模式识别的理论与方法

Pattern Recognition

裴继红

Chapter 3 (Part 3):

Maximum Likelihood and Bayesian Parameter Estimation

(Sections 3.6-3.8)

本讲内容：FDA

- Fisher判别分析 FDA
- Fisher多重判别分析 MDA

特征空间降维的方法

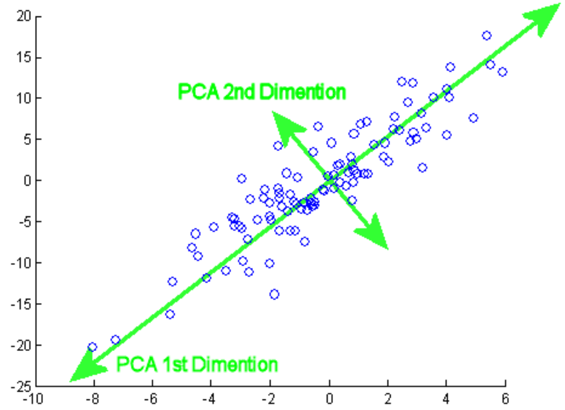
- 可以采用特征线性组合的方法减少特征空间中特征的维数。将高维数据投影到较低维空间上。主成分分析和判别分析是两类有效的线性组合变换方法

1. 主成分分析（ Principal Component Analysis，PCA ）
   - 基本思路：寻找在最小均方误差意义下最能代表数据特性的投影方向，用这些方向矢量表示数据。
2. **Fisher**判别分析（Fisher Discriminant Analysis， FDA）
   - 基本思路：在最小均方误差意义一下，寻找最能够分开各个类别数据的最佳方向。

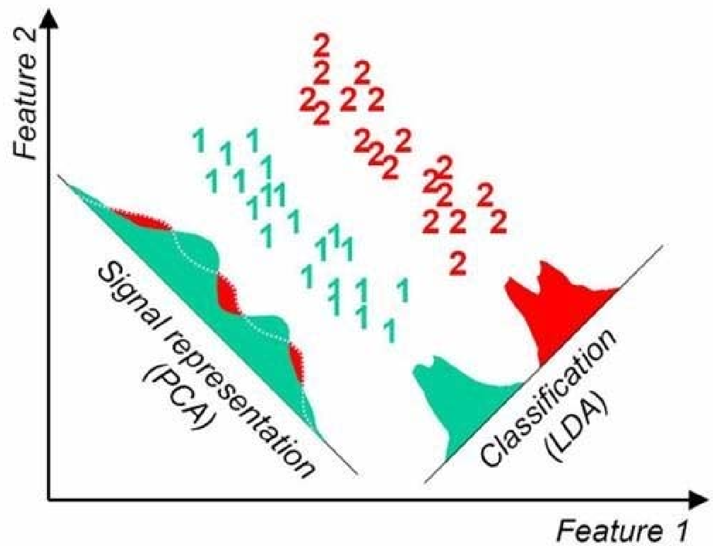
FDA

Fisher 线性判别分析
（Fisher Linear Discriminant Analysis，FDA）

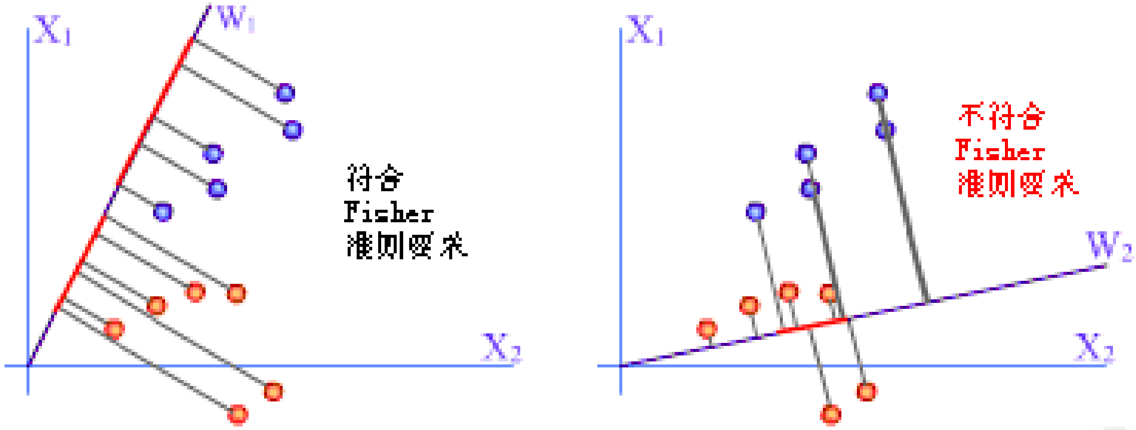
- PCA 寻找的是用来有效表示同一类样本共同特点的主轴方向，也即PCA方法对于表示同一类数据样本的共同特征是非常有效的。但PCA方法不适合用于区分不同的样本类。

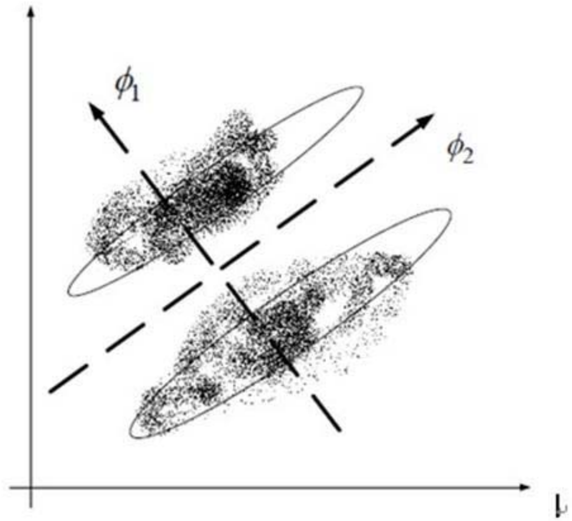


- Fisher线性判别分析（FDA）是用于寻找对不同样本类样本进行区分的最有效方向。



FDA的基本原理





- 同一组样本点在不同方向投影后可分性不同

FDA的基本原理：向直线投影

- 可分性分析的目标：考虑将d维空间中的点投影到一条直线上。通过适当地选择直线的方向，有可能找到能够最大限度地区分各类样本数据点的投影方向。

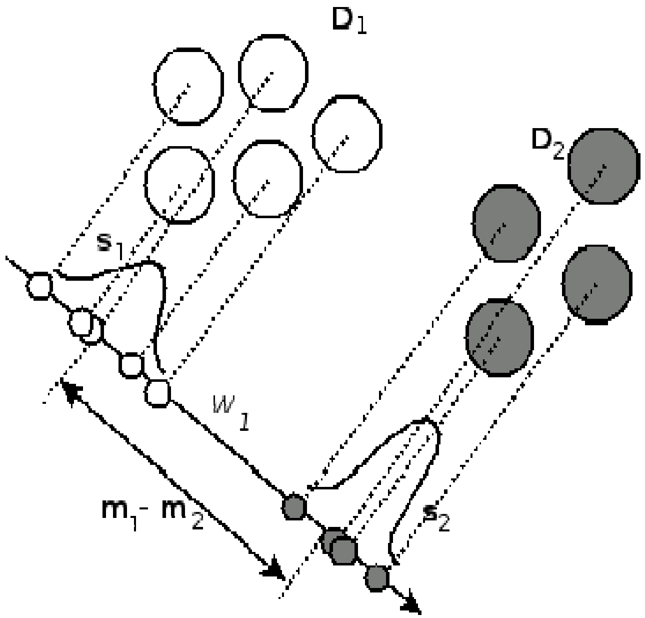
假设：一组由 n 个 d 维样本组成的数据集合 $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$，它们分属于两个不同的类 ω_1 和 ω_2，其中 ω_1 类的样本子集为 D_1，包含 n_1 个样本点；ω_2 类的样本子集为 D_2，包含 n_2 个样本点。

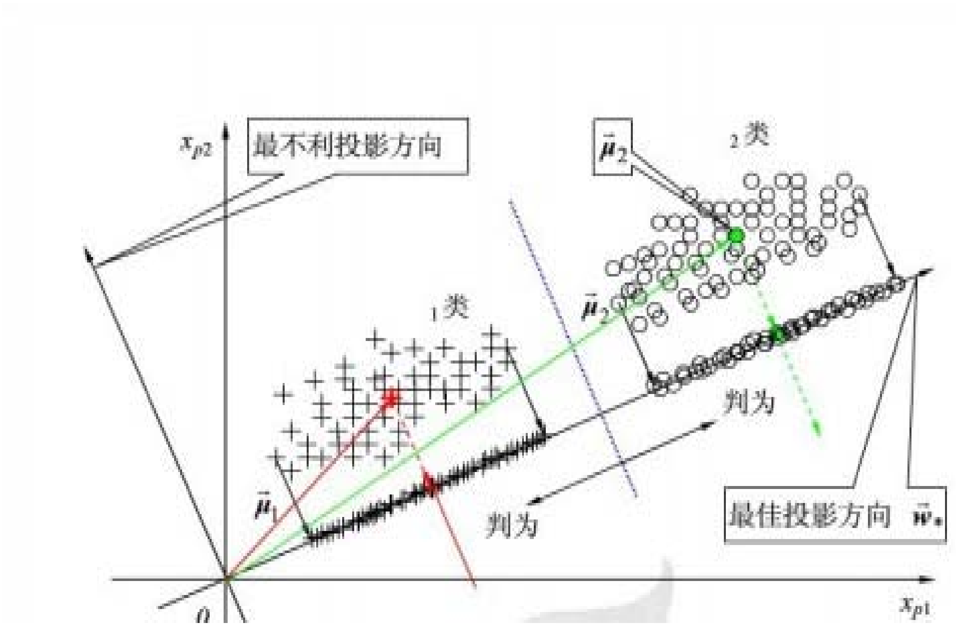
将每一个样本点投影到方向为w的直线上，其中$\|w\|=1$，

$$y = \mathbf{w}^t \mathbf{x}$$

则：得到 n 个投影标量 $y_1, y_2, \ldots, y_n$，对应于ω_1和ω_2，分别属于集合Y_1和Y_2

FDA





FDA的基本原理：向直线投影

- 假设m_i是ω_i类的样本均值，即

$$\mathbf{m}_i = \frac{1}{n_i}\sum_{x\in D_i}\mathbf{x}$$

则投影后的样本均值为：

$$\tilde{\mathbf{m}}_i = \frac{1}{n_i}\sum_{y\in Y_i} y = \frac{1}{n_i}\sum_{x\in D_i}\mathbf{w}^t\mathbf{x} = \mathbf{w}^t\mathbf{m}_i$$

投影后的样本均值之差为：

$$\left|\tilde{\mathbf{m}}_1 - \tilde{\mathbf{m}}_2\right| = \left|\mathbf{w}^t\left(\mathbf{m}_1 - \mathbf{m}_2\right)\right|$$

FDA原理：
类内散布矩阵、Fisher线性可分性准则

- 投影后，ω_i 类的类内散布为

$$\tilde{s}_i^2 = \sum_{y \in Y_i} \left(y - \tilde{\mathbf{m}}_i \right)^2$$

在投影 **y = w ^{t}x** 下，**Fisher线性可分性准则函数**为

$$J(\mathbf{w}) = \frac{\left| \tilde{\mathbf{m}}_1 - \tilde{\mathbf{m}}_2 \right|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

$\tilde{s}_1^2 + \tilde{s}_2^2$ 称为投影样本的总类内散布

FDA原理：
投影前的类内散布矩阵、类间散布矩阵

- 投影前， 类内散布矩阵为

$$\mathbf{S}_i = \sum_{x \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^t$$

总类内散布矩阵为

$$\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2$$

总类间散布矩阵为

$$\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^t$$

FDA原理：

投影后的类内散布矩阵、总类内散布矩阵

- 投影后，类内散布矩阵为

$$\tilde{s}_i^2 = \sum_{\mathbf{x}\in D_i}\left(\mathbf{w}^t\mathbf{x}-\mathbf{w}^t\mathbf{m}_i\right)^2 = \sum_{\mathbf{x}\in D_i}\mathbf{w}^t\left(\mathbf{x}-\mathbf{m}_i\right)\left(\mathbf{x}-\mathbf{m}_i\right)^t\mathbf{w} = \mathbf{w}^t\mathbf{S}_i\mathbf{w}$$

总类内散布矩阵为

$$\tilde{s}_1^2+\tilde{s}_2^2 = \mathbf{w}^t\mathbf{S}_W\mathbf{w}$$

FDA原理：
投影后样本均值之差的类间散布矩阵表示

- 投影样本均值之差可以表示为

$$\left(\tilde{\mathbf{m}}_1 - \tilde{\mathbf{m}}_2\right)^2 = \left(\mathbf{w}^t\mathbf{m}_1 - \mathbf{w}^t\mathbf{m}_2\right)^2$$

$$= \mathbf{w}^t\left(\mathbf{m}_1 - \mathbf{m}_2\right)\left(\mathbf{m}_1 - \mathbf{m}_2\right)^t\mathbf{w} = \mathbf{w}^t\mathbf{S}_B\mathbf{w}$$

即 $$\left(\tilde{\mathbf{m}}_1 - \tilde{\mathbf{m}}_2\right)^2 = \mathbf{w}^t\mathbf{S}_B\mathbf{w}$$

其中 $$\mathbf{S}_B = \left(\mathbf{m}_1 - \mathbf{m}_2\right)\left(\mathbf{m}_1 - \mathbf{m}_2\right)^t$$

FDA原理：

Fisher判别准则函数——广义瑞利商

- 用类内总散布矩阵和类间总散布矩阵表示的Fisher判别准则函数为

$$J(\mathbf{w}) = \frac{\mathbf{w}^t \mathbf{S}_B \mathbf{w}}{\mathbf{w}^t \mathbf{S}_W \mathbf{w}}$$

该准则函数也称为广义瑞利商。

Fisher判别分析的目的就是要找到使得该准则函数 $J(.)$ 最大化时的直线w

FDA原理：基于Fisher判别准则函数优化的最大类可分性投影方向W的求解

- 使 $J(.)$ 最大化时的直线w 必须满足方程：

$$\mathbf{S}_B\mathbf{w} = \lambda\,\mathbf{S}_W\mathbf{w}$$

若 S_W 非奇异，则 w 是$S_W^{-1}S_B$ 的特征矢量，即满足

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

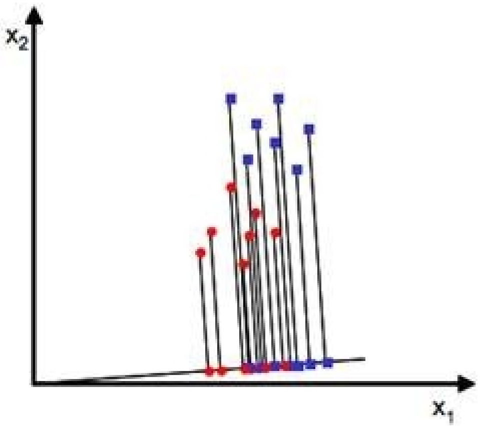
由于

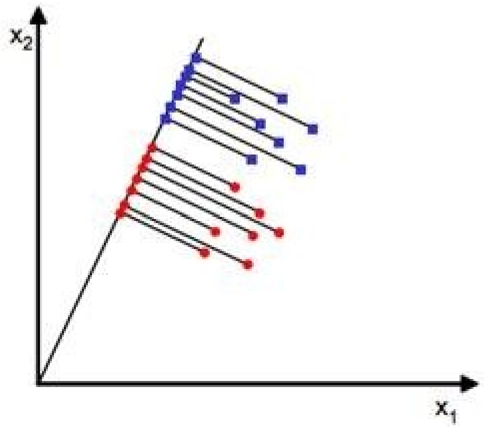
$$\mathbf{S}_B\mathbf{w} = (\mathbf{m}_1-\mathbf{m}_2)(\mathbf{m}_1-\mathbf{m}_2)^t\,\mathbf{w} = (\mathbf{m}_1-\mathbf{m}_2)\left[(\mathbf{m}_1-\mathbf{m}_2)^t\,\mathbf{w}\right] = a\cdot(\mathbf{m}_1-\mathbf{m}_2)$$

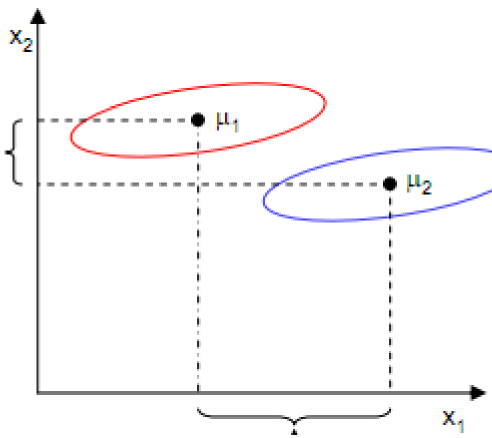
即 S_Bw 的方向与矢量 m_1-m_2 的方向相同， 而我们主要关心的是w 的方向， w 的模对问题的求解无关紧要，因此可得

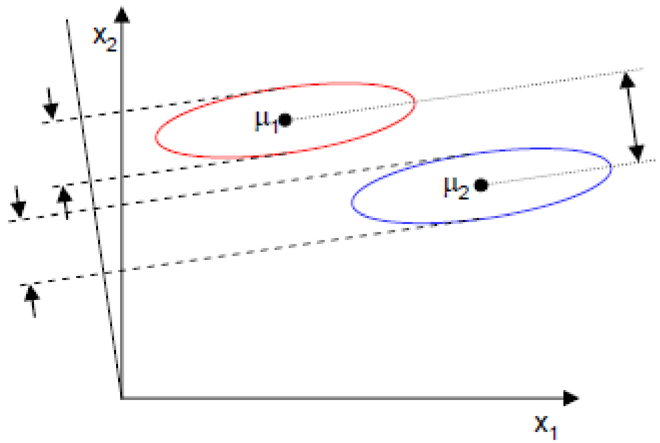
使 $J(.)$ 最大化时的w为

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


x2
x1


x2
x1

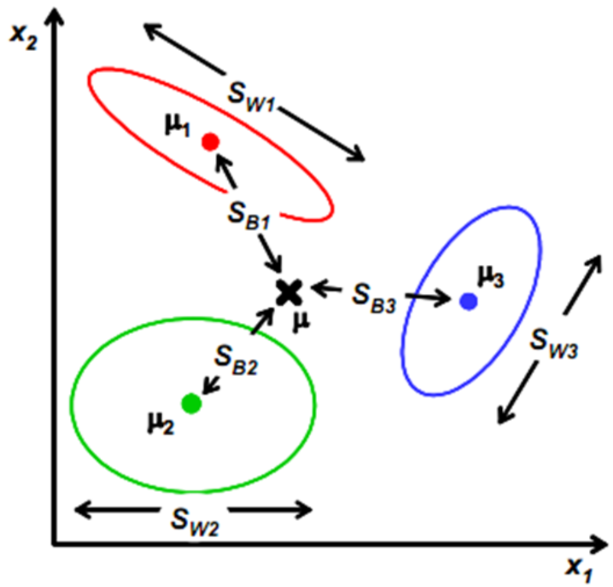

x2
μ1
μ2
x1

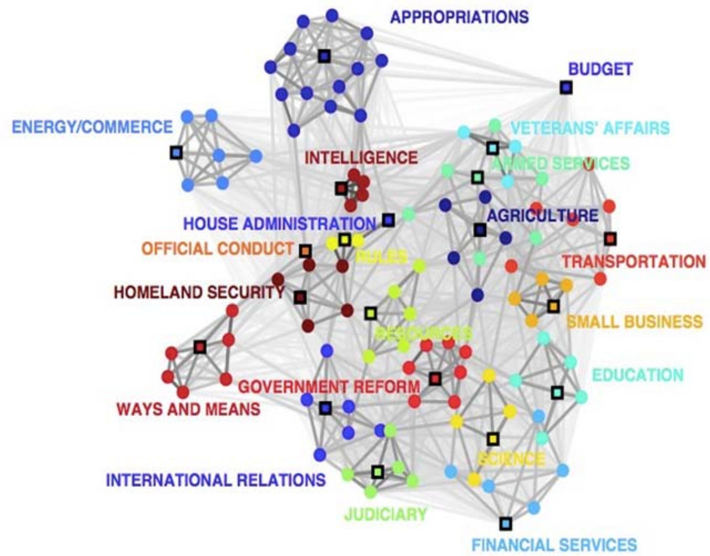

x2
μ1
μ2
x1

MDA

多重判别分析（MDA）

- 对于c>2类的 分类问题，为推广Fisher线性判别准则，需要c-1个判别函数。
- 此时需要从d维空间向c-1维空间进行投影。





MDA：类内散布矩阵

➢ 假设d≥c，此时类内散布矩阵为

$$\mathbf{S}_W = \sum_{i=1}^{c} \mathbf{S}_i$$

其中：

$$\mathbf{S}_i = \sum_{x \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^t$$

而

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{x \in D_i} \mathbf{x}$$

MDA：
总体均值向量、类内散布矩阵、总体散布矩阵

总体均值向量：
$$\mathbf{m} = \frac{1}{n}\sum_{\mathbf{x}} \mathbf{x} = \frac{1}{n}\sum_{i=1}^{c} n_i \mathbf{m}_i$$

类间散布矩阵：
$$\mathbf{S}_B = \sum_{i=1}^{c} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^t$$

总体散布矩阵：
$$\mathbf{S}_T = \sum_{\mathbf{x}} (\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^t$$
$$= \mathbf{S}_W + \sum_{i=1}^{c} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^t$$

$$\mathbf{S}_T = \mathbf{S}_W + \mathbf{S}_B$$

MDA： d维空间向c-1维空间投影

- 从d维空间向c-1维空间进行投影，通过 c-1个分类过程进行

$$y_i = \mathbf{w}_i^t \mathbf{x}$$ 其中 i =1,2,…c-1

或写成矩阵形式 $\mathbf{y} = \mathbf{W}^t \mathbf{x}$

其中，W为d×(c-1)的矩阵，矩阵的每一行是d维的矢量w_i

MDA：
向c-1维投影后的均值向量、散布矩阵

- 从d维空间向c-1维空间进行投影后的均值向量

$$\tilde{\mathbf{m}}_i = \frac{1}{n_i}\sum_{\mathbf{y}\in Y_i}\mathbf{y} \qquad \tilde{\mathbf{m}} = \frac{1}{n}\sum_{i=1}^{c} n_i\tilde{\mathbf{m}}_i$$

散布矩阵

$$\tilde{\mathbf{S}}_W = \sum_{i=1}^{c}\sum_{y\in Y_i}(\mathbf{y}-\tilde{\mathbf{m}}_i)(\mathbf{y}-\tilde{\mathbf{m}}_i)^t$$

$$\tilde{\mathbf{S}}_B = \sum_{i=1}^{c} n_i(\tilde{\mathbf{m}}_i-\tilde{\mathbf{m}})(\tilde{\mathbf{m}}_i-\tilde{\mathbf{m}})^t$$

MDA：投影前、后散布矩阵之间的关系

- 易证明投影前后散布矩阵之间的关系

$$\tilde{\mathbf{S}}_W = \mathbf{W}^t \mathbf{S}_W \mathbf{W}$$

$$\tilde{\mathbf{S}}_B = \mathbf{W}^t \mathbf{S}_B \mathbf{W}$$

MDA：准则函数

- MDA准则函数

$$J(\mathbf{w}) = \frac{|\tilde{\mathbf{S}}_B|}{|\tilde{\mathbf{S}}_W|} = \frac{|\mathbf{w}^t \mathbf{S}_B \mathbf{w}|}{|\mathbf{w}^t \mathbf{S}_W \mathbf{w}|}$$

最优解的特征矢量 $\mathbf{w}_i$ 满足的条件

$$\mathbf{S}_B \mathbf{w}_i = \lambda_i \mathbf{S}_W \mathbf{w}_i$$

因为，S_B的秩 $\leq$ c-1，

所以，上述公式中非零特征值的个数最多只有 c-1 个，

而最优解的特征矢量 $\mathbf{w}_i$ 满足这些非零特征值对应的特征向量

MDA：求解最优解的特征矢量 w_i

$$\mathbf{S}_B \mathbf{w}_i = \lambda_i \mathbf{S}_W \mathbf{w}_i$$

① 若$\mathrm{S_w}$ 可逆，最优解的特征矢量 w_i 满足

$$\mathbf{S}_W^{-1} \mathbf{S}_B \mathbf{w}_i = \lambda_i \mathbf{w}_i$$

则 $\mathrm{S^{-1}{}_w\,S_B}$的c-1个非零特征值对应的特征向量w_i就是最优解的特征矢量

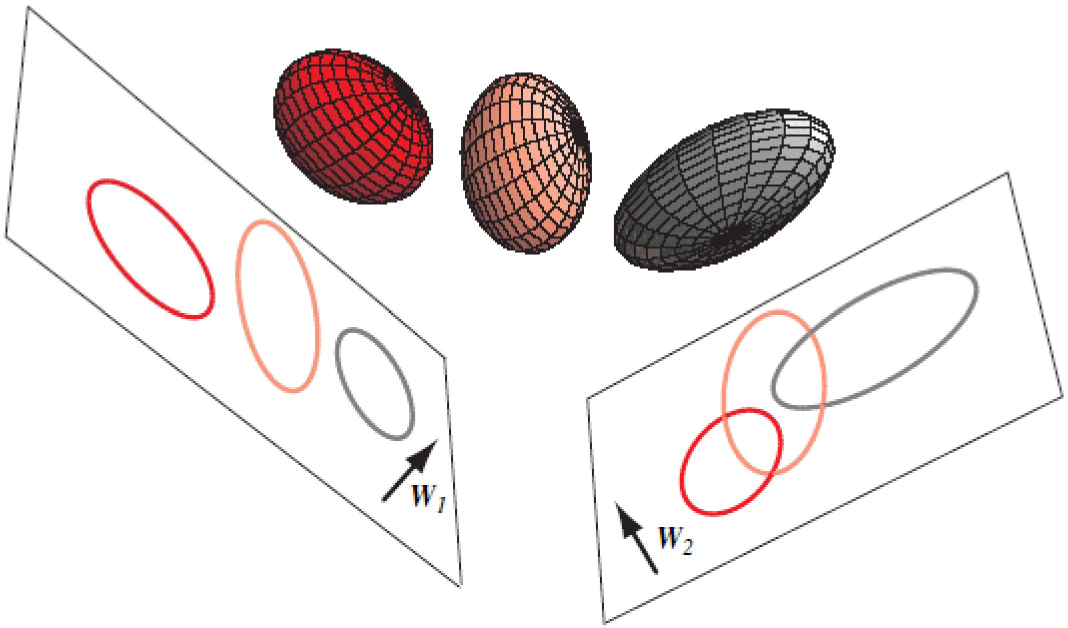
② 若类内散布矩阵各向同性，此时c-1个特征矢量 w_i 可以由c-1个向量：

$$\mathbf{m}_i\text{-}\mathbf{m}，\quad i = 1, 2, \ldots c-1$$

通过Gram-Schmidt正交化方法得到。

MDA投影判别示意：Figure3-6

- 三个三维分布的样本被分别投影到用法向量W_1、W_2所表示的二维平面上，在不同的投影面上，样本投影结果有不同的分离度



几点说明

- 若实际可以得到的训练样本非常少，就需要向更低维的空间作投影；而若样本数量较多，则适合向较高维的空间进行投影。
- 在两类情况下，多重判别分析提供了一种合理的降维方法。
- 除了 PCA 和 MDA 可以进行降维，还存在一些其他的线性和非线性降维方法，如局部线性嵌入方法等。