深圳大学研究生课程：模式识别理论与方法

课程作业实验报告

实验名称：Fisher 线性判别分析 FDA

实验编号：Proj03-02

签　　名：

姓　　名：夏荣杰

学　　号：2170269107

截止提交日期：2018 年 4 月 27 日

摘要：Fisher 线性判别分析（FDA）是一种将高维数据投影到低维数据的线性组合变换方法，这一方法是在最小均方误差意义下，寻找最能够分开各个类别数据的最佳方向。作为 FDA 的高维推广，多重判别分析（MDA）是用于寻找对不同类样本进行区分的最有效子空间。本次实验分为两部分，第一部分是 FDA 分析：将两类高维分布的数据降维为一维数据，并使用贝叶斯分类器对其进行分类；第二部分是 MDA 分析：将三类高维分布的数据降维为二维数据，并使用贝叶斯分类器对其进行分类。实验结果表明，采用特征线性组合的方法可以减少特征空间中特征的维数。FDA 和 MDA 通过适当地选择投影方向，能够找到最大限度地区分各类样本数据点的投影方向，从而在低维空间中将数据进行分类。实验过程中，我还比较分析了 FDA 寻找的最佳投影方向和非最优方向的差异，以及 MDA 寻找的最优投影子空间和非最优子空间的差异。由于训练样本较少，某种程度下，数据在最优子空间的分类准确率比在非最优子空间的分类准确率低。但只要训练样本的数量足够大，最优子空间的分类效果是要比在非最优子空间下的分类效果更优的。

一、 背景技术 或 基本原理

1. FDA 分析

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 的基本原理是考虑将 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。

假设有一组 n 个 d 维的样本 $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$，这些样本属于两个类别 w_1 和 w_2，其中 w_1 类的样本子集为 $\mathbf{D}_1$，包含 n_1 个样本点；w_2 类的样本子集为 $\mathbf{D}_2$，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，

$$y = \mathbf{w}^t \mathbf{x} \tag{1-1}$$

则：得到 n 个投影标量 $y_1, y_2, \cdots, y_n$，对应于 $\mathbf{w}_1$ 和 $\mathbf{w}_2$ ，分别属于集合 $\mathbf{Y}_1$ 和 $\mathbf{Y}_2$。

假设 $\mathbf{m}_i$ 是 w_i 类的样本均值，即

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{x \in D_i} \mathbf{x} \tag{1-2}$$

则投影后的样本均值为：

$$\tilde{m}_i = \frac{1}{n_i} \sum_{y \in Y_i} y = \frac{1}{n_i} \sum_{x \in D_i} \mathbf{w}^t \mathbf{x} = \mathbf{w}^t \mathbf{m}_i \tag{1-3}$$

投影后，w_i 类的类内散布为：

$$\tilde{s}_i^2 = \sum_{y \in Y_i} \left(y - \tilde{\mathbf{m}}_i \right)^2 \tag{1-4}$$

在投影 $y = \mathbf{w}^t \mathbf{x}$ 下，Fisher 线性可分性准则函数为

$$J\left(\mathbf{w}\right) = \frac{\left|\tilde{\mathbf{m}}_1 - \tilde{\mathbf{m}}_2\right|^2}{\tilde{s}_1^2 + \tilde{s}_2^2} \tag{1-5}$$

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

- 首先，计算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

$$\mathbf{S}_i = \sum_{x \in D_i} \left(\mathbf{x} - \mathbf{m}_i \right) \left(\mathbf{x} - \mathbf{m}_i \right)^t \tag{1-6}$$

总类内散布矩阵为

$$\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2 \tag{1-7}$$

总类间散布矩阵为

$$\mathbf{S}_B=\left(\mathbf{m}_1-\mathbf{m}_2\right)\left(\mathbf{m}_1-\mathbf{m}_2\right)^t \tag{1-8}$$

➢ 接着，计算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

$$\mathbf{S}=\sum_{x\in D_i}\left(\mathbf{w}^t\mathbf{x}-\mathbf{w}^t\mathbf{m}_i\right)^2=\sum_{x\in D_i}\mathbf{w}^t\left(\mathbf{x}-\mathbf{m}_i\right)\left(\mathbf{x}-\mathbf{m}_i\right)^t\mathbf{w}=\mathbf{w}^t\mathbf{S}_i\mathbf{w} \tag{1-9}$$

总类内散布矩阵为

$$\tilde{s}_1^2+\tilde{s}_2^2=\mathbf{w}^t\mathbf{S}_W\mathbf{w} \tag{1-10}$$

所以，式（5）中投影样本均值之差可以表示为

$$\begin{aligned}\left(\tilde{\mathbf{m}}_1-\tilde{\mathbf{m}}_2\right)^2&=\left(\mathbf{w}^t\mathbf{m}_1-\mathbf{w}^t\mathbf{m}_2\right)^2\\&=\mathbf{w}^t\left(\mathbf{m}_1-\mathbf{m}_2\right)\left(\mathbf{m}_1-\mathbf{m}_2\right)^t\mathbf{w}\\&=\mathbf{w}^t\mathbf{S}_B\mathbf{w}\end{aligned} \tag{1-11}$$

其中，$\mathbf{S}_B=\left(\mathbf{m}_1-\mathbf{m}_2\right)\left(\mathbf{m}_1-\mathbf{m}_2\right)^t$，称为总类间散布矩阵。

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

$$J\left(\mathbf{w}\right)=\frac{\mathbf{w}^t\mathbf{S}_B\mathbf{w}}{\mathbf{w}^t\mathbf{S}_W\mathbf{w}} \tag{1-12}$$

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**。

使 J(.) 最大化时的直线 **w** 必须满足方程：

$$\mathbf{S}_B\mathbf{w}=\lambda\mathbf{S}_W\mathbf{w} \tag{1-13}$$

若 $\mathbf{S}_W$ 非奇异，使 J(.) 最大化时的 **w** 为

$$\mathbf{w}=\mathbf{S}_W^{-1}\left(\mathbf{m}_1-\mathbf{m}_2\right) \tag{1-14}$$

2. 多重判别分析（MDA 分析）

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。

➢ 首先，计算总体均值向量、类内散布矩阵、类间散布矩阵和总体散布矩阵。

假设 d≥c，此时总类内散布矩阵为：

$$\mathbf{S}_W=\sum_{i=1}^{c}\mathbf{S}_i \tag{2-1}$$

其中，

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

$$\mathbf{m}_i=\frac{1}{n_i}\sum_{x\in D_i}\mathbf{x} \tag{2-3}$$

总体均值向量：

$$\mathbf{m}=\frac{1}{n_i}\sum_{\mathbf{x}}\mathbf{x}=\frac{1}{n}\sum_{i=1}^{c}n_i\mathbf{m}_i \tag{2-4}$$

类间散布矩阵：

$$\mathbf{S}_B=\sum_{i=1}^{c}n_i\left(\mathbf{m}_i-\mathbf{m}\right)\left(\mathbf{m}_i-\mathbf{m}\right)^t \tag{2-5}$$

总体散布矩阵：

$$\begin{aligned}\mathbf{S}_T&=\sum_{\mathbf{x}}\left(\mathbf{x}-\mathbf{m}\right)\left(\mathbf{x}-\mathbf{m}\right)^t\\&=\mathbf{S}_W+\sum_{i=1}^{c}n_i\left(\mathbf{m}_i-\mathbf{m}\right)\left(\mathbf{m}_i-\mathbf{m}\right)^t\\&=\mathbf{S}_W+\mathbf{S}_B\end{aligned} \tag{2-6}$$

- 接着，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

$$\mathbf{y}=\mathbf{W}^t\mathbf{x} \tag{2-7}$$

其中，**W** 为 d×(c-1)的矩阵，矩阵的每一行是 d 维的矢量 $\mathbf{w}_i$。

- 然后，计算向 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

$$\tilde{\mathbf{m}}_i=\frac{1}{n_i}\sum_{y\in Y_i}\mathbf{y}\text{，}\quad\tilde{\mathbf{m}}=\frac{1}{n}\sum_{i=1}^{c}n_i\tilde{\mathbf{m}}_i \tag{2-8}$$

散布矩阵为：

$$\tilde{\mathbf{S}}_W=\sum_{i=1}^{c}\sum_{y\in Y_i}\left(y-\tilde{\mathbf{m}}_i\right)\left(y-\tilde{\mathbf{m}}_i\right)^t=\mathbf{W}^t\mathbf{S}_W\mathbf{W} \tag{2-9}$$

$$\tilde{\mathbf{S}}_B=\sum_{i=1}^{c}n_i\left(\tilde{\mathbf{m}}_i-\tilde{\mathbf{m}}\right)\left(\tilde{\mathbf{m}}_i-\tilde{\mathbf{m}}\right)^t=\mathbf{W}^t\mathbf{S}_B\mathbf{W} \tag{2-10}$$

则，MDA 准则函数为：

$$J\left(\mathbf{w}\right)=\frac{\left|\tilde{\mathbf{S}}_B\right|}{\left|\tilde{\mathbf{S}}_W\right|}=\frac{\left|\mathbf{w}^t\mathbf{S}_B\mathbf{w}\right|}{\left|\mathbf{w}^t\mathbf{S}_W\mathbf{w}\right|} \tag{2-11}$$

最优解的特征矢量 $\mathbf{w}_i$ 满足的条件

$$\mathbf{S}_B\mathbf{w}_i=\lambda\mathbf{S}_W\mathbf{w}_i \tag{2-12}$$

若 $\mathbf{S}_W$ 非奇异，使 J(.) 最大化时的 $\mathbf{w}_i$ 为

$$\mathbf{S}_W^{-1}\mathbf{S}_B\mathbf{w}_i=\lambda_i\mathbf{w}_i \tag{2-13}$$

则 $\mathbf{S}_W^{-1}\mathbf{S}_B$ 的 c-1 个非零特征值对应的特征向量 $\mathbf{w}_i$ 就是最优解的特征矢量。

二、 实验方法 或 算法流程步骤

实验用到的方法和步骤如下：

1. 基本实验：FDA 分析

① 根据式（1-2）和式（1-6），分别计算类别 w_2 和 w_3 数据的均值向量 $\mathbf{m}_i$ 和类内散布矩阵 $\mathbf{S}_i$；

② 根据式（1-7）计算总类散布矩阵；

③ 根据式（1-14），对表格 2 中的类别 w_2 和 w_3，计算最优方向矢量 **w**；

④ 根据式（1-1）计算类别 w_2 和 w_3 的所有数据点在矢量方向 **w** 上的投影；

⑤ 在一幅图中用不同的颜色分别画出类别 w_2 和 w_3 数据的三维散点图，同时画出表示方向矢量 **w** 的直线，并且标记出投影后的点在直线上的位置；

⑥ 对投影后得到的两个一维数据集合，假设它们都满足一维高斯分布，分别计算出它们的均值和方差。并假设它们的先验概率 $P(w_2)=P(w_3)=0.5$，设计一个方向 **w** 上的一维贝叶斯分类器；

⑦ 用设计得到的贝叶斯分类器对 w_2 和 w_3 的所有数据点进行分类，并计算分类器的训练误差，即错分点的个数；

⑧ 假设 $\mathbf{v} =(1.0,2.0,-1.5)^T$，令非最优方向矢量 $\mathbf{w} = \mathbf{v}/\|\mathbf{v}\|$，重复上面步骤，计算在这个非最优子空间中，分类器的训练误差。

2. 拓展实验：MDA 分析

① 根据式（2-1）到（2-3），分别计算类别 w_1，w_2 和 w_3 数据的均值向量 $\mathbf{m}_i$，类内散布矩阵 $\mathbf{S}_i$ 和总类内散布矩阵 $\mathbf{S}_W$；

② 根据式（2-4）计算总体均值向量 **m**;

③ 根据式（2-5）计算样本的类间散布矩阵 $\mathbf{S}_B$；

④ 根据式（2-13），对表格 2 中的类别 w_1，w_2 和 w_3，计算最优方向矢量 $\mathbf{W}=[\mathbf{w}_1，\mathbf{w}_2]$，其中两个方向基矢量 $\mathbf{w}_1$、$\mathbf{w}_2$ 分别是公式（2-13）中矩阵 $\mathbf{S}_W^{-1}\mathbf{S}_B$ 的最大特征值和次大特征值对应的特征矢量，方向基矢量 $\mathbf{w}_1$、$\mathbf{w}_2$ 均为单位矢量；

⑤ 根据式（2-7）计算类别 w_1，w_2 和 w_3 的所有数据点在矢量方向 **W** 上的投影；

⑥ 在一幅图中用不同的颜色画出表 2 中三个类别数据的三维散点图；

⑦ 在一幅图中用不同颜色画出⑤中计算出的三个类别数据的二维散点图；

⑧ 对投影后得到的三类二维数据集合，假设它们都满足高斯分布，分别计算出它们的均值向量和协方差矩阵。并假设它们的先验概率 $P(w_1)=P(w_2)=P(w_3)=1/3$，设计子空间 **W** 上的贝叶斯分类器；

⑨ 用设计得到的贝叶斯分类器对 w_1，w_2 和 w_3 的所有数据点进行分类，并计算分类器的训练误差，即错分点的个数；

⑩ 假设 $v =(1.0,2.0,-1.5)^T$，$v2 =(-1.0,0.5,-1.0)^T$，令非最优方向矢量 $\mathbf{w}_i = \mathbf{v}_i/\|\mathbf{v}_i\|$，重复上面步骤，计算在这个非最优子空间 $\mathbf{W}=[\mathbf{w}_1，\mathbf{w}_2]$中，设计的贝叶斯分类器的训练误差。

三、 实验结果

1. 基本实验：FDA 分析

对表格 2 中的类别 w_2 和 w_3，计算得到最优方向矢量 **w** 为：

$$\mathbf{w} = [-0.8603, 0.4798, -0.1723]$$

在一幅图中用不同的颜色分别画出类别 w_2 和 w_3 数据的三维散点图，同时画出表示方向

矢量 **w** 的直线，并且标记出投影后的点在直线上的位置，如图 1-1 所示：

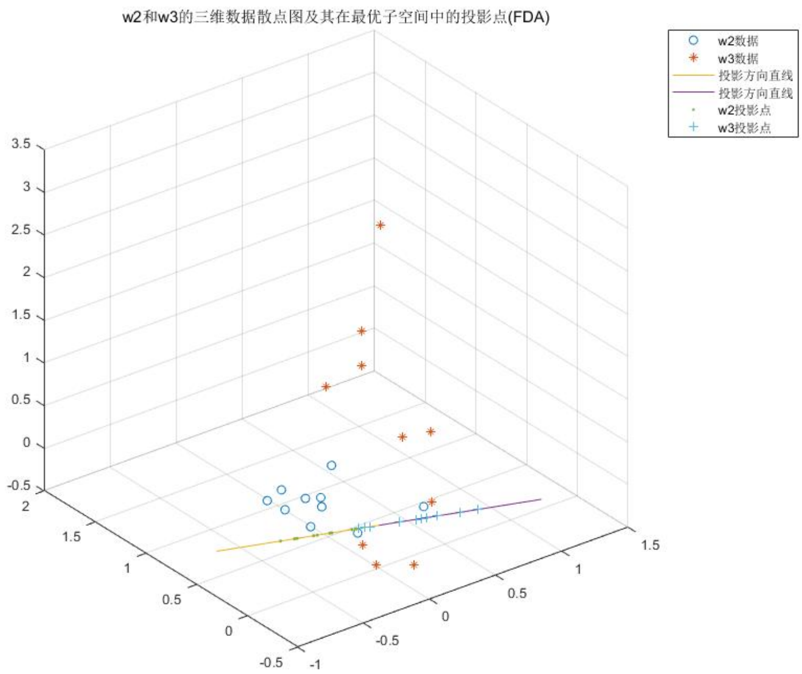


图 1-1. w2 和 w3 的三维数据散点图及其在最优子空间中的投影点（FDA）

对投影后得到的两个一维数据集合，假设它们都满足一维高斯分布，分别计算出它们的均值和方差如下：

	均值	方差
w_2 投影后的一维数据	0.3025	0.0665
w_3 投影后的一维数据	-0.2092	0.0611

假设它们的先验概率 $P(w_2) = P(w_3) = 0.5$，设计一个方向 **w** 上的一维贝叶斯分类器；用设计得到的贝叶斯分类器对 w_2 和 w_3 的所有数据点进行分类，并计算分类器的训练误差，即错分点的个数，结果如下：

在最优子空间中(FDA)，
使用贝叶斯分类器对w2和w3的所有数据点进行分类
------------w2样本点：第2类------------
------------w2样本点：第2类------------
------------w2样本点：第3类------------
------------w2样本点：第2类------------
------------w2样本点：第2类------------
------------w2样本点：第2类------------
------------w2样本点：第2类------------
------------w2样本点：第2类------------
------------w2样本点：第2类------------
------------w2样本点：第2类------------
w2数据的训练误差：错分点为1个

------------w3样本点：第2类------------
------------w3样本点：第3类------------
------------w3样本点：第2类------------
------------w3样本点：第3类------------
------------w3样本点：第3类------------
------------w3样本点：第2类------------
------------w3样本点：第3类------------
------------w3样本点：第3类------------
------------w3样本点：第3类------------
------------w3样本点：第3类------------
w3数据的训练误差：错分点为3个

对比实验：假设 $\mathbf{v}=(1.0,2.0,-1.5)^T$，令非最优方向矢量 $\mathbf{w}=\mathbf{v}/\|\mathbf{v}\|$，重复上面步骤，在一幅图中用不同的颜色分别画出类别 w_2 和 w_3 数据的三维散点图，同时画出表示方向矢量 $\mathbf{w}$ 的直线，并且标记出投影后的点在直线上的位置，如图 1-2 所示：

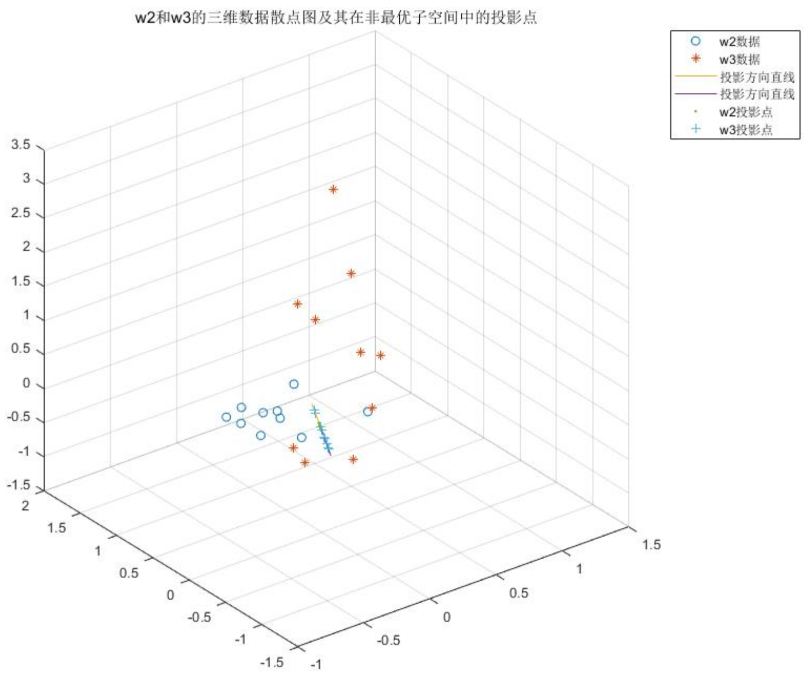


图 1-2. w2 和 w3 的三维数据散点图及其在非最优子空间中的投影点

对投影后得到的两个一维数据集合，假设它们都满足一维高斯分布，分别计算出它们的均值和方差如下：

	均值	方差
w_2 投影后的一维数据	0.2754	0.0245
w_3 投影后的一维数据	-0.0531	1.3930

假设它们的先验概率 $P(w_2)=P(w_3)=0.5$，设计一个方向 $\mathbf{w}$ 上的一维贝叶斯分类器；并使用设计得到的贝叶斯分类器对 w_2 和 w_3 的所有数据点进行分类，并计算分类器的训练误差，即错分点的个数，结果如下：

在非最优子空间中，
使用贝叶斯分类器对w2和w3的所有数据点进行分类
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
w2数据的训练误差：错分点为0个

------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
------------w3样本点：第2类-------------
------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
------------w3样本点：第2类-------------
w3数据的训练误差：错分点为2个

2. 拓展实验：MDA 分析

对表格 2 中的类别 w_1，w_2 和 w_3，计算得到最优二维投影子空间 $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2]$如下：

$$\mathbf{W} = \begin{bmatrix} -0.1386 & 0.9081 \\ -0.5653 & -0.4046 \\ -0.8132 & 0.1079 \end{bmatrix}$$

在一幅图中用不同的颜色画出表 2 中三个类别数据的三维散点图，如图 2-1 所示：

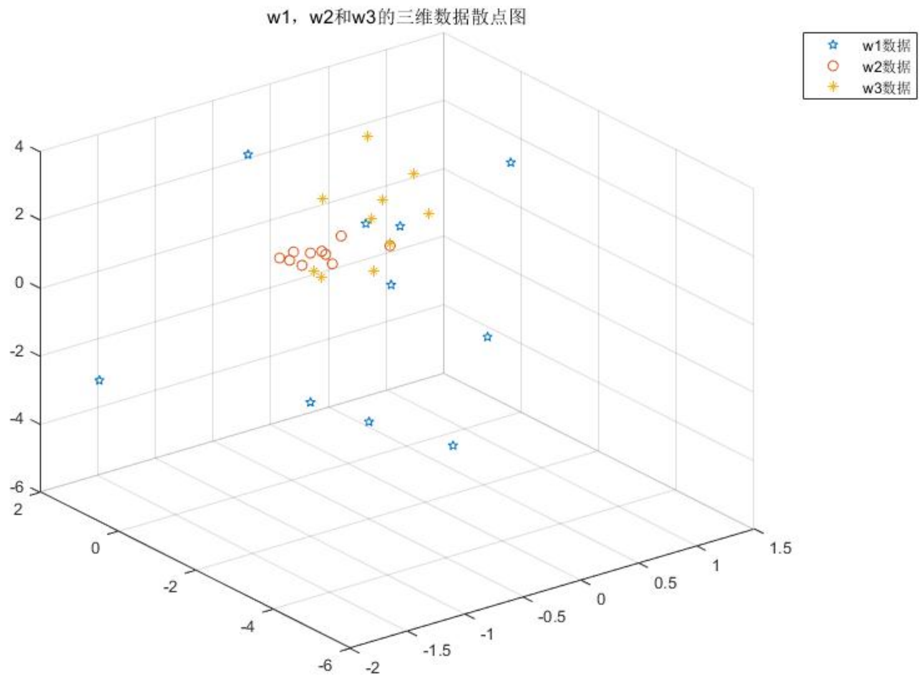


图 2-1. w1，w2 和 w3 的三维数据散点图

在一幅图中用不同颜色画出计算出的三个类别数据的二维散点图，如图 2-2 所示：

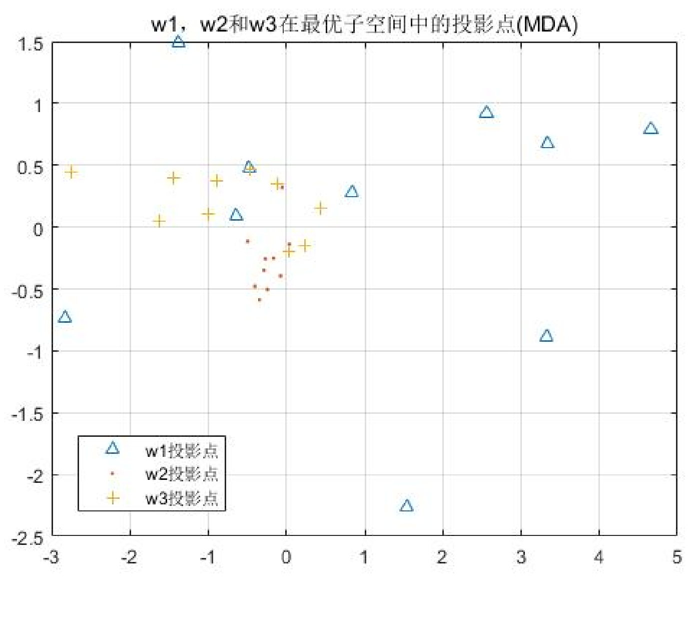


图 2-2. w1，w2 和 w3 在最优子空间中的投影点（MDA）

对投影后得到的三类二维数据集合，假设它们都满足高斯分布，分别计算出它们的均值向量和协方差矩阵。

	均值向量	协方差矩阵
w_1 投影后的二维数据	$[1.0924, 0.0820]$	$\begin{bmatrix} 7.3202 & 6.2991 \\ 6.2991 & 7.3202 \end{bmatrix}$
w_2 投影后的二维数据	$[-0.2304, -0.2758]$	$\begin{bmatrix} 0.0880 & 0.0860 \\ 0.0860 & 0.0880 \end{bmatrix}$
w_3 投影后的二维数据	$[-0.7595, 0.2013]$	$\begin{bmatrix} 1.8613 & 0.9382 \\ 0.9382 & 1.8613 \end{bmatrix}$

假设它们的先验概率 $P(w_1) = P(w_2) = P(w_3) = 1/3$，设计子空间 **W** 上的贝叶斯分类器；并使用设计得到的贝叶斯分类器对 w_1，w_2 和 w_3 的所有数据点进行分类，并计算分类器的训练误差，即错分点的个数，结果如下：

```
在最优子空间中(MDA)，
使用贝叶斯分类器对w1，w2和w3的所有数据点进行分类
------------w1样本点：第3类-------------
------------w1样本点：第1类-------------
------------w1样本点：第3类-------------
------------w1样本点：第3类-------------
------------w1样本点：第1类-------------
------------w1样本点：第1类-------------
------------w1样本点：第3类-------------
------------w1样本点：第3类-------------
------------w1样本点：第1类-------------
------------w1样本点：第1类-------------
w1数据的训练误差：错分点为5个
```

```
------------w2样本点：第3类-------------
------------w2样本点：第3类-------------
------------w2样本点：第3类-------------
------------w2样本点：第2类-------------
------------w2样本点：第3类-------------
------------w2样本点：第3类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
w2数据的训练误差：错分点为5个
```

```
------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
------------w3样本点：第2类-------------
------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
w3数据的训练误差：错分点为1个
```

对比实验：假设 $v = (1.0, 2.0, -1.5)^T$，$v2 = (-1.0, 0.5, -1.0)^T$，令非最优子空间基矢量 $\mathbf{w}_i = \mathbf{v}_i / \|\mathbf{v}_i\|$，重复上面步骤，在一幅图中用不同的颜色画出表 2 中三个类别数据的三维散点图，如图 2-3 所示：

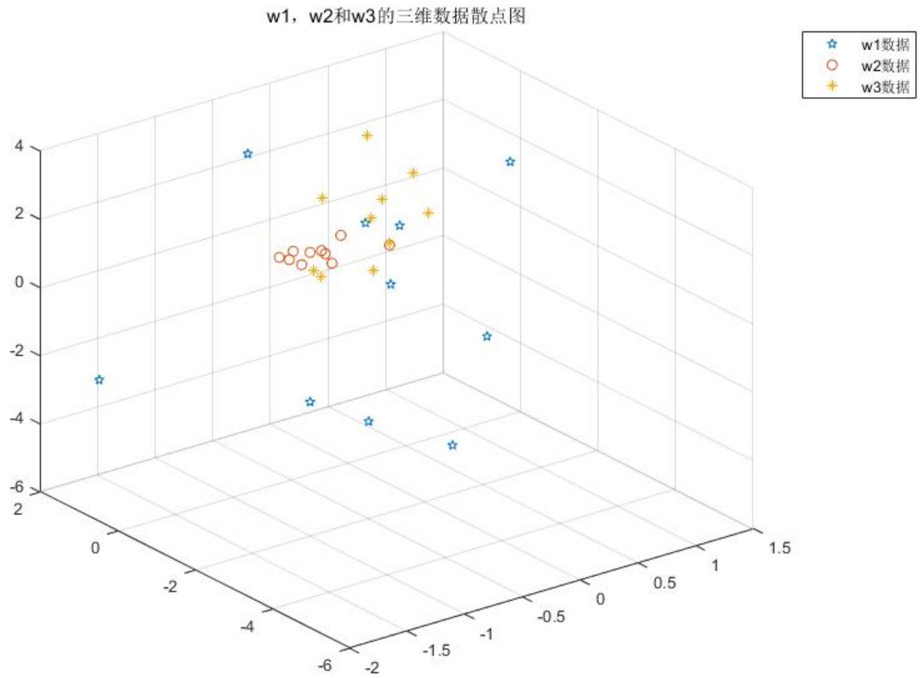


图 2-3. w1，w2 和 w3 的三维数据散点图

在一幅图中用不同颜色画出计算出的三个类别数据的二维散点图，如图 2-4 所示：

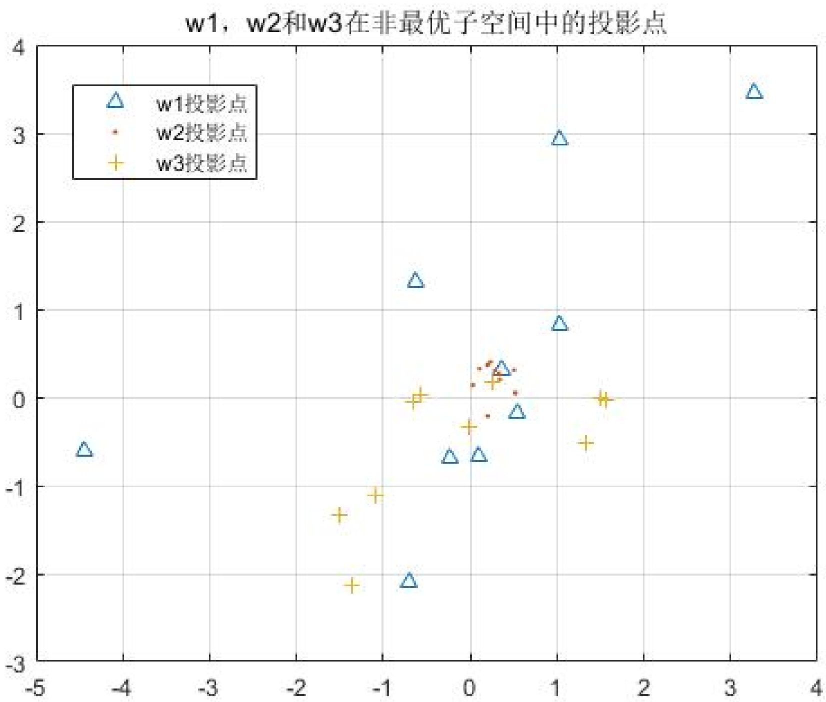


图 2-4. w1，w2 和 w3 在非最优子空间中的投影点

对投影后得到的三类二维数据集合，假设它们都满足高斯分布，分别计算出它们的均值向量和协方差矩阵。

	均值向量	协方差矩阵
w_1 投影后的二维数据	$[0.0320, 0.4530]$	$\begin{bmatrix} 6.2324 & 6.0552 \\ 6.0552 & 6.2324 \end{bmatrix}$
w_2 投影后的二维数据	$[0.2754, 0.2159]$	$\begin{bmatrix} 0.0560 & 0.0525 \\ 0.0525 & 0.0560 \end{bmatrix}$
w_3 投影后的二维数据	$[-0.0531, -0.5355]$	$\begin{bmatrix} 2.0024 & 1.7697 \\ 1.7697 & 2.0024 \end{bmatrix}$

假设它们的先验概率 $P(w_1) = P(w_2) = P(w_3) = 1/3$，设计子空间 **W** 上的贝叶斯分类器；并使用设计得到的贝叶斯分类器对 w_1，w_2 和 w_3 的所有数据点进行分类，并计算分类器的训练误差，即错分点的个数，结果如下：

```
在非最优子空间中，
使用贝叶斯分类器对w1，w2和w3的所有数据点进行分类
------------w1样本点：第3类-------------
------------w1样本点：第1类-------------
------------w1样本点：第3类-------------
------------w1样本点：第3类-------------
------------w1样本点：第1类-------------
------------w1样本点：第1类-------------
------------w1样本点：第3类-------------
------------w1样本点：第2类-------------
------------w1样本点：第1类-------------
------------w1样本点：第3类-------------
w1数据的训练误差：错分点为6个
```

```
------------w2样本点：第2类-------------
------------w2样本点：第1类-------------
------------w2样本点：第3类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
------------w2样本点：第3类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
------------w2样本点：第2类-------------
w2数据的训练误差：错分点为3个
```

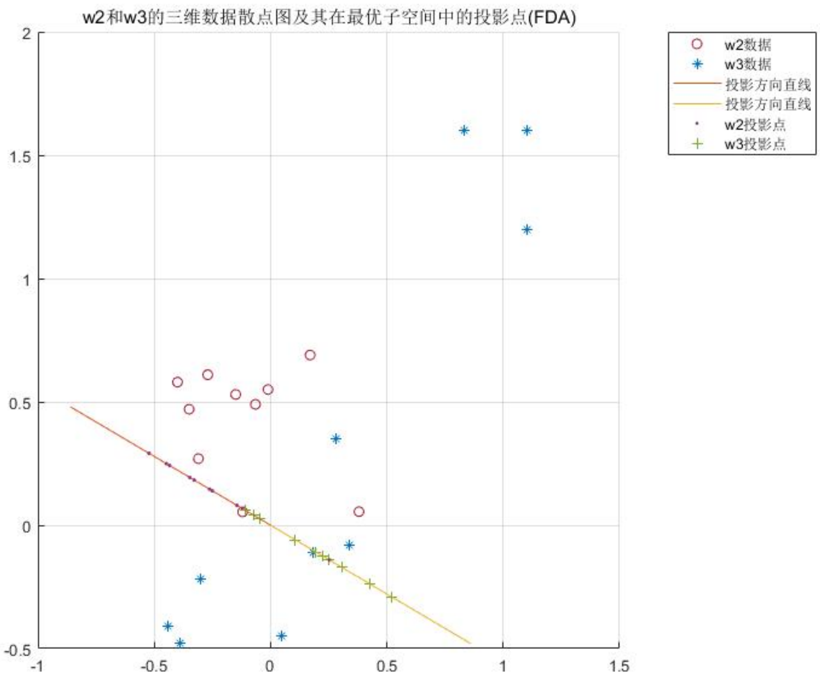
```
------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
------------w3样本点：第1类-------------
------------w3样本点：第3类-------------
------------w3样本点：第3类-------------
------------w3样本点：第1类-------------
------------w3样本点：第3类-------------
------------w3样本点：第1类-------------
------------w3样本点：第3类-------------
------------w3样本点：第2类-------------
w3数据的训练误差：错分点为4个
```

四、 讨论与分析

1. 基本实验：FDA 分析

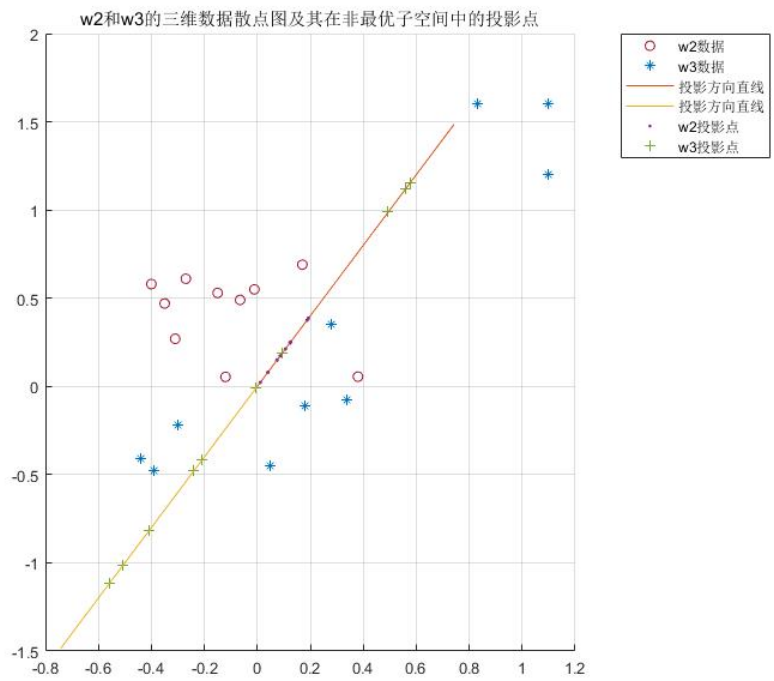
① 由实验结果中的图 1-1 所示，w_2 和 w_3 的所有三维数据点经过在矢量方向 **w** 上的投影，得到了在矢量方向 **w** 上的两个一维数据集合，实现了 FDA 的降维操作。

② 将图 1-1 进行旋转并放大，得到如下图所示：



观察 w_2 和 w_3 数据投影后的点在直线上的位置，可以发现，w_2 投影点和 w_3 投影点很容易区分开来，即不同类样本之间距离较远，而同类样本之间聚合得较紧密，这正符合 FDA 准则函数（式（1-12））中关于类间散布矩阵与类内散布矩阵之间的比值的解释；

对比在非最优子空间中的投影，这里将图 1-2 进行旋转并放大，得到如下图所示：



观察 w_2 和 w_3 数据投影后的点在直线上的位置，可以发现，w_2 投影点和 w_3 投影点

混在一起，而不容易（用一个线性分类器）将其分开。

以上的对比结果说明了同一组样本点在不同方向投影后可分性不同；验证了 FDA 是用于寻找对不同样本类样本进行区分的最有效方向；同时也说明了使用降维的方法能够使得不同类别的数据有一个更好的分类界面。

③ 由上面实验结果对 w2 和 w3 的所有数据点进行分类的贝叶斯分类器的训练误差，可分别计算在最优子空间（FDA）和非最优子空间中的贝叶斯分类器的分类准确率如下：

最优子空间（FDA）	错分点个数	分类准确率
w_2 数据	1	90%
w_3 数据	3	70%

非最优子空间	错分点个数	分类准确率
w_2 数据	0	100%
w_3 数据	2	80%

对比 w_2 和 w_3 数据在最优子空间（FDA）与非最优子空间上的分类结果，可得结论：w_2 和 w_3 数据在最优子空间（FDA）的分类准确率比在非最优子空间的分类准确率都低。分析其原因：这是因为训练样本只有两类总共 20 个样本，无法体现真实的误差情况。因此，需要使用更多的样本点才能准确验证最优子空间的分类结果确实是最优。

2. 拓展实验：MDA 分析

① 由实验结果中的图 2-2 所示，w_1，w_2 和 w_3 的所有三维数据点经过在二维子空间 **W** 上的投影，得到了在该二维子空间 **W** 上的三类二维数据集合，实现了 MDA 的降维操作。

② 对比图 2-2 和图 2-4，可以发现，相比图 2-4 中 w_1，w_2 和 w_3 数据在非最优子空间上的投影点混淆在一起，图 2-2 中 w_1，w_2 和 w_3 数据在最优子空间上的投影点不同类之间的距离较远，而类内相对聚合得较紧密，三类数据的投影点较容易区分开来。对比结果同样说明了 MDA 是 FDA 的高维推广，用于寻找对不同样本类样本进行区分的最有效子空间。

③ 由上面实验结果对 w1，w2 和 w3 的所有数据点进行分类的贝叶斯分类器的训练误差，可分别计算在最优子空间（MDA）和非最优子空间中的贝叶斯分类器的分类准确率如下：

最优子空间（MDA）	错分点个数	分类准确率
w_1 数据	5	50%
w_2 数据	5	50%
w_3 数据	1	90%

非最优子空间	错分点个数	分类准确率
w_1 数据	6	40%
w_2 数据	3	70%
w_3 数据	2	80%

对比 w_1，w_2 和 w_3 数据在最优子空间（MDA）与非最优子空间上的分类结果，可得：w_1 和 w_3 数据在最优子空间（MDA）中的分类准确率比在非最优子空间的分类准确率高，而 w_2 数据在最优子空间（MDA）中的分类准确率比在非最优子空间的分类准确率低。相比于两类数据的分类实验中的结论：在最优子空间（FDA）的分类准确率比在非最优子空间的分类准确率低，这里使用更多的样本点（三类数据），能验证得：在某种程度上，最优子空间的分类结果确实比非最优子空间的分类结果要优。

综上所述，实验结果表明，采用特征线性组合的方法可以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。判别分析是一类有效的线性组合变换方法。FDA 是用于寻找对不同样本类样本进行区分的最有效方向；MDA 是 FDA 的高维推广，用于寻找对不同样本类样本进行区分的最有效子空间。它们通过适当地选择投影方向，找到能够最大限度地区分各类样本数据点的投影方向，从而在低维空间中将数据进行分类。由于训练样本较少，某种程度下，数据在最优子空间的分类准确率比在非最优子空间的分类准确率低。但只要训练样本的数量足够大，最优子空间的分类效果是要比在非最优子空间下的分类效果更优的。

附录.

1. 基本实验：FDA 分析

```
%%---------------Proj03-02：Fisher 线性判别分析 FDA--------------%%
%%--------------------Proj03-02-exp1-------------------%%
clear; clc;
N = 10;
c = 3;
k = 1;%投影直线系数
%%第一类
w1 = [0.42 -0.087 0.58; -0.2 -3.3 -3.4; 1.3 -0.32 1.7; 0.39 0.71 0.23; -1.6 -5.3 -0.15;
    -0.029 0.89 -4.7; -0.23 1.9 2.2; 0.27 -0.3 -0.87; -1.9 0.76 -2.1; 0.87 -1.0 -2.6];
%%第二类
w2 = [-0.4 0.58 0.089; -0.31 0.27 -0.04; 0.38 0.055 -0.035; -0.15 0.53 0.011; -0.35 0.47 0.034;
    0.17 0.69 0.1; -0.011 0.55 -0.18; -0.27 0.61 0.12; -0.065 0.49 0.0012; -0.12 0.054 -0.063];
%%第三类
w3 = [0.83 1.6 -0.014; 1.1 1.6 0.48; -0.44 -0.41 0.32; 0.047 -0.45 1.4; 0.28 0.35 3.1;
    -0.39 -0.48 0.11; 0.34 -0.079 0.14; -0.3 -0.22 2.2; 1.1 1.2 -0.46; 0.18 -0.11 -0.49];
%%计算均值
m1 = mean(w1)';
m2 = mean(w2)';
m3 = mean(w3)';
%%计算类内散布矩阵
S1 = Intraclass_DM(w1, m1, N);
S2 = Intraclass_DM(w2, m2, N);
S3 = Intraclass_DM(w3, m3, N);
%%类别 w2 和 w3，计算最优方向矢量 w
Sw = S2 + S3;
w = inv(Sw) * (m2 - m3);
w = w / norm(w);%单位化
%%类别 w2 和 w3 的所有数据点在矢量方向 w 上的投影
y2 = w' * w2';%%得到的是标量，即投影的长度
y3 = w' * w3';
y2 = y2' * w';%%投影长度乘上方向矢量，即为投影矢量
y3 = y3' * w';
figure(1); W2 = plot3(w2(:, 1), w2(:, 2), w2(:, 3), 'o');%%画 w2 和 w3 的三维数据散点图
hold on; grid on; W3 = plot3(w3(:, 1), w3(:, 2), w3(:, 3), '*');
```

```
Line1 = plot3(k * [0 w(1)], k * [0 w(2)], k * [0 w(3)]);%%画表示方向矢量 w 的直线（投影方向直线）
Line2 = plot3(-k * [0 w(1)], -k * [0 w(2)], -k * [0 w(3)]);
Y2 = plot3(y2(:, 1), y2(:, 2), y2(:, 3), '.');%%标记出投影后的点在直线上的位置
Y3 = plot3(y3(:, 1), y3(:, 2), y3(:, 3), '+');
legend([W2, W3, Line1, Line2, Y2, Y3], 'w2 数据', 'w3 数据', '投影方向直线', '投影方向直线', 'w2 投影点', 'w3 投影点');
title('w2 和 w3 的三维数据散点图及其在最优子空间中的投影点(FDA)');

%%计算投影后得到的两个一维数据的均值和方差
miu2 = mean(y2)'; miu3 = mean(y3)';
s2 = var(y2); s3 = var(y3);
%%利用最小错误率贝叶斯分类器对训练类样本进行分类
p_w2 = 0.5; p_w3 = 0.5;%%贝叶斯分类器的先验概率
[f_max2, pre_b2] = Bayes_cla(y2, miu2, miu3, s2, s3, p_w2, p_w3);
[f_max3, pre_b3] = Bayes_cla(y3, miu2, miu3, s2, s3, p_w2, p_w3);
%%计算分类器的训练误差，即错分点的个数
label2 = 2 * ones(N, 1); label3 = 3 * ones(N, 1);%标签
error2 = length(find((pre_b2 + 1 - label2)~=0));%w2 错误分类的个数
error3 = length(find((pre_b3 + 1 - label3)~=0));%w3 错误分类的个数
fprintf('在最优子空间中(FDA)，\n 使用贝叶斯分类器对 w2 和 w3 的所有数据点进行分类\n');
fprintf('------------w2 样本点: 第%d 类-------------\n', pre_b2 + 1);
fprintf('w2 数据的训练误差：错分点为%d 个\n\n', error2);
fprintf('------------w3 样本点: 第%d 类-------------\n', pre_b3 + 1);
fprintf('w3 数据的训练误差：错分点为%d 个\n', error3);

%% 对比实验:在非最优子空间中，使用贝叶斯分类器计算 w2 和 w3 数据的训练误差
v = [1.0 2.0 -1.5]';
ww = v / norm(v);%单位化
%%类别 w2 和 w3 的所有数据点在矢量方向 w 上的投影
yy2 = ww' * w2';%%得到的是标量，即投影的长度
yy3 = ww' * w3';
yy2 = yy2' * ww';%%投影长度乘上方向矢量，即为投影矢量
yy3 = yy3' * ww';
figure(2); WW2 = plot3(w2(:, 1), w2(:, 2), w2(:, 3), 'o');%%画 w2 和 w3 的三维数据散点图
hold on; grid on; WW3 = plot3(w3(:, 1), w3(:, 2), w3(:, 3), '*');
line1 = plot3(2 * k * [0 ww(1)], 2 * k * [0 ww(2)], 2 * k * [0 ww(3)]);%%画表示方向矢量 w 的直线（投影方向直线）
line2 = plot3(-2 * k * [0 ww(1)], -2 * k * [0 ww(2)], -2 * k * [0 ww(3)]);
YY2 = plot3(yy2(:, 1), yy2(:, 2), yy2(:, 3), '.');%%标记出投影后的点在直线上的位置
YY3 = plot3(yy3(:, 1), yy3(:, 2), yy3(:, 3), '+');
legend([WW2, WW3, line1, line2, YY2, YY3], 'w2 数据', 'w3 数据', '投影方向直线', '投影方向直线', 'w2 投影点', 'w3 投影点');
```

```
title('w2 和 w3 的三维数据散点图及其在非最优子空间中的投影点');

%%计算投影后得到的两个一维数据的均值和方差
Miu2 = mean(yy2)'; Miu3 = mean(yy3)';
ss2 = var(yy2); ss3 = var(yy3);
%%利用最小错误率贝叶斯分类器对训练类样本进行分类
p_w2 = 0.5; p_w3 = 0.5;%%贝叶斯分类器的先验概率
[F_max2, Pre_b2] = Bayes_cla(yy2, Miu2, Miu3, ss2, ss3, p_w2, p_w3);
[F_max3, Pre_b3] = Bayes_cla(yy3, Miu2, Miu3, ss2, ss3, p_w2, p_w3);
%%计算分类器的训练误差，即错分点的个数
Error2 = length(find((Pre_b2 + 1 - label2)~=0));%w2 错误分类的个数
Error3 = length(find((Pre_b3 + 1 - label3)~=0));%w3 错误分类的个数
fprintf('\n 在非最优子空间中，\n 使用贝叶斯分类器对 w2 和 w3 的所有数据点进行分类\n');
fprintf('------------w2 样本点：第%d 类-------------\n', Pre_b2 + 1);
fprintf('w2 数据的训练误差：错分点为%d 个\n\n', Error2);
fprintf('------------w3 样本点：第%d 类-------------\n', Pre_b3 + 1);
fprintf('w3 数据的训练误差：错分点为%d 个\n', Error3);

%% ---------------------子函数------------------------- %%
function S = Intraclass_DM(x, m, N) %%计算类内散布矩阵；x 为矩阵，m 为向量，N 为样本数
目（标量）
S = zeros(size(m, 1));
for i = 1: N
    A = (x(i, :)' - m) * (x(i, :)' - m)';
    S = A + S;
end
end

%%设计一个方向 w 上的一维贝叶斯分类器
function [f_max, pre_b] = Bayes_cla(x, m1, m2, S1, S2, p_w1, p_w2)
N = size(x, 1);
p1 = zeros(N, 1); p2 = zeros(N, 1);
g1 = zeros(N, 1); g2 = zeros(N, 1);
f_max = zeros(N, 1); pre_b = zeros(N, 1);
for i = 1 : N
    p1(i) = mvnpdf(x(i, :), m1', S1); %条件概率
    p2(i) = mvnpdf(x(i, :), m2', S2);
    g1(i) = p1(i) .* p_w1; %分类器函数
    g2(i) = p2(i) .* p_w2;
    [f_max(i), pre_b(i)] = max([g1(i); g2(i)]);
end
end
```

2. 拓展实验：MDA 分析

```
%%----------------Proj03-02：Fisher 线性判别分析 FDA--------------%%
%%---------------Proj03-02-exp2 多重判别分析(MDA)---------------%%
clear; clc;
N = 10;
c = 3;
k = 1;%投影直线系数
%%第一类
w1 = [0.42 -0.087 0.58; -0.2 -3.3 -3.4; 1.3 -0.32 1.7; 0.39 0.71 0.23; -1.6 -5.3 -0.15;
    -0.029 0.89 -4.7; -0.23 1.9 2.2; 0.27 -0.3 -0.87; -1.9 0.76 -2.1; 0.87 -1.0 -2.6];
%%第二类
w2 = [-0.4 0.58 0.089; -0.31 0.27 -0.04; 0.38 0.055 -0.035; -0.15 0.53 0.011; -0.35 0.47 0.034;
    0.17 0.69 0.1; -0.011 0.55 -0.18; -0.27 0.61 0.12; -0.065 0.49 0.0012; -0.12 0.054 -0.063];
%%第三类
w3 = [0.83 1.6 -0.014; 1.1 1.6 0.48; -0.44 -0.41 0.32; 0.047 -0.45 1.4; 0.28 0.35 3.1;
    -0.39 -0.48 0.11; 0.34 -0.079 0.14; -0.3 -0.22 2.2; 1.1 1.2 -0.46; 0.18 -0.11 -0.49];
%%计算均值
m1 = mean(w1)';
m2 = mean(w2)';
m3 = mean(w3)';
m = (N * m1 + N * m2 + N * m3) / (3 * N);%%总体均值向量
%%计算类内散布矩阵
S1 = Intraclass_DM(w1, m1, N);
S2 = Intraclass_DM(w2, m2, N);
S3 = Intraclass_DM(w3, m3, N);

%%类别 w1，w2 和 w3，计算最优方向矢量 w
Sw = S1 + S2 + S3;
Sb = Interclass_DM(m1, m2, m3, m, N, c);%计算类间散布矩阵
S = inv(Sw) * Sb;
[V, D] = eig(S);%%D 的对角线元素是特征值，V 的列是相应的特征向量
[D_sort, index] = sort(diag(D),'descend');
V_sort = V(:,index);
W1 = V_sort(:, 1); W2 = V_sort(:, 2);
W1 = W1 / norm(W1); W2 = W2 / norm(W2);%单位化
W = [W1 W2];
y1 = W' * w1';%%三个类别 w1、w2 和 w3 的所有数据点在二维子空间 W 上的投影
y2 = W' * w2';
```

```
y3 = W' * w3';
figure(1); WW1 = plot3(w1(:, 1), w1(:, 2), w1(:, 3), 'p');%%画 w1,w2 和 w3 的三维数据散点图
hold on; grid on; WW2 = plot3(w2(:, 1), w2(:, 2), w2(:, 3), 'o');
WW3 = plot3(w3(:, 1), w3(:, 2), w3(:, 3), '*');
legend([WW1, WW2, WW3], 'w1 数据', 'w2 数据', 'w3 数据');
title('w1，w2 和 w3 的三维数据散点图');
figure(2); Y1 = plot(y1(1, :), y1(2, :), '^');%%三个类别数据的二维散点图
hold on; grid on; Y2 = plot(y2(1, :), y2(2, :), '.');
Y3 = plot(y3(1, :), y3(2, :), '+');
legend([Y1, Y2, Y3], 'w1 投影点', 'w2 投影点', 'w3 投影点');
title('w1，w2 和 w3 在最优子空间中的投影点(MDA)');

%%计算投影后得到的三类二维数据的均值向量和协方差矩阵
miu1 = mean(y1')'; miu2 = mean(y2')'; miu3 = mean(y3')';
s1 = Cov(y1, miu1, N); s2 = Cov(y2, miu2, N); s3 = Cov(y3, miu3, N);
% s1 = cov(y1'); s2 = cov(y2'); s3 = cov(y3');%可以用这个函数计算协方差矩阵吗？
%%利用最小错误率贝叶斯分类器对投影后得到的三类二维数据集合进行分类
p_w1 = 1/3; p_w2 = 1/3; p_w3 = 1/3;%%贝叶斯分类器的先验概率
[f_max1, pre_b1] = Bayes_cla(y1, miu1, miu2, miu3, s1, s2, s3, p_w1, p_w2, p_w3);
[f_max2, pre_b2] = Bayes_cla(y2, miu1, miu2, miu3, s1, s2, s3, p_w1, p_w2, p_w3);
[f_max3, pre_b3] = Bayes_cla(y3, miu1, miu2, miu3, s1, s2, s3, p_w1, p_w2, p_w3);
%%计算分类器的训练误差，即错分点的个数
label1 = ones(N, 1); label2 = 2 * ones(N, 1); label3 = 3 * ones(N, 1);%标签
error1 = length(find((pre_b1 - label1)~=0));%w1 错误分类的个数
error2 = length(find((pre_b2 - label2)~=0));%w2 错误分类的个数
error3 = length(find((pre_b3 - label3)~=0));%w3 错误分类的个数
fprintf('在最优子空间中(MDA)，\n 使用贝叶斯分类器对 w1，w2 和 w3 的所有数据点进行分类\n');
fprintf('------------w1 样本点: 第%d 类-------------\n', pre_b1);
fprintf('w1 数据的训练误差：错分点为%d 个\n\n', error1);
fprintf('------------w2 样本点: 第%d 类-------------\n', pre_b2);
fprintf('w2 数据的训练误差：错分点为%d 个\n\n', error2);
fprintf('------------w3 样本点: 第%d 类-------------\n', pre_b3);
fprintf('w3 数据的训练误差：错分点为%d 个\n\n', error3);

%% 对比实验:在非最优子空间中，使用贝叶斯分类器计算 w2 和 w3 数据的训练误差
v1 = [1.0 2.0 -1.5]'; v2 = [-1.0 0.5 -1.0]';
ww1 = v1 / norm(v1);%单位化
ww2 = v2 / norm(v2);
%%类别 w2 和 w3 的所有数据点在矢量方向 w 上的投影
W = [ww1 ww2];
yy1 = W' * w1';%%三个类别 w1、w2 和 w3 的所有数据点在二维子空间 W 上的投影
yy2 = W' * w2';
yy3 = W' * w3';
```

```
figure(3); YY1 = plot(yy1(1, :), yy1(2, :), '^');%%标记出投影后的点在直线上的位置
hold on; grid on; YY2 = plot(yy2(1, :), yy2(2, :), '.');
YY3 = plot(yy3(1, :), yy3(2, :), '+');
legend([YY1, YY2, YY3], 'w1 投影点', 'w2 投影点', 'w3 投影点');
title('w1，w2 和 w3 在非最优子空间中的投影点');

%%计算投影后得到的三个二维数据的均值和方差
Miu1 = mean(yy1')'; Miu2 = mean(yy2')'; Miu3 = mean(yy3')';
ss1 = Cov(yy1, Miu1, N); ss2 = Cov(yy2, Miu2, N); ss3 = Cov(yy3, Miu3, N);
% ss1 = cov(yy1'); ss2 = cov(yy2'); ss3 = cov(yy3');%可以用这个函数计算协方差矩阵
吗？
%%利用最小错误率贝叶斯分类器对训练类样本进行分类
p_w1= 1/3; p_w2 = 1/3; p_w3 = 1/3;%%贝叶斯分类器的先验概率
[F_max1, Pre_b1] = Bayes_cla(yy1, Miu1, Miu2, Miu3, ss1, ss2, ss3, p_w1, p_w2,
p_w3);
[F_max2, Pre_b2] = Bayes_cla(yy2, Miu1, Miu2, Miu3, ss1, ss2, ss3, p_w1, p_w2,
p_w3);
[F_max3, Pre_b3] = Bayes_cla(yy3, Miu1, Miu2, Miu3, ss1, ss2, ss3, p_w1, p_w2,
p_w3);
%%计算分类器的训练误差，即错分点的个数
Error1 = length(find((Pre_b1 - label1)~=0));%w1 错误分类的个数
Error2 = length(find((Pre_b2 - label2)~=0));%w1 错误分类的个数
Error3 = length(find((Pre_b3 - label3)~=0));%w1 错误分类的个数
fprintf('\n 在非最优子空间中，\n 使用贝叶斯分类器对 w1，w2 和 w3 的所有数据点进行分类\n');
fprintf('------------w1 样本点: 第%d 类-------------\n', Pre_b1);
fprintf('w1 数据的训练误差：错分点为%d 个\n\n', Error1);
fprintf('------------w2 样本点: 第%d 类-------------\n', Pre_b2);
fprintf('w2 数据的训练误差：错分点为%d 个\n\n', Error2);
fprintf('------------w3 样本点: 第%d 类-------------\n', Pre_b3);
fprintf('w3 数据的训练误差：错分点为%d 个\n\n', Error3);

%% ---------------------子函数------------------------- %%
function S = Intraclass_DM(x, m, N) %%计算类内散布矩阵；x 为矩阵，m 为向量，N 为样本数
目（标量）
S = zeros(size(m, 1));
for i = 1: N
    A = (x(i, :)' - m) * (x(i, :)' - m)';
    S = A + S;
end
end

function Sb = Interclass_DM(m1, m2, m3, m, N, c) %%计算类间散布矩阵；x 为矩阵，m 为
向量，c 为样本类别数量
mm = [m1 m2 m3];
```

```matlab
Sb = zeros(size(c, 1));
for i = 1: c
    A = N .* (mm(:, i) - m) * (mm(:, i) - m)';
    Sb = A + Sb;
end
end

%%协方差矩阵计算函数
function S = Cov(x, m, N) %%x 为矩阵，m 为向量，N 为样本数目（标量）
S = zeros(size(m, 1));
for i = 1: N
    A = (1 / N) .* ((x(:, i)' - m) * (x(:, i)' - m)');
    S = A + S;
end
end

%%设计一个方向 w 上的一维贝叶斯分类器
function [f_max, pre_b] = Bayes_cla(x, m1, m2, m3, S1, S2, S3, p_w1, p_w2, p_w3)
N = size(x, 2);
p1 = zeros(N, 1); p2 = zeros(N, 1); p3 = zeros(N, 1);
g1 = zeros(N, 1); g2 = zeros(N, 1); g3 = zeros(N, 1);
f_max = zeros(N, 1); pre_b = zeros(N, 1);
for i = 1 : N
    p1(i) = mvnpdf(x(:, i)', m1', S1); %条件概率
    p2(i) = mvnpdf(x(:, i)', m2', S2);
    p3(i) = mvnpdf(x(:, i)', m3', S3);
    g1(i) = p1(i) .* p_w1; %分类器函数
    g2(i) = p2(i) .* p_w2;
    g3(i) = p3(i) .* p_w3;
    [f_max(i), pre_b(i)] = max([g1(i); g2(i); g3(i)]);
end
end
```