分类问题的一个常用思路
- 一对一、一对多、有向无环图三种方式都可以用于解决多分类问题
- 三种方法各有优缺点，实际问题中可根据各自的特点进行方法的选择

作业

已知

甲类：$[0\ \ 3]^T$、$[2\ \ 4]^T$、$[1\ \ 3]^T$、$[2\ \ 3]^T$、$[0\ \ 2]^T$

乙类：$[4\ \ 1]^T$、$[3\ \ 2]^T$、$[2\ \ 1]^T$、$[3\ \ 0]^T$、$[3\ \ 1]^T$

待分类样本为 $x = [5\ \ 0]^T$

试用垂直平分判别函数对x进行决策，判断其属于哪一类。

作业

对于二维线性判别函数$g(x) = 3x_1 + 2x_2 - 5$

1）将判别函数写成向量形式$g(x) = w^T x + w_0$

2）映射成广义齐次线性函数$f(y) = w_1^T y$, 其中

$$y = (y_1, y_2, y_3)^T = (x_1, x_2, 1)^T$$

3）证明上述X空间只是Y空间的一个子空间，且$f(y) = 0$对X空间的划分结果与$g(x) = 0$相同。

作业

已 知 两 类 $\omega_1:(0,0)^T,(1,2)^T,\omega_2:(1,-1)^T,(3,0)^T$ ， 请 用 Fisher准则构造分类器，确定最佳投影方向。