

0. 影响聚类结果的主要因素有那些？ (2) 证明 马氏距离是平移不变的、非奇异线性变换不变的。

答：(1) 分类准则，模式相似性测度，特征量的选择，量纲。

1. 监督学习与非监督学习的区别：

监督学习方法用来对数据实现分类，分类规则通过训练获得。该训练集由带分类号的数据集组成，因此监督学习方法的训练过程是离线的。

非监督学习方法不需要单独的离线训练过程，也没有带分类号（标号）的训练数据集，一般用来对数据集进行分析，如聚类，确定其分布的主分量等。

（实例：道路图）就道路图像的分割而言，监督学习方法则先在训练用图像中获取道路像素与非道路像素集，进行分类器设计，然后用所设计的分类器对道路图像进行分割。

使用非监督学习方法，则依据道路路面像素与非道路像素之间的聚类分析进行聚类运算，以实现道路图像的分割。

2. 动态聚类 是指对当前聚类通过迭代运算改善聚类；

分级聚类则是将样本个体，按相似度标准合并，随着相似度要求的降低实现合并。

3. 线性分类器三种最优准则：

**Fisher 准则**：根据两类样本一般类内密集，类间分离的特点，寻找线性分类器最佳的法线向量方向，使两类样本在该方向上的投影满足类内尽可能密集，类间尽可能分开。

该种度量通过类内离散矩阵  $S_w$  和类间离散矩阵  $S_b$  实现。

**感知准则函数**：准则函数以使错分类样本到分界面距离之和最小为原则。

其优点是通过对错分类样本提供的信息对分类器函数进行修正，这种准则是人工神经网络多层感知器的基础。

**支持向量机**：基本思想是在两类线性可分条件下，所设计的分类器界面使两类之间的间隔为最大，它的基本出发点是使期望泛化风险尽可能小。

一、试问“模式”与“模式类”的含义。如果一位姓王的先生是位老年人，试问“王先生”和“老头”谁是模式，谁是模式类？

答：在模式识别学科中，就“模式”与“模式类”而言，模式类是一类事物的代表，概念或典型，而“模式”则是某一事物的具体体现，如“老头”是模式类，而王先生则是“模式”，是“老头”的具体化。

二、试说明 **Mahalanobis** 距离平方的定义，到某点的 Mahalanobis 距离平方为常数的轨迹的几何意义，它与欧氏距离的区别与联系。

答：Mahalanobis 距离的平方定义为：
$$r^2(x, u) = (x - u)^T \Sigma^{-1} (x - u)$$

其中  $x, u$  为两个数据， $\Sigma^{-1}$  是一个正定对称矩阵（一般为协方差矩阵）。根据定义，距某一点的 Mahalanobis 距离相等点的轨迹是超椭圆，如果是单位矩阵，则 Mahalanobis 距离就是通常的欧氏距离。

三、试说明用监督学习与非监督学习两种方法对道路图像中道路区域的划分的基本做法，以说明这两种学习方法的定义与它们间的区别。

答：监督学习方法用来对数据实现分类，分类规则通过训练获得。该训练集由带分类号的数据集组成，因此监督学习方法的训练过程是离线的。

非监督学习方法不需要单独的离线训练过程，也没有带分类号（标号）的训练数据集，一般用来对数据集进行分析，如聚类，确定其分布的主分量等。

就道路图像的分割而言，监督学习方法则先在训练用图像中获取道路像素与非道路像素集，进行分类器设计，然后用所设计的分类器对道路图像进行分割。

使用非监督学习方法，则依据道路路面像素与非道路像素之间的聚类分析进行聚类运算，以实现道路图像的分割。

四、试述动态聚类与分级聚类这两种方法的原理与不同。

答：动态聚类是指对当前聚类通过迭代运算改善聚类；分级聚类则是将样本个体，按相似度标准合并，随着相似度要求的降低实现合并。

五、如果观察一个时序信号时在离散时刻序列得到的观察量序列表示为  $O = \{o_1, \dots, o_L\}$ ，而该时序信号的内在状态序列表示成  $S = \{s_1, \dots, s_L\}$ 。如果计算在给定  $O$  条件下出现  $S$  的概率，试问此概率是何种概率。如果从观察序列来估计状态序列的最大似然估计，这与 Bayes 决策中基于最小错误率的决策有什么关系。

答：在给定观察序列  $O = \{o_1, \dots, o_L\}$  条件下分析它由某个状态序列  $S$  产生的概率似后验概率，写成  $P(S|O)$ ，而通过  $O$  求对状态序列的最大似然估计，与贝叶斯决策的最小错误率决策相当。

六、已知一组数据的协方差矩阵为  $\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$ ，试问

1. 协方差矩阵中各元素的含义。
2. 求该数组的两个主分量。
3. 主分量分析或称 K-L 变换，它的最佳准则是什么？
4. 为什么说经主分量分析后，消除了各分量之间的相关性。

答：协方差矩阵为  $\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$ ，则

1) 对角元素是各分量的方差，非对角元素是各分量之间的协方差。

2) 主分量，通过求协方差矩阵的特征值，用  $\begin{pmatrix} \lambda-1 & -1/2 \\ -1/2 & \lambda-1 \end{pmatrix}$  得  $(\lambda-1)^2 = \frac{1}{4}$ ，则  $\lambda = \begin{cases} \frac{1}{2} \\ \frac{3}{2} \end{cases}$ ，相应的特征向量

为： $\lambda = \frac{3}{2}$ ，对应特征向量为  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ， $\lambda = \frac{1}{2}$ ，对应  $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ 。  
这两个特征向量即为主分量。

3) K-L 变换的最佳准则为：对一组数据进行按一组正交基分解，在只取相同数量分量的条件下，以均方误差计算截尾误差最小。

4) 在经主分量分解后，协方差矩阵成为对角矩阵，因而各主分量间相关消除。

七、试说明以下问题求解是基于监督学习或是非监督学习：

1. 求数据集的主分量 2. 汉字识别 3. 自组织特征映射 4. CT 图像的分割

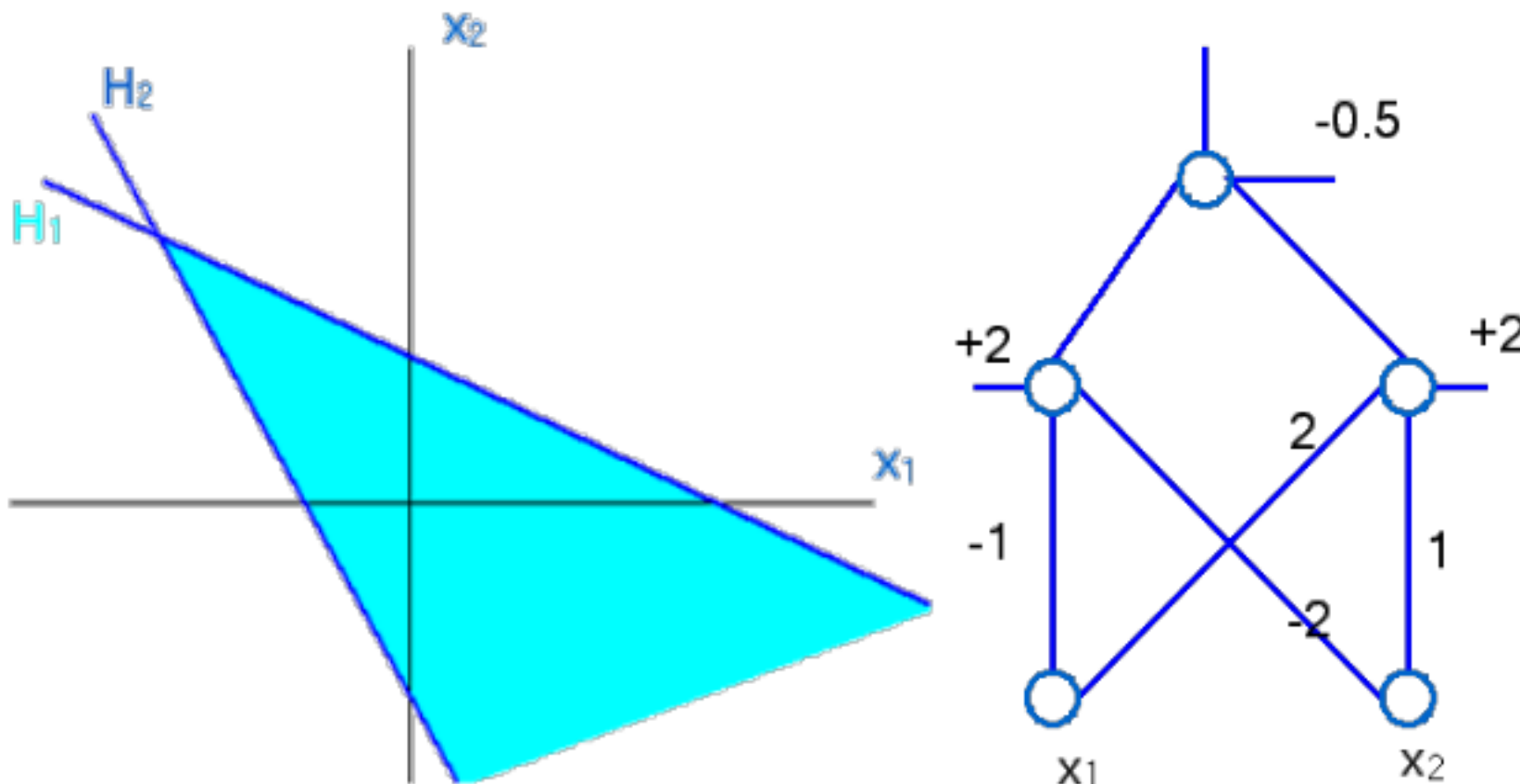
答：1、求数据集的主分量是非监督学习方法； 2、汉字识别对待识别字符加上相应类别号——有监督学习方法；  
3、自组织特征映射——将高维数组按保留近似度向低维映射——非监督学习；  
4、CT 图像分割——按数据自然分布聚类——非监督学习方法；

八、在一二维特征空间，两类决策域由两条直线  $H_1$  和  $H_2$  分界，

其中  $H_1: x_1 + 2x_2 - 2 = 0$   $H_2: 2x_1 + x_2 + 2 = 0$

而包含  $H_1$  与  $H_2$  的锐角部分为第一类，其余为第二类。

试求：1. 用一双层感知器构造该分类器 2. 用凹函数的并构造该分类器



答：按题意要求

1)  $H_1$  与  $H_2$  将空间划分成四个部分，按使  $H_1$  与  $H_2$  大于零与小于零表示成四个区域，而第一类属于  $(- +)$  区域，为方便起见，令  $H_1: -x_1 - 2x_2 + 2 = 0$  则第一类在  $(+ +)$  区域。用双层感知器，神经元用  $\pm 1$  域值，则在第一类样本输入时，两隐层结点的输出均为 +1，其余则分别为  $(+ -)$ ， $(- -)$ ， $(- +)$ ，故可按图设置域值。

2) 用凹函数的并表示： $H_1^{(x)} \cap H_2^{(x)}$  或表示成  $P^{(x)} = \min(H_1^{(x)}, H_2^{(x)})$ ，如  $P^{(x)} > 0$ ，则  $x \in \omega_1$ ，否则  $x \in \omega_2$

十、对一副道路图像，希望把道路部分划分出来，可以采用以下两种方法：

1. 在该图像中分别在道路部分与非道路部分画出一个窗口，把在这两个窗口中的像素数据作为训练集，用 Fisher 准则方法求得分类器参数，再用该分类器对整幅图进行分类。

2. 将整幅图的每个像素的属性记录在一张数据表中，然后用某种方法将这些数据按它们的自然分布状况划分成两类。因此每个像素就分别得到相应的类别号，从而实现了道路图像的分割。试问以上两种方法哪一种是监督学习，哪个是非监督学习？

答：第一种方法中标记了两类样本的标号，需要人手工干预训练过程，属于监督学习方法；

第二种方法只是依照数据的自然分布，把它们划分成两类，属于非监督学习方法。

十一、已知有两类数据，分别为

$$\omega_1: (1, 0), (2, 0), (1, 1)$$

$$\omega_2: (-1, 0), (0, 1), (-1, 1)$$

试求：该组数据的类内及类间离散矩阵  $S_w$  及  $S_b$ 。答：第一类的均值向量为

$$m_1 = \left(\frac{4}{3}, \frac{1}{3}\right), \quad m_2 = \left(-\frac{2}{3}, \frac{2}{3}\right) \quad \therefore S_1 = \frac{1}{9} \begin{pmatrix} 6 & -5 \\ -5 & 6 \end{pmatrix}, \quad S_2 = \frac{1}{3} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$S_w = S_1 + S_2 = \frac{1}{9} \begin{pmatrix} 12 & -2 \\ -2 & 12 \end{pmatrix} \quad S_b = \begin{pmatrix} \frac{6}{3} \\ -\frac{1}{3} \end{pmatrix} \begin{pmatrix} \frac{6}{3} & -\frac{1}{3} \end{pmatrix} = \frac{1}{9} \begin{pmatrix} 36 & -6 \\ -6 & 1 \end{pmatrix}$$

十三、试分析五种常用决策规则思想方法的异同。

答、五种常用决策是：

1. 基于最小错误率的贝叶斯决策，利用概率论中的贝叶斯公式，得出使得错误率最小的分类规则。
2. 基于最小风险的贝叶斯决策，引入了损失函数，得出使决策风险最小的分类。当在 0 - 1 损失函数条件下，基于最小风险的贝叶斯决策变成基于最小错误率的贝叶斯决策。
3. 在限定一类错误率条件下使另一类错误率最小的两类别决策。
4. 最大最小决策：类先验概率未知，考察先验概率变化对错误率的影响，找出使最小贝叶斯奉献最大的先验概率，以这种最坏情况设计分类器。
5. 序贯分类方法，除了考虑分类造成的损失外，还考虑特征获取造成的代价，先用一部分特征分类，然后逐步加入性特征以减少分类损失，同时平衡总的损失，以求得最有效益。

十五、既然有线性判别函数，为什么还要引进非线性判别函数？试分析由“线性判别函数”向“非线性判别函数”推广的思想和方法。

答：实际中有很多模式识别问题并不是线性可分的，这时就需要采用非线性分类器，比如当两类样本不具有多峰性质并互相交错时，简单的线性判别函数往往会带来较大的分类错误。这时，树分类器作为一种分段线性分类器，常常能有效地应用于这种情况。

十六、1. 什么是特征选择？ 2. 什么是 Fisher 线性判别？

答：1. 特征选择就是从一组特征中挑选出一些最有效的特征以达到降低特征空间维数的目的。

2. Fisher 线性判别：可以考虑把 d 维空间的样本投影到一条直线上，形成一维空间，即把维数压缩到一维，这在数学上容易办到，然而，即使样本在 d 维空间里形成若干紧凑的互相分得开的集群，如果把它们投影到一条任意的直线上，也可能使得几类样本混在一起而变得无法识别。但是在一般情况下，总可以找到某个方向，使得在这个方向的直线上，样本的投影能分开得最好。问题是如何根据实际情况找到这条最好的、最易于分类的投影线，这就是 Fisher 算法所要解决的基本问题。

十八、请论述模式识别系统的主要组成部分及其设计流程，并简述各组成部分中常用方法的主要思想。

信息获取：通过测量、采样和量化，可以用矩阵或向量表示二维图像或以为波形。

预处理：去除噪声，加强有用的信息，并对输入测量仪器或其他因素造成的退化现象进行复原。

特征选择和提取：为了有效地实现分类识别，就要对原始数据进行变换，得到最能反映分类本质的特征。

分类决策：在特征空间中用统计方法把识别对象归为某一类。

二十、定性说明基于参数方法和非参数方法的概率密度估计有什么区别？

答：基于参数方法：是由已知类别的样本集对总体分布的某些参数进行统计推断

非参数方法：已知样本所属类别，但未知总体概率密度函数形式

二十二、简述支持向量机的基本思想。

答：SVM 从线性可分情况下的最优分类面发展而来。

最优分类面就是要求分类线不但能将两类正确分开（训练错误率为 0），且使分类间隔最大。

SVM 考虑寻找一个满足分类要求的超平面，并且使训练集中的点距离分类面尽可能的远，也就是寻找一个分类面使它两侧的空白区域（margin）最大。

过两类样本中离分类面最近的点，且平行于最优分类面的超平面上  $H_1, H_2$  的训练样本就叫支持向量。

五、（12 分，每问 4 分）在目标识别中，假定有农田和装甲车两种类型，类型  $\omega_1$  和类型  $\omega_2$  分别代表农田和装甲车，它们的先验概率分别为 0.8 和 0.2，损失函数如表 1 所示。现在做了三次试验，获得三个样本的类概率密度如下：

$$p(x/\omega_1) : 0.3, 0.1, 0.6$$

$$p(x/\omega_2) : 0.7, 0.8, 0.3$$



- (1) 试用 贝叶斯最小误判概率准则 判决三个样本各属于哪一个类型；  
 (2) 假定只考虑前两种判决，试用 贝叶斯最小风险准则 判决三个样本各属于哪一类；  
 (3) 把拒绝判决考虑在内，重新考核三次试验的结果。

表 1

判决 \ 损失 \ 类型	$\omega_1$	$\omega_2$
$\alpha_1$	1	4
$\alpha_2$	5	1
$\alpha_3$	1	1

解：由题可知： $P(\omega_1)=0.7, P(\omega_2)=0.3$ ， $\frac{P(\omega_2)}{P(\omega_1)}=\frac{3}{7}$ ， $\frac{P(x_1|\omega_1)}{P(x_1|\omega_2)}=\frac{3}{7}$ ，  
 $\frac{P(x_2|\omega_1)}{P(x_2|\omega_2)}=\frac{1}{8}$ ， $\frac{P(x_3|\omega_1)}{P(x_3|\omega_2)}=2$

(1) (4 分) 根据贝叶斯最小误判概率准则知：

$\frac{P(x_1|\omega_1)}{P(x_1|\omega_2)} = \frac{P(\omega_2)}{P(\omega_1)}$ ，则可以任判；  
 $\frac{P(x_2|\omega_1)}{P(x_2|\omega_2)} < \frac{P(\omega_2)}{P(\omega_1)}$ ，则判为  $\omega_2$ ；  
 $\frac{P(x_3|\omega_1)}{P(x_3|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$ ，则判为  $\omega_1$ ；

(2) (4 分) 由题可知： $\frac{P(\omega_2)(\lambda_{21}-\lambda_{22})}{P(\omega_1)(\lambda_{12}-\lambda_{11})} = \frac{0.3(5-1)}{0.7(4-1)} = \frac{4}{7}$

则  $\frac{P(x_1|\omega_1)}{P(x_1|\omega_2)} < \frac{4}{7}$ ，判为  $\omega_2$ ；  
 $\frac{P(x_2|\omega_1)}{P(x_2|\omega_2)} < \frac{4}{7}$ ，判为  $\omega_2$ ；  
 $\frac{P(x_3|\omega_1)}{P(x_3|\omega_2)} > \frac{4}{7}$ ，判为  $\omega_1$ ；

(3) (4 分) 对于两类问题，对于样本  $x$ ，假设  $P(x)$  已知，有

$$R(\alpha_j|x) = \lambda(\alpha_j|\omega_1)P(\omega_1|x) + \lambda(\alpha_j|\omega_2)P(\omega_2|x) =$$

$$= \frac{\lambda(\alpha_j|\omega_1)P(x|\omega_1)P(\omega_1) + \lambda(\alpha_j|\omega_2)P(x|\omega_2)P(\omega_2)}{P(x)}$$

则对于第一个样本，

$$R(\alpha_1|x) = \frac{5 \times 0.21}{P(x)}, R(\alpha_2|x) = \frac{4 \times 0.21}{P(x)}, R(\alpha_3|x) = \frac{2 \times 0.21}{P(x)} \text{，则拒判；}$$

$$R(\alpha_1|x) = \frac{1.03}{P(x)}, R(\alpha_2|x) = \frac{0.59}{P(x)}, R(\alpha_3|x) = \frac{0.24}{P(x)} \text{，则拒判；}$$

$$R(\alpha_1|x) = \frac{0.78}{P(x)}, R(\alpha_2|x) = \frac{2.19}{P(x)}, R(\alpha_3|x) = \frac{0.51}{P(x)} \text{，拒判。}$$

十四、假设在某个地区细胞识别中正常 ( $w_1$ ) 和异常 ( $w_2$ ) 两类先验概率分别为  $P(w_1)=0.9$ ， $P(w_2)=0.1$ ，现有一待识别的细胞，其观察值为  $x$ ，从类条件概率密度分布曲线上查得

$P(x|w_1) = 0.2$ ， $P(x|w_2) = 0.4$ ，并且已知  $\lambda_{11} = 0$ ， $\lambda_{12} = 6$ ，

$\lambda_{21} = 1$ ， $\lambda_{22} = 0$  试对该细胞  $x$  用一下两种方法进行分类： 1 基于最小错误率的贝叶斯决策； 2 基于最小风险的贝叶

斯决策；请分析两种结果的异同及原因。

答：1.

解：利用贝叶斯公式，分别计算出  $\omega_1$  及  $\omega_2$  的后验概率。

$$P(\omega_1 | \mathbf{x}) = \frac{p(\mathbf{x} | \omega_1)P(\omega_1)}{\sum_{j=1}^2 p(\mathbf{x} | \omega_j)P(\omega_j)} = \frac{0.2 \times 0.9}{0.2 \times 0.9 + 0.4 \times 0.1} = 0.818$$

$$P(\omega_2 | \mathbf{x}) = 1 - p(\omega_1 | \mathbf{x}) = 0.182$$

根据贝叶斯决策规则式(2-2)，有

$$P(\omega_1 | \mathbf{x}) = 0.818 > P(\omega_2 | \mathbf{x}) = 0.182$$

所以合理的决策是把  $\mathbf{x}$  归类于正常状态。

2.

解：已知条件为

$$\begin{aligned} P(\omega_1) &= 0.9, & P(\omega_2) &= 0.1 \\ p(\mathbf{x} | \omega_1) &= 0.2, & p(\mathbf{x} | \omega_2) &= 0.4 \\ \lambda_{11} &= 0, & \lambda_{12} &= 6 \\ \lambda_{21} &= 1, & \lambda_{22} &= 0 \end{aligned}$$

根据 1 的计算结果可知后验概率为

$$P(\omega_1 | \mathbf{x}) = 0.818, \quad P(\omega_2 | \mathbf{x}) = 0.182$$

再计算出条件风险

$$R(\alpha_1 | \mathbf{x}) = \sum_{j=1}^2 \lambda_{1j} P(\omega_j | \mathbf{x}) = \lambda_{12} P(\omega_2 | \mathbf{x}) = 1.092$$

$$R(\alpha_2 | \mathbf{x}) = \lambda_{21} P(\omega_1 | \mathbf{x}) = 0.818$$

由于  $R(\alpha_1 | \mathbf{x}) > R(\alpha_2 | \mathbf{x})$

即决策为  $\omega_2$  的条件风险小于决策为  $\omega_1$  的条件风险，因此我们采取决策行动  $\alpha_2$ ，即判断待识别的细胞  $\mathbf{x}$  为  $\omega_2$  类——异常细胞。

将 1 与 2 相对比，其分类结果正好相反，这是因为这里影响决策结果的因素又多了一个，即“损失”。而且两类错误决策所造成的损失相差很悬殊，因此“损失”就起了主导作用。

十九、有两类样本集

$$\mathbf{x}_1^1 = [0, 0, 0]^T, \quad \mathbf{x}_1^2 = [1, 0, 0]^T, \quad \mathbf{x}_1^3 = [1, 0, 1]^T, \quad \mathbf{x}_1^4 = [1, 1, 0]^T$$

$$\mathbf{x}_2^1 = [0, 0, 1]^T, \quad \mathbf{x}_2^2 = [0, 1, 0]^T, \quad \mathbf{x}_2^3 = [0, 1, 1]^T, \quad \mathbf{x}_2^4 = [1, 1, 1]^T$$

1. 用 K-L 变换求其二维特征空间，并求出其特征空间的坐标轴；
2. 使用 Fisher 线性判别方法给出这两类样本的分类面。

$$(1) \quad E[\mathbf{x}\mathbf{x}^T] = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}, \text{其对应的本征值和本征向量为:}$$

$$\Lambda = \begin{bmatrix} \frac{1}{4} & & \\ & \frac{1}{4} & \\ & & 1 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} 0 & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

$$\text{降到二维空间取} \mathbf{U} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

$$\text{对应的坐标} \quad \omega_1: \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{2}{\sqrt{6}} \\ \frac{-2}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} \frac{2\sqrt{2}}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} \frac{2\sqrt{2}}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \end{bmatrix} \quad \omega_2: \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} \frac{2\sqrt{2}}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} \frac{3\sqrt{2}}{\sqrt{6}} \\ 0 \end{bmatrix}$$

$$(2) \quad \mathbf{m}_1 = \begin{bmatrix} \frac{3}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}, \quad \mathbf{m}_2 = \begin{bmatrix} \frac{1}{4} \\ \frac{3}{4} \\ \frac{3}{4} \\ \frac{1}{4} \end{bmatrix}, \quad S_w = \frac{1}{12} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

$$\mathbf{w}^* = S_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2) = \begin{bmatrix} 6 \\ -6 \\ -6 \end{bmatrix}, \quad y_0 = \frac{\mathbf{w}^{*T}\mathbf{m}_1 + \mathbf{w}^{*T}\mathbf{m}_2}{2} = -3$$

$$\text{所以判别函数为: } g(\mathbf{x}) = \begin{bmatrix} 6 \\ -6 \\ -6 \end{bmatrix}^T \mathbf{x} + 3$$