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

分类模型

第一部分 线性分类模型

张朝晖

2018-2019学年 20180911

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

1.分类模块

产生式分类模型 A.贝叶斯分类模型

判别式分类模型

线性分类模型
- B. Fisher判别分类
- C. 感知器分类模型
- D. 大间隔分类模型(线性*SVM*)

非线性分类模型
- E. 核*SVM*(非线性*SVM*)
- F. 核Fisher判别分类
- G. 神经网络

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

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

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

4.集成学习
- *U*.*Bagging*
- *V*.随机森林
- *W*.*Boosting*

5.特征工程
- *X*.主成分分析(*PCA*)
- …

6.评价模块 混淆矩阵(及其相关指标)、ROC曲线、交叉验证

主要内容

A.引言

B. Fisher线性判别分析

C.感知器

D. 线性支持向量机(线性SVM)

A.引言

1. 有关约定

关键词：

- 线性运算、非线性运算
- 函数、方程
- 线性/非线性
- 特征空间及其划分
- 决策域、决策边界(决策面、分类边界)
- 产生式分类模型、判别式(鉴别式)分类模型
- 线性/非线性

  分类器、分类模型、分类边界、判别函数
- 两类别分类、多类别分类

2.基于模型的分类器设计 -- 代表模型：贝叶斯分类模型

基本思路：

(1)基于训练集，估计描述每个类别的概率密度(参数)

(2)观测样本的类别决策

训练样本 → $\boxed{\text{估计}\,p(x\,|\,\omega_j)}$ → 贝叶斯判决

如何估计了类条件概率密度函数

(1)**参数估计法**

需明确类条件概率密度函数形式；

参数估计需要大量样本

(2)**非参数估计法**

不足：$p(x\,|\,\omega_j)$的估计需要大量样本；

随特征空间维数增加，占用大量存储空间

==>**贝叶斯决策尽管最优，但实现困难**

3. 基于样本（数据）的直接分类器设计

例：鉴别式分类模型，由训练样本集，直接确定决策域划分

这里我们首先关注利用样本数据设计鉴别式分类模型！！！

基本思路

明确
- 训练样本集
- 分类器（判别函数）类型
  - 线性判别函数
  - 非线性判别函数
- 分类器设计的目标准则
- 算法

$\longrightarrow$ $\boxed{\text{利用样本，估计判别函数的未知参数使所选目标准则最优}}$

分类器设计 $\Leftarrow$ $\boxed{\text{判别函数参数估计}}$ $\Rightarrow$ 准则函数极值解求取

4. 鉴别式分类模型 $\begin{cases}\text{线性分类模型}\\ \text{非线性分类模型}\end{cases}$

先从简单情况开始：
面向两类别分类的“线性分类模型”

线性分类模型的分类边界是关于$\boldsymbol{x}$的线性方程.

线性判别函数

$$\begin{cases}\text{对于多类：} g_i(\boldsymbol{x}) = \boldsymbol{w}_i^T\boldsymbol{x} + \omega_{i0} \qquad \boldsymbol{i} = 1,\ldots,\boldsymbol{c} \\ \qquad\qquad \boldsymbol{c}\text{个判别函数} \\ \text{对于两类：} g(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + \omega_0 \qquad 1\text{个判别函数}\end{cases}$$

两类问题 $\boldsymbol{c} = 2$

$$\begin{cases} d\text{维特征向量(样本向量)} \quad \boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_d\end{bmatrix}^T \\ \text{判别函数} \quad g(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + \omega_0 \quad \left[\boldsymbol{g}(\boldsymbol{x}) = g_1(\boldsymbol{x}) - g_2(\boldsymbol{x})\right] \\ \text{决策边界} \quad \boldsymbol{w}^T\boldsymbol{x} + \omega_0 = 0 \\ \qquad \text{法向量(权向量)}\boldsymbol{w} = \begin{bmatrix} w_1 & w_2 & \cdots & w_d\end{bmatrix}^T \\ \qquad \text{阈值权}\omega_0 \\ \text{决策域：第1类决策域、第2类决策域}\end{cases}$$

两类别分类模型--决策边界$\boldsymbol{H}$(分类超平面)的确定：

$$g(\boldsymbol{x})=0 \quad \left(g(\boldsymbol{x})=\boldsymbol{w}^T\boldsymbol{x}_1+w_0\right)$$

对于$\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{H}$并且$\boldsymbol{x}_1 \neq \boldsymbol{x}_2$ 有 $g(\boldsymbol{x}_1)=0$，$g(\boldsymbol{x}_2)=0$

即 $\boldsymbol{w}^T\boldsymbol{x}_1+w_0=\boldsymbol{w}^T\boldsymbol{x}_2+w_0$

$$\longrightarrow \quad \boldsymbol{w}^T(\boldsymbol{x}_1-\boldsymbol{x}_2)=0$$

权向量w与超平面H的任意向量正交(w垂直于H)

权向量w是超平面H的法向量

对于$\boldsymbol{x} \in \Re_1$，有$g(\boldsymbol{x})>0$,所以 **决策面H的法向量指向**$\Re_1$

$$\longrightarrow \begin{cases} \text{决策面}\boldsymbol{H}\text{的正侧：} \Re_1 \\ \text{决策面}\boldsymbol{H}\text{的负侧：} \Re_2 \end{cases}$$

两类别分类模型

决策规则

$$\text{对于观测}\boldsymbol{x}\ \begin{cases}\text{若}g(\boldsymbol{x})>0, & \text{则决策}\boldsymbol{x}\text{为}\omega_1\text{类}\\ \text{若}g(\boldsymbol{x})<0, & \text{则决策}\boldsymbol{x}\text{为}\omega_2\text{类}\\ \text{若}g(\boldsymbol{x})=0, & \text{则拒绝, 或结合实际问题分到某1类}\end{cases}$$

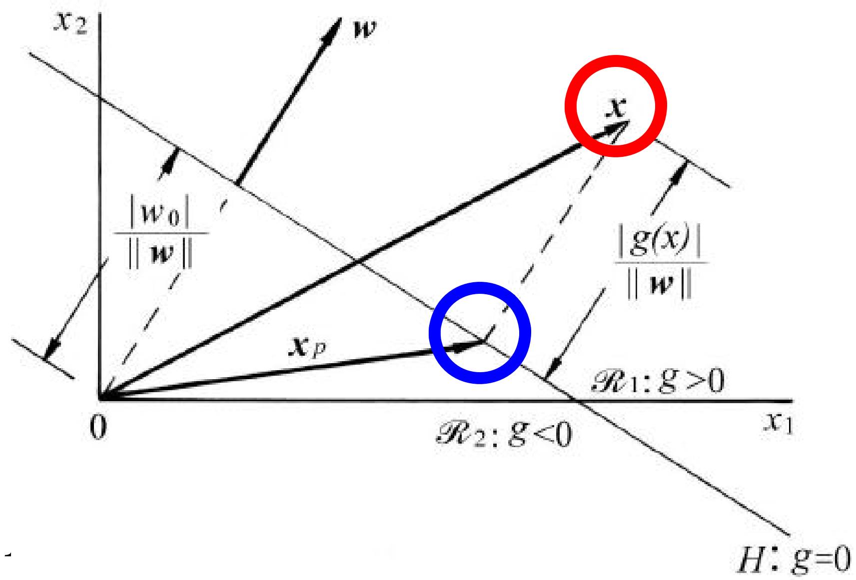
决策面（决策边界） $\boldsymbol{H}$ $\quad g(\boldsymbol{x})=0 \quad$ （超平面）

决策域 $\begin{cases}\Re_1=\{\boldsymbol{x}:\boldsymbol{x}\in\Re\text{且}g(\boldsymbol{x})>0\}\\ \Re_2=\{\boldsymbol{x}:\boldsymbol{x}\in\Re\text{且}g(\boldsymbol{x})<0\}\end{cases}$

可将判别函数$g(x)$视为观测x到决策面H距离的代数度量。

判别函数 $g(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + \omega_0$

三角形三边关系 $\boldsymbol{x} = \boldsymbol{x}_p + r\dfrac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$



其中：

$\boldsymbol{x}_p$　　$\boldsymbol{x}$在$\boldsymbol{H}$的投影向量，$\boldsymbol{x}_p \in$

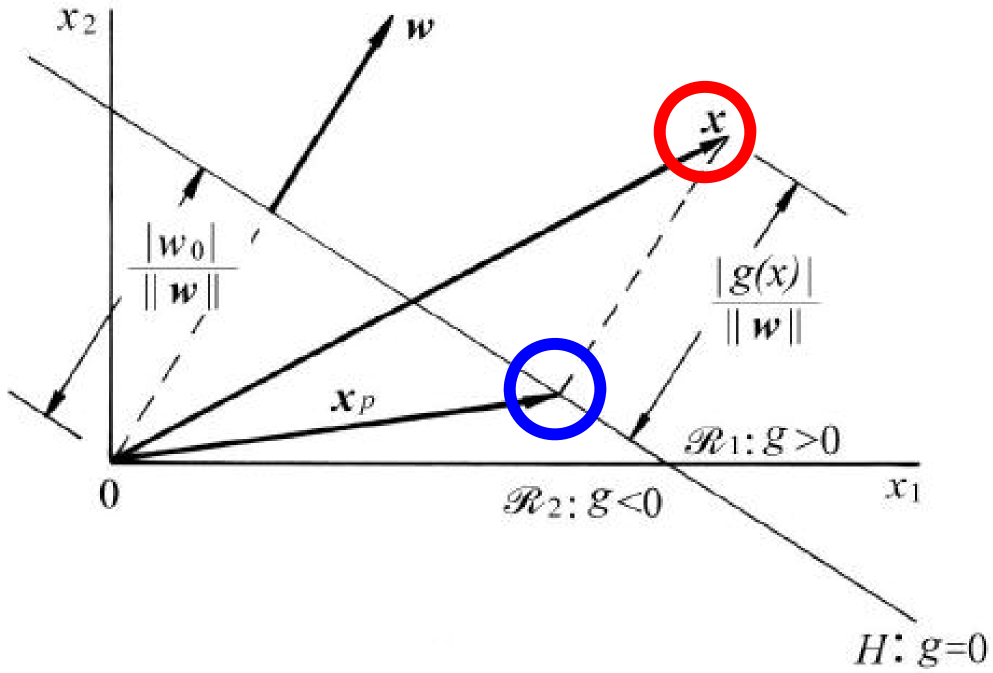
$$g(\boldsymbol{x}_p) = 0 = \boldsymbol{w}^T \boldsymbol{x}_p + \omega_0$$

r　　$\boldsymbol{x}$点到$\boldsymbol{H}$面的代数垂直距离

若$r > 0$,则$\boldsymbol{x}$在$\boldsymbol{H}$正侧。

$\dfrac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$　　$\boldsymbol{w}$方向的单位向量

$$\begin{cases} g(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + \omega_0 \\ \boldsymbol{g}(\boldsymbol{x}_p) = 0 = \boldsymbol{w}^T \boldsymbol{x}_p + \omega_0 \\ \boldsymbol{x} = \boldsymbol{x}_p + r\dfrac{\boldsymbol{w}}{\|\boldsymbol{w}\|} \end{cases}$$



$$\Rightarrow g(\boldsymbol{x}) = \boldsymbol{w}^T \overbrace{\left(\boldsymbol{x}_p + r\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\right)}^{\boldsymbol{x}} + \omega_0$$

$$= \boldsymbol{w}^T \boldsymbol{x}_p + r\boldsymbol{w}^T \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|} + \omega_0 = \boldsymbol{g}(\boldsymbol{x}_p) + r\|\boldsymbol{w}\| = r\|\boldsymbol{w}\|$$

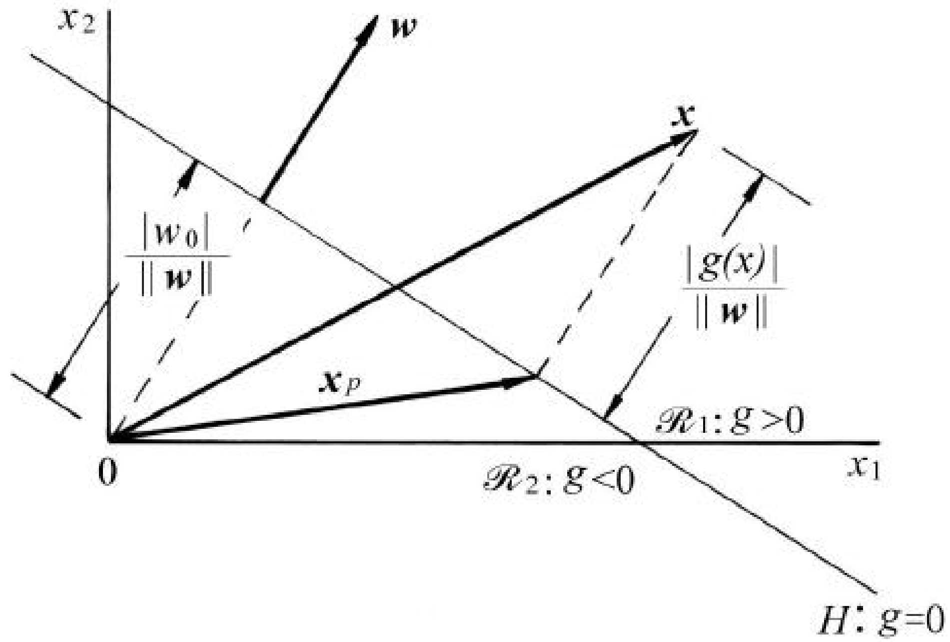
即： $g(\boldsymbol{x}) = r\|\boldsymbol{w}\|$ 或 $r = \dfrac{\boldsymbol{g}(\boldsymbol{x})}{\|\boldsymbol{w}\|}$ $\|\boldsymbol{x} - \boldsymbol{x}_p\|_2 = \dfrac{|\boldsymbol{g}(\boldsymbol{x})|}{\|\boldsymbol{w}\|}$

即：判别函数$g(x)$正比于x到超平面的代数距离r

问题：如何确定任意观测样本到分类边界的距离？

$$g(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + \omega_0 = r\|\boldsymbol{w}\|$$

$$\boldsymbol{x} = \boldsymbol{x}_p + r\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$$



对于原点 $\boldsymbol{x} = \boldsymbol{0}$，有：

$$\boldsymbol{g}(\boldsymbol{x} = \boldsymbol{0}) = \omega_0,\quad r = \frac{\omega_0}{\|\boldsymbol{w}\|}$$

$$\begin{cases} \omega_0 > 0，\text{原点}\boldsymbol{x} = \boldsymbol{0}\text{在}\boldsymbol{H}\text{正侧}; \\ \omega_0 < 0，\text{原点}\boldsymbol{x} = \boldsymbol{0}\text{在}\boldsymbol{H}\text{负侧} \\ \omega_0 = 0，\boldsymbol{H}\text{经过原点，判别函数}g(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x}\text{。} \end{cases}$$

问题：如何确定原点到分类边界的距离？

超平面（分类面，决策面）$\boldsymbol{H}$的法向量为$\boldsymbol{w}$;

超平面$\boldsymbol{H}$的偏移（超平面位置）由ω_0决定；

判别函数$\boldsymbol{g}(\boldsymbol{x})$正比于观测$\boldsymbol{x}$到超平面$\boldsymbol{H}$的代数距离；

决策域$\Re_1, \Re_2$分别位于超平面$\boldsymbol{H}$的正负侧。

设计线性分类器的关键：

基于何种准则，估计判别函数的参数$\boldsymbol{w}, \omega_0$;

B. Fisher判别分类模型

掌握：

- Fisher判别分类模型设计的基本思想
- 准则函数
- Fisher判别分类模型的实现步骤

问题：

你能否将Fisher判别分类模型的“Fisher 判别比”用于分类任务中的特征选择？

主要内容

第一部分 类别数目C=2时 FLDA分类模型

关键：确定最佳投影直线

第二部分 类别数目C>2时FLDA分类模型

许多机器学习问题涉及样本的“降维”

”降维”的目的不止一个，如：

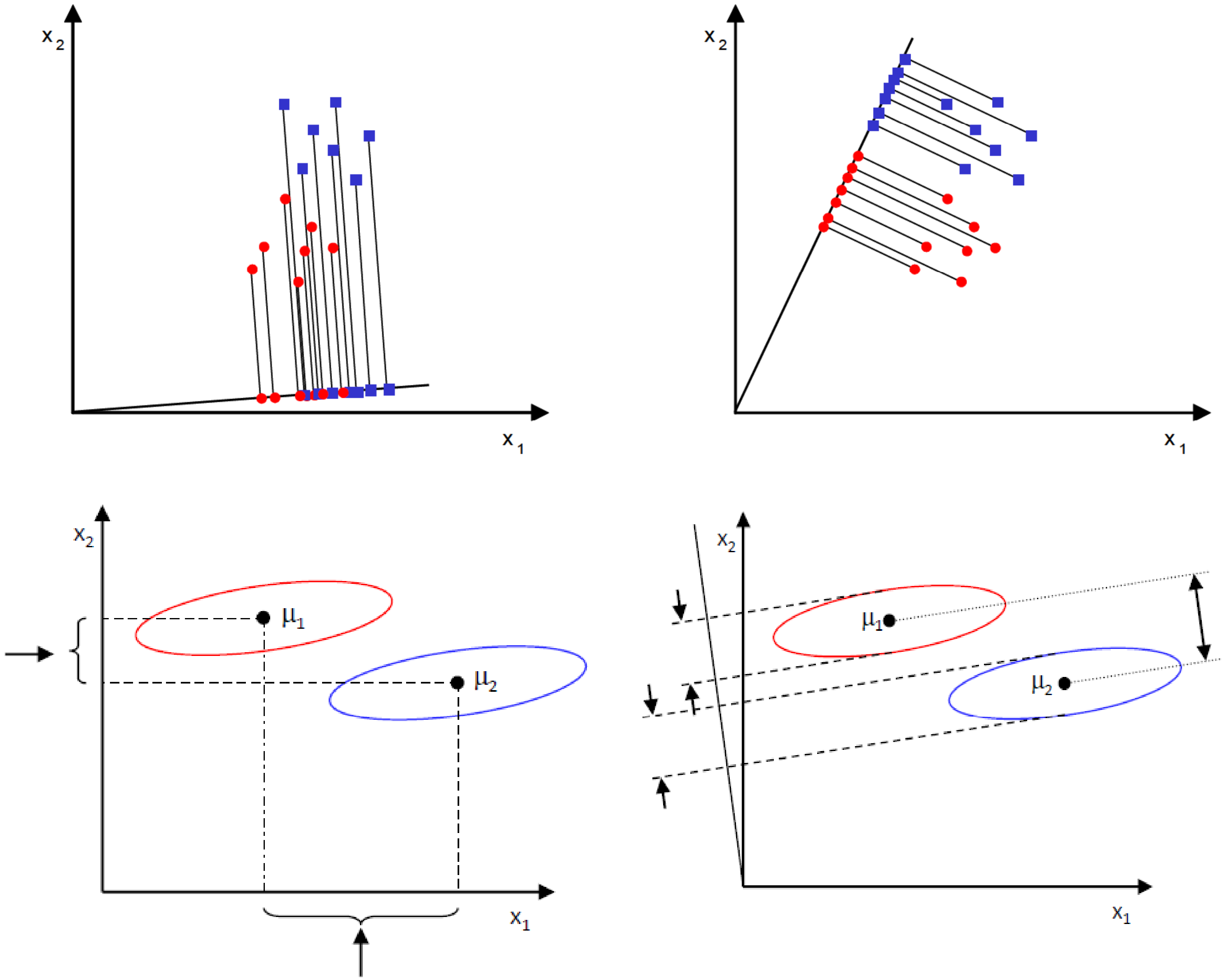
(1) *PCA*−−**数据描述**(信息压缩)

目的：寻找有效表示特征的主轴方向

(2) ***Fisher*线性判别**−−**分类器设计**

目的：寻找能有效分类的方向

什么样的投影方向更有助于分类?



1 基本思想

将d维特征空间的样本向某低维子空间做投影(两类分类对应1维空间)；在该投影空间，能最大限度区分各类数据点,以有效分类.

$$\begin{cases}\textbf{类间样本} \ \text{尽可能远离} \\ \textbf{类内样本} \ \text{尽可能聚集}\end{cases}$$

分类器的设计 $\Rightarrow$ **寻找投影直线的最佳方向** $(C = 2)$

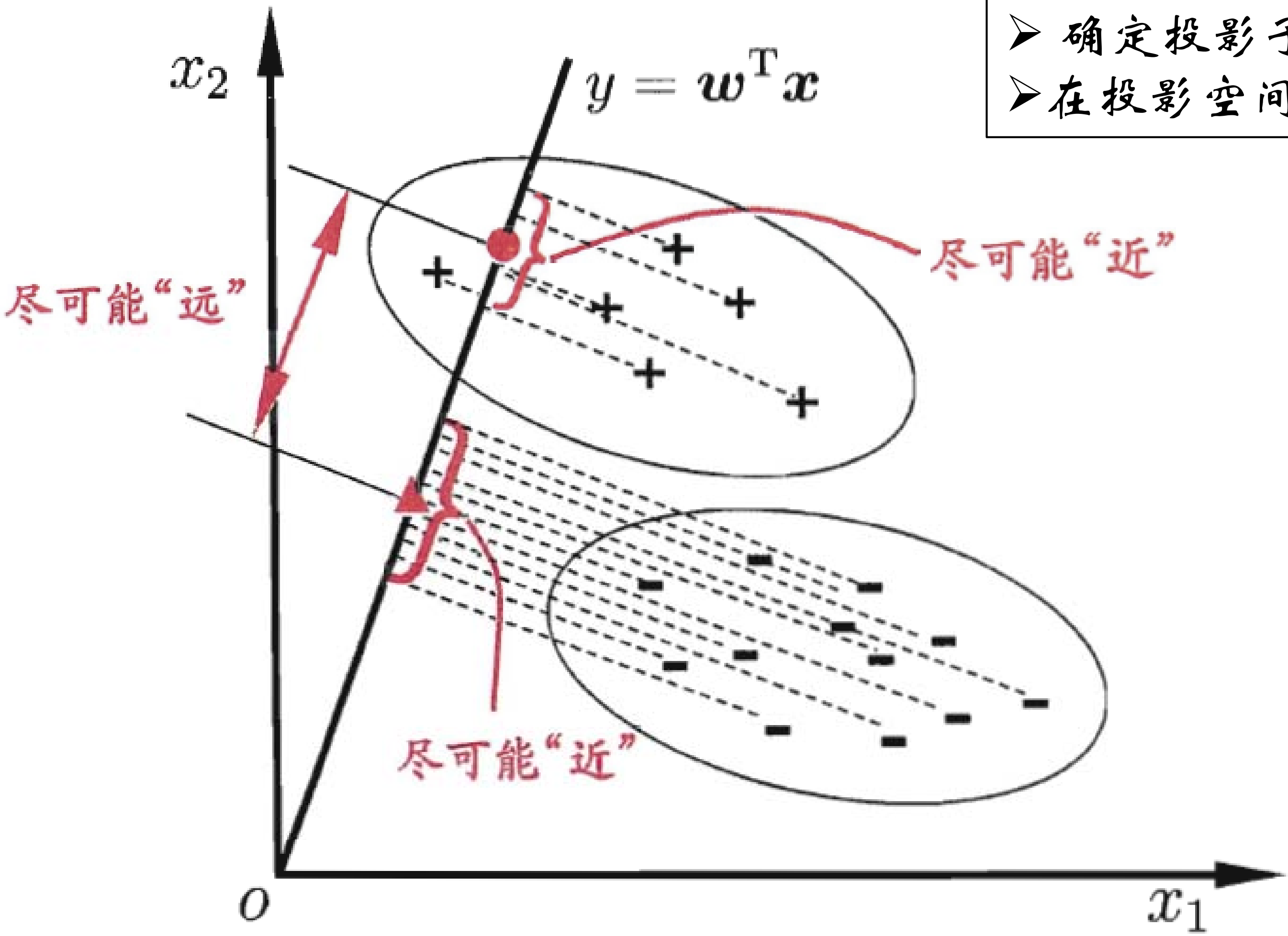
类间样本 尽可能远离

类内样本 尽可能聚集

二维特征、两类别Fisher判别分析 示意
引自周志华《机器学习》page60

任务分解：

- 确定投影子空间
- 在投影空间，设计分类模型



2 问题描述

已知：$\boldsymbol{d}$维训练样本$\mathcal{X}=\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N\}$

$$
\text{其中}\quad\begin{cases}\omega_1\text{类：}\ \mathcal{X}_1=\{\boldsymbol{x}_1^{1},\ldots,\boldsymbol{x}_{N_1}^{1}\}\\ \omega_2\text{类：}\ \mathcal{X}_2=\{\boldsymbol{x}_1^{2},\ldots,\boldsymbol{x}_{N_2}^{2}\}\end{cases}
$$

$$
\mathcal{X}=\mathcal{X}_1\cup\mathcal{X}_2\qquad \boldsymbol{N}=\boldsymbol{N}_1+\boldsymbol{N}_2
$$

$$
\xrightarrow{\text{对}\boldsymbol{x}\text{投影：}\ y=\boldsymbol{w}^T\boldsymbol{x}}\mathcal{Y}=\{y_1,\ldots,y_N\}\quad\begin{cases}\omega_1\text{类}\ \ \mathcal{Y}_1,\ \ \boldsymbol{N}_1\\ \omega_2\text{类}\ \ \mathcal{Y}_2,\ \ \boldsymbol{N}_2\end{cases}
$$

其中 $y_i=\boldsymbol{w}^T\boldsymbol{x}_i,\boldsymbol{i}=1,2,\ldots,\boldsymbol{N}$

求解：最佳投影方向$\boldsymbol{w}^*$，使两类分类效果最好。

3.准备工作(定义几个必要的参量)

[1]投影前$--d$维特征空间

各类均值向量

$$m_i = \frac{1}{N_i} \sum_{x \in \mathcal{X}_i} x \qquad i = 1, 2$$

总体均值向量(样本集的中心)

$$m = \frac{m_1 N_1 + m_2 N_2}{N_1 + N_2}$$

类内散度矩阵$(within - class \ \ scatter \ \ matrix)$

$$S_i = \frac{1}{N_i} \sum_{x \in \mathcal{X}_i} (x - m_i)(x - m_i)^T \qquad i = 1, 2$$

[1]投影前 $--d$维特征空间

总类内散度矩阵$\left(\textit{pooled within}-\textit{class scatter matrix}\right)$

$$S_w=\frac{1}{N_1+N_2}\sum_{i=1}^{2}N_iS_i$$

S_w对称半正定；若$N>d$, S_w通常非奇异

类间散度矩阵$\left(\textit{between}-\textit{class scatter}\right)$

$$\begin{cases}\text{不考虑先验概率} & S_b=(m_1-m_2)(m_1-m_2)^T\\ \text{若考虑先验概率} & S_b=P(\omega_1)P(\omega_2)(m_1-m_2)(m_1-m_2)^T\end{cases}$$

$$S_b=\frac{1}{N_1+N_2}\sum_{i=1}^{2}N_i(m_i-m)(m_i-m)^T=\sum_{i=1}^{2}P(\omega_i)(m_i-m)(m_i-m)^T$$

或者 $S_b=\frac{1}{2}\sum_{i=1}^{2}(m_i-m)(m_i-m)^T$

[2]**投影后**－1维空间

样本均值

$$\widetilde{m}_i = \frac{1}{N_i}\sum_{y\in\mathcal{Y}_i} y = \frac{1}{N_i}\sum_{x\in\mathcal{X}_i} w^T x = w^T\left(\frac{1}{N_i}\sum_{x\in\mathcal{X}_i} x\right) = w^T m_i$$

$$i = 1, 2$$

总体样本均值

$$\widetilde{m} = \frac{\widetilde{m}_1 N_1 + \widetilde{m}_2 N_2}{N_1 + N_2} = w^T m$$

类内离散度

$$\widetilde{S}_i^2 = \frac{1}{N_i}\sum_{y_j\in\mathcal{Y}_i}\left(y_j - \widetilde{m}_i\right)^2$$

$$= \frac{1}{N_i}\sum_{x_j\in\mathcal{X}_i} w^T\left(x_j - m_i\right)\left(x_j - m_i\right)^T w = w^T S_i w \qquad i = 1, 2$$

[2]投影后 −1维空间

总类内离散度

$$\tilde{S}_w = \tilde{S}_1^2 + \tilde{S}_2^2 = w^T S_1 w + w^T S_2 w = w^T \left(S_1 + S_2 \right) w = w^T S_w w$$

$$\tilde{S}_w = \frac{1}{N_1 + N_2} \sum_{i=1}^{2} N_i \tilde{S}_i^2 = \frac{1}{N_1 + N_2} \sum_{i=1}^{2} N_i w^T S_i w = w^T S_w w$$

类间离散度

$$\tilde{S}_b^2 = \left(\tilde{m}_1 - \tilde{m}_2 \right)^2 = \left(w^T m_1 - w^T m_2 \right)^2$$

$$= \left(w^T m_1 - w^T m_2 \right) \left(m_1^T w - m_2^T w \right)$$

$$= w^T \left(m_1 - m_2 \right) \left(m_1 - m_2 \right)^T w = w^T S_b w$$

$$\tilde{S}_b^2 = \frac{1}{N_1 + N_2} \sum_{i=1}^{2} N_i \left(\tilde{m}_i - \tilde{m} \right) \left(\tilde{m}_i - \tilde{m} \right)^T$$

$$= \frac{1}{N_1 + N_2} \sum_{i=1}^{2} N_i w^T \left(m_i - m \right) \left(m_i - m \right)^T w = w^T S_b w$$

4.构造*Fisher*准则函数

高维$(d$维$)$样本投影到1维空间

$$\begin{cases} \textbf{类间样本尽量分开} \Rightarrow \textbf{投影后类间离散度} \uparrow \\ \textbf{类内样本尽量聚集} \Rightarrow \textbf{投影后总类内离散度} \downarrow \end{cases}$$

$\Rightarrow$

*Fisher*准则函数 $$J_F(w) = \frac{\left|\tilde{m}_1 - \tilde{m}_2\right|^2}{\tilde{S}_w} = \frac{\left(\tilde{m}_1 - \tilde{m}_2\right)^2}{\tilde{S}_w}$$

*Fisher*准则 $$\max_{w} J_F(w)$$

$$w^* = \arg\max_{w} J_F(w)$$

如何求解？

$$Fisher\text{线性分类器设计}\begin{cases}(1)\text{最佳投影空间的估计} \\ (2)\text{基于训练样本在最佳投影空间} \\ \quad\ \text{投影设计最优边界。}\end{cases}$$

5.确定最优投影方向--求解$w^* = \arg\max_{w} J_F(w)$

[1]确定$J_F(w)$关于w的显式函数形式

$$J_F(w) = \frac{\left(\tilde{m}_1 - \tilde{m}_2\right)^2}{\tilde{S}_1^2 + \tilde{S}_2^2}$$

$$\left(\tilde{m}_1 - \tilde{m}_2\right)^2 = \left(w^T m_1 - w^T m_2\right)^2$$

$$= w^T \left(m_1 - m_2\right)\left(m_1 - m_2\right)^T w = w^T S_B w$$

$$\tilde{S}_w = \tilde{S}_1^2 + \tilde{S}_2^2 = w^T S_1 w + w^T S_2 w = w^T \left(S_1 + S_2\right) w = w^T S_w w$$

所以 $\boxed{J_F(w) = \frac{w^T S_B w}{w^T S_w w}}$ ----**广义瑞利**($Rayleigh$)**商**.

[2]估计$\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}} \boldsymbol{J}_F(\boldsymbol{w}) = \arg\max_{\boldsymbol{w}} \dfrac{\boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{S}_w \boldsymbol{w}}$

$$\Rightarrow \quad \boxed{\begin{array}{ll} \max\limits_{\boldsymbol{w}} & \boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w} \\ s.t. & \boldsymbol{w}^T \boldsymbol{S}_w \boldsymbol{w} = \boldsymbol{c} \neq 0 \end{array}}$$

采用***Lagrange***条件极值法，引入辅助函数：

$$\begin{aligned} L(\boldsymbol{w}, \lambda) &= \boldsymbol{J}_F(\boldsymbol{w}) - \lambda\left(\boldsymbol{w}^T \boldsymbol{S}_w \boldsymbol{w} - \boldsymbol{c}\right) \\ &= \boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w} - \lambda\left(\boldsymbol{w}^T \boldsymbol{S}_w \boldsymbol{w} - \boldsymbol{c}\right) \end{aligned}$$

λ为标量***Lagrange***乘子

$$L(w,\lambda)=w^T S_B w-\lambda\left(w^T S_w w-c\right)$$

$$\frac{\partial\left(x^T A x\right)}{\partial x}=\left(A+A^T\right)x \overset{A\text{是对称矩阵}}{=} 2Ax$$

$$\frac{\partial L}{\partial w}=2\left(S_B w-\lambda S_w w\right)$$

令 $\frac{\partial L}{\partial w}=0$

得 $S_B w^*=\lambda S_w w^*$ $\xrightarrow{N>d，S_w\text{非奇异}}$ $\boxed{S_w^{-1} S_B w^*=\lambda w^*}$

相应地：$J(w^*)=\frac{w^{*T} S_B w^*}{w^{*T} S_w w^*}=\frac{\lambda w^{*T} S_w w^*}{w^{*T} S_w w^*}=\lambda$

思路1 要使$J(w^*)=\lambda$最大，

应使λ取$S_w^{-1} S_B$矩阵的最大本征值，

w^*为$S_w^{-1} S_B$矩阵相应本征列向量

$[2]$**确定** $\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}} J_F(\boldsymbol{w})$

思路2：直接找到 $\boldsymbol{w}^*$

类间散度矩阵 $\boldsymbol{S}_B = (\boldsymbol{m}_1 - \boldsymbol{m}_2)(\boldsymbol{m}_1 - \boldsymbol{m}_2)^T$

$$\boldsymbol{S}_B \boldsymbol{w}^* = (\boldsymbol{m}_1 - \boldsymbol{m}_2)\underbrace{(\boldsymbol{m}_1 - \boldsymbol{m}_2)^T \boldsymbol{w}^*}_{\text{标量}R} = (\boldsymbol{m}_1 - \boldsymbol{m}_2)R$$

所以 $\boldsymbol{S}_B \boldsymbol{w}^*$ 方向与向量 $(\boldsymbol{m}_1 - \boldsymbol{m}_2)$ 方向一致。

$$\boxed{\lambda \boldsymbol{w}^* = \boldsymbol{S}_w^{-1} \boldsymbol{S}_B \boldsymbol{w}^* = \boldsymbol{S}_w^{-1} (\boldsymbol{m}_1 - \boldsymbol{m}_2) R}$$

$$\Rightarrow \quad \boxed{\boldsymbol{w}^* = \frac{R}{\lambda} \boldsymbol{S}_w^{-1} (\boldsymbol{m}_1 - \boldsymbol{m}_2)}$$

忽略标量 $\frac{R}{\lambda}$，得**最优投影方向** $\boxed{\boldsymbol{w}^* = \boldsymbol{S}_w^{-1} (\boldsymbol{m}_1 - \boldsymbol{m}_2)}$

[3] $w^* = \arg\max_{w} J_F(w)$ 计算流程

STEP1. 获取来自两类 ω_1 / ω_2 的训练样本集 $\mathcal{X}$

$$\begin{cases} \omega_1 \text{类：} \ \mathcal{X}_1, \ N_1 \text{个样本} \\ \omega_2 \text{类：} \ \mathcal{X}_2, \ N_2 \text{个样本} \end{cases}$$

STEP2. 计算各类样本均值 $\quad \boldsymbol{m}_i = \dfrac{1}{N_i} \sum_{x \in \mathcal{X}_i} \boldsymbol{x} \quad i = 1, 2$

STEP3. 类内散度矩阵

$$\boldsymbol{S}_i = \frac{1}{N_i} \sum_{x \in \mathcal{X}_i} (\boldsymbol{x} - \boldsymbol{m}_i)(\boldsymbol{x} - \boldsymbol{m}_i)^T \qquad i = 1, 2$$

STEP4. 总类内散布矩阵 $\boldsymbol{S}_w$ 及逆 $\boldsymbol{S}_w^{-1}$.

$$\boldsymbol{S}_w = \frac{1}{N_1 + N_2} \sum_{i=1}^{2} N_i \boldsymbol{S}_i$$

STEP5. 计算 $\boldsymbol{w}^*$ $\qquad \boldsymbol{w}^* = \boldsymbol{S}_w^{-1}(\boldsymbol{m}_1 - \boldsymbol{m}_2)$

利用上述过程，可直接确定投影直线 $\boldsymbol{w}^*$.

6.确定最优分类超平面

$$\left(\boldsymbol{w}^*\right)^T \boldsymbol{x} + \omega_0 = 0$$

$$明确\begin{cases} 法向量 \qquad \boldsymbol{w}^* = \arg\max_{\boldsymbol{w}} \boldsymbol{J}_F\left(\boldsymbol{w}\right) \\ \omega_0 = ? \end{cases}$$

[1] 一维最佳投影空间的分类边界
基于最小错误率贝叶斯决策

第1，两类样本正态分布，且分布形状一致
--线性分类器

第2，若维数$\boldsymbol{d}$及样本数目$\boldsymbol{N}$足够大时，$\boldsymbol{y} = \boldsymbol{w}^{*T}\boldsymbol{x}$近似正态分布，可在投影空间利用"**两步贝叶斯决策**"法。

A.确定阈值点（分类边界）y_0

先验概率相等 $$y_0 = \frac{\widetilde{m}_1 + \widetilde{m}_2}{2} = \frac{(m_1 + m_2) S_w^{-1} (m_1 - m_2)}{2}$$

先验概率不等

$$y_0 = \frac{\widetilde{m}_1 + \widetilde{m}_2}{2} + \log\frac{P(\omega_2)}{P(\omega_1)} = \frac{(m_1 + m_2) S_w^{-1} (m_1 - m_2)}{2} + \log\frac{P(\omega_2)}{P(\omega_1)}$$

投影后样本数据中心

$$y_0 = \frac{\widetilde{m}_1 N_1 + \widetilde{m}_2 N_2}{N_1 + N_2} = \frac{(m_1 N_1 + m_1 N_2) S_w^{-1} (m_1 - m_2)}{N_1 + N_2}$$

B.决策

对于任意观测$\boldsymbol{x}$

$$\begin{cases} \text{投影空间：} \quad y = \boldsymbol{w}^{*T} \boldsymbol{x} \\ \text{判决：} \text{若} y > y_0, \text{则} \boldsymbol{x} \text{为} \omega_1 \text{类；} \\ \qquad\quad \text{若} y < y_0, \text{则} \boldsymbol{x} \text{为} \omega_2 \text{类。} \end{cases}$$

分类边界满足：$y - y_0 = 0$，即 $\boldsymbol{w}^{*T} \boldsymbol{x} - y_0 = 0$

原始特征空间

判别函数：$\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{w}^{*T} \boldsymbol{x} + \omega_0 = \boldsymbol{w}^{*T} \boldsymbol{x} - y_0$

分类边界：$\boldsymbol{w}^{*T} \boldsymbol{x} - y_0 = 0$

决策规则：$\begin{cases} \text{若} \boldsymbol{g}(\boldsymbol{x}) > 0, \text{则} \boldsymbol{x} \text{为} \omega_1 \text{类} \\ \text{若} \boldsymbol{g}(\boldsymbol{x}) < 0, \text{则} \boldsymbol{x} \text{为} \omega_2 \text{类} \end{cases}$