哈爾濱工業大學（深圳）
HARBIN INSTITUTE OF TECHNOLOGY, SHENZHEN

# 《模式识别》课程

# 实验报告

# （实验二：基于 KL 变换的特征提取）


姓　　名：杨承翰

学　　号：210210226

班　　级：通信 2 班

指导教师：顾术实

哈尔滨工业大学（深圳）

# 实践环节评分及标准

| 考察项目 | 考察内容 | 教师评价 | 备注（评语） |
|---|---|---|---|
| A 预习（可选）100 分 | 预习认真，数据记录表格设计全面、准确（90-100 分） | | |
| | 预习较认真，数据记录表格设计较为全面（80-89 分） | | |
| | 预习深度不够，数据记录表格设计不全面（60-79 分） | | |
| | 未充分预习（60 分以下） | | |
| B 过程表现（必选）100 分 | 态度认真、积极主动、操作规范，实验完毕后仪器断电、归位（90-100 分） | | |
| | 态度认真、积极主动、操作正确，实验完毕后仪器断电、归位（80-89 分） | | |
| | 态度较认真、操作基本正确，实验完毕后仪器断电、归位（60-79 分） | | |
| | 态度不端正、操作不规范，实验完毕后仪器未断电、归位（60 分以下） | | |
| C 结果验收（必选）100 分 | 实验数据处理准确规范、结果合理、结论正确（90-100 分） | | |
| | 实验数据处理准确、结果合理、结论正确（80-89 分） | | |
| | 实验数据处理较准确、结果基本合理、结论基本正确（60-79 分） | | |
| | 有大量错误，未达到实验目的（60 分以下） | | |
| D 答辩（可选）100 分 | 理解充分，表达清晰，问题回答全面准确（90-100 分） | | |
| | 理解较充分，表达清晰，问题回答正确（80-89 分） | | |
| | 理解较充分，表达基本清晰，问题回答基本正确（60-79 分） | | |
| | 表达模糊，问题回答不正确（60 分以下） | | |
| E 实验报告内容（必选）100 分 | 格式规范；内容充实；分析深刻，见解独到（90-100 分） | | |
| | 格式规范；内容完整；分析正确，有个人见解（80-89 分） | | |
| | 格式较规范；内容较完整；分析基本正确（60-79 分） | | |
| | 格式不规范；内容不完整（60 分以下） | | |
| 综合评分 | $\lambda_1A+\lambda_2B+\lambda_3C+\lambda_4D+\lambda_5E$<br>式中 $\Sigma\lambda_i=1$ | | |
| 教师评语 | 备注：比较高分数和比较低分数的实验报告必须要给出评语，指明给分的具体依据（此条打印时请删除） | | |

## 1、实验目的

（1）掌握 KL 变换的求取方法
（2）了解特征提取的目的
（3）掌握基于利用 KL 变换进行特征提取的原理

## 2 实验内容及要求

1、利用**类均值向量法对两组数据提取特征**，选取最优的投影将数据降到 1 维，考察投影后样本分布情况，并利用 Fisher 准则一节中阈值求取方法求阈值用该投影方向进行分类。

2、利用**基于类平均向量中判别信息的最优压缩方法**对两组数据提取特征，选取最优的投影，考察投影后样本分布情况，并利用 Fisher 准则一节中阈值求取方法求阈值用该投影方向进行分类。

数据：使用公开数据集 **Iris** 数据中的前两类数据处理，为了图形显示方便，建议只用前两个特征处理。

Iris 数据介绍：

▪UcI(国际上常用的标准测试数据集)中的 Iris(鸢尾属植物)数据，用于聚类分析。Iris 鸢尾花数据集是一个经典数据集，在统计学习和机器学习领域都经常被用作示例。

▪数据集内包含 3 类（iris-setosa 山鸢尾, iris-versicolour 变色鸢尾, iris-virginica 维吉尼亚鸢尾）共 150 条记录，每类各 50 个数据。其中的一个种类与另外两个种类是线性可分离的，后两个种类是非线性可分离的。

▪每朵鸢尾花都用四个特征进行描述：萼片长度(cm)，萼片宽度(cm)，花瓣长度(cm)，花瓣宽度(cm)

▪（load iris.dat ）； X 有 150 列。这些列代表 150 组鸢尾花属性。

▪它具有 5 行，前 4 行表示四个特征值。

▪第 5 个则是类别属性。用 1，2，3 表示

**变色鸢尾　　　　　　　　　　山鸢尾　　　　　　　　　　维吉尼亚鸢尾**

## 3、实验原理

**1、利用类均值向量提取特征**

类条件均值 $u_i$ 包含大量的分类信息。为了降低维数，且尽可能地保持原有特征的分类信息，应选择一种变换，使得变换后的类条件均值向量的分量比其他的变换保持更多的分类信息。

基于欧氏距离特征提取：判据是从使类内尽可能密集，类间尽可能分开的思想出发的。但是，各类均值向量各分量的分类性能，不仅与取决于各均值向量各分量之间的距离，而且，还和其方差以及各分量的相关程度有关。

**利用类均值向量提取特征算法步骤如下：**

1）　为了估计各分量（特征）对于分类的单独作用，先按总的类内离散度矩阵 $\mathbf{S_w}$ 作为产生矩阵产生相应的 K-L 坐标系统，从而把包含在原向量中各分量的相关性消除或者：

$$\mathbf{US_w} = \mathbf{\Lambda U} \text{ ，这类 } \mathbf{U} = \left[\mathbf{u_1} \quad \mathbf{u_2} \quad \mathbf{L} \quad \mathbf{u_d}\right], \quad \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_d \end{bmatrix}$$

并得到在新坐标系中各分量 $\mathbf{u}_j$ 上的类内离散度： $\mathbf{u}_j^T \mathbf{S_w} \mathbf{u}_j = \lambda_j$

2）　然后对均值向量在这些新坐标中分离的程度作出判断，决定在各坐标轴分量均值向量所能提供的相对可分性信息。据此选取 K-L 变换的基向量作为特征提取变换矩阵。

计算在 $u_j$ 轴上的总的类间离散度 ： $\mathbf{u}_j^T \mathbf{S_b} \mathbf{u}_j = S_b^j$，

这里，$\mathbf{S}_b = \sum_{i=1}^{c} P(\omega_i)(\mathbf{u}_i - \mathbf{u})(\mathbf{u}_i - \mathbf{u})^T$

3）　计算判据在 $u_j$ 坐标上的判据:

$$J(X_i)=\frac{S_b^j}{S_w^j}=\frac{u_j^T S_b u_j}{u_j^T S_w u_j}=\frac{u_j^T S_b u_j}{\lambda_j}$$

$J(X_i)$越大，表明在新坐标系中该坐标抽包含较多可分性信息。

4） 为了降低特征空间的维数，可以将各分量按判据大小重新排列，使

$$J(X_1)\geq J(X_2)\geq J(X_3)\ \ldots\geq J(X_D)$$

取与前面 d 个最大的 $J(X_i)\geq$值相对应的本征向量 $\mathbf{u}_j$，$j=1,...,d$; 作为特征空间的基向量。

$$\mathbf{W=[u_1,u_2,\ldots,u_d]}$$

**2、基于类平均向量中判别信息的最优压缩**

这是一种充分利用类均值向量所包含的判别信息的方法，这种方法分成两步：

1）白化处理

2）特征提取

**具体算法如下：**

- 先用原坐标系中 Sw 作为产生矩阵： $\mathbf{U^T S_w U = \Lambda}$
- 令: $\mathbf{B=U\Lambda^{-1/2}}$
- 计算经过 B 变换后的类间离散度矩阵 **S'b**：$\mathbf{S'_b=B^T S_b B}$

（前 3 步是白化过程）

- 以 $S'_b$ 作为产生矩阵，作第二次 K-L 变换

$$\mathbf{V^T S_b' = \Lambda' V}$$

- 选择最大 d 个本征值对应的 d （≤c-1）个本征向量

  $\mathbf{V'=[v1,v2,\ldots,v_d]}$
- 计算最终的变换矩阵：$\mathbf{W=BV'=U\Lambda^{-1/2}V'}$

这种方法主要用在类别数 C 比原特征向量的维数 D 小得多的情况，由于 $\mathbf{S_b}$ 的秩最多为 C-1，因此可使特征维数降至 C 维以下。

**3、分类**

**数据**降维后，可利用线性分类器等设计方法设计分类器，对数据进行分类处理。

本实验是将数据降到 1 维，因此，得到的是一维上的投影数据，求出来了投影方向，即线性分类器的权向量 **W***。若按线性分类器分类，还需求取阈值权，阈值权可以用 Fisher 准则一节中的方法，如：

阈值权 $W_0$ 确定 ：$W_0=-\dfrac{m_1^{'}+m_2^{'}}{2}$；

这里： $m_i^{'}=\mathbf{W}^{*T}m_i \quad i=1,2$，是投影后两类的均值。

## 4、 实验结果及分析

# 原理概述

基于 KL 变换的特征提取算法是一种常用的线性特征提取方法，用于解决分类问题。它通过最大化类间散度和最小化类内散度的方式，找到最佳的投影方向，从而实现有效的特征提取。

算法原理如下：

1. 数据预处理：假设有 N 个样本，每个样本有 D 维特征。首先将数据集分为 C 个类别，计算每个类别的均值向量 μc 和协方差矩阵 Σc。

2. 计算类内散度矩阵 Sw：Sw 衡量了同一类别内样本的散布程度，可以通过以下公式计算：

$$Sw = \Sigma c = \sum(Xi - \mu c)(Xi - \mu c)'$$

3. 计算类间散度矩阵 Sb：Sb 衡量了不同类别之间样本的差异性，可以通过以下公式计算：

$$Sb = \Sigma(\mu c - \mu)(\mu c - \mu)'$$

其中，μ 为所有样本的均值向量，μ_c 为每个类别的均值向量。

4. 求解广义特征值问题：通过求解广义特征值问题，即求解 Sw^(-1)Sb 的特征值和特征向量，得到投影方向。

5. 特征选择：根据特征值的大小，选择前 k 个特征向量作为最佳投影方向。

6. 特征提取：将原始数据投影到最佳投影方向上，得到新的特征表示。

7. 分类：使用新的特征表示进行分类任务，例如使用最近邻分类器或支持向量机等。

基于 KL 变换的特征提取算法通过最大化类间散度和最小化类内散度，寻找到在投影空间中能够最好地区分不同类别的投影方向。这样可以有效地降低维度，并提取出具有判别性的特征，从而改善分类性能。

类均值向量提取特征是一种常用的特征提取方法，其原理如下：

数据集准备：首先，我们需要收集一组已经标记好的训练样本数据，每个样本都属于某个已知的类别。例如，如果我们要识别手写数字，我们可以收集一组已经标记好的手写数字图像。

特征提取：对于每个样本，我们需要从原始数据中提取特征。类均值向量特征提取方法假设每个类别的样本数据分布在特征空间中近似于一个多元正态分布。因此，我们可以计算每个类别的样本数据的均值向量，作为该类别的特征表示。均值向量是一个包含了各个特征维度上的平均值的向量。

特征表示：对于每个样本，我们将其原始数据映射到特征空间中，并用类均值向量表示该样本。这可以通过计算原始数据与每个类别的均值向量之间的距离来实现。通常，欧氏距离或马哈拉诺比斯距离被用来度量样本与均值向量之间的距离。

分类决策：在特征表示完成后，我们可以使用分类算法（如最近邻分类器）对新的未知样本进行分类。具体地，我们计算未知样本与每个类别的均值向量之间的距离，并将其归类到距离最近的类别。

通过类均值向量提取特征，我们能够将原始数据转化为更具代表性的特征向量，并且能够通过计算距离来度量样本之间的相似性。这种方法在许多模式识别任务中都得到了广泛应用，例如人脸识别、手写字体识别等。

基于类平均向量的最优压缩原理是一种特征选择方法，旨在从原始特征集中选择出最具有判别性的特征子集，以减少特征的维度并保留最重要的信息。

该方法的原理如下：

1. 数据集准备：首先，我们需要收集一组已经标记好的训练样本数据，每个样本都属于某个已知的类别。

2. 特征提取：对于每个样本，我们从原始数据中提取特征。这可以是任何合适的特征提取方法，如常用的统计特征、频域特征或图像纹理特征等。

3. 类平均向量计算：对于每个类别，我们计算该类别下所有样本的特征均值向量。这个均值向量代表了该类别的平均特征。

4. 类内散布矩阵计算：对于每个类别，我们计算该类别下所有样本的协方差矩阵，并将它们相加得到类内散布矩阵。类内散布矩阵反映了同一类别内部样本的变化情况。

5. 类间散布矩阵计算：我们计算所有类别的平均特征向量的协方差矩阵，称之为类间散布矩阵。类间散布矩阵反映了不同类别之间的差异程度。

6. 特征选择：最优压缩的目标是选择一组特征，使得类内散布矩阵尽可能小，而类间散布矩阵尽可能大。这可以通过计算特征子空间的投影向量或线性变换来实现。通常，我们可以通过求解广义瑞利商的最大化问题来获得最优的投影向量或线性变换。

7. 分类决策：在特征选择完成后，我们可以使用分类算法（如最近邻分类器）对新的未知样本进行分类。具体地，我们将未知样本投影到所选特征子空间中，并使用相应的分类算法进行分类。

基于类平均向量的最优压缩原理能够选择出最具有判别性的特征子集，从而减少特征的维度并保留最重要的信息。这种方法在模式识别和机器学习任务中常被用于降低维度、提高分类性能和减少计算复杂度。

# 代码解析

该程序是基于 KL 变换的特征提取，用于鸢尾花数据集的分类问题。程序首先加载数据集，将数据集分为三个类别，然后对数据进行预处理，计算协方差矩阵和均值向量等统计量。接着，程序通过计算类内散度矩阵和类间散度矩阵，求出投影方向，从而实现特征提取。程序最后通过绘图展示样本在原始 2 维空间和新的投影空间上的分布情况，并计算可分性指标 J 和错误率，用于评估特征提取效果。

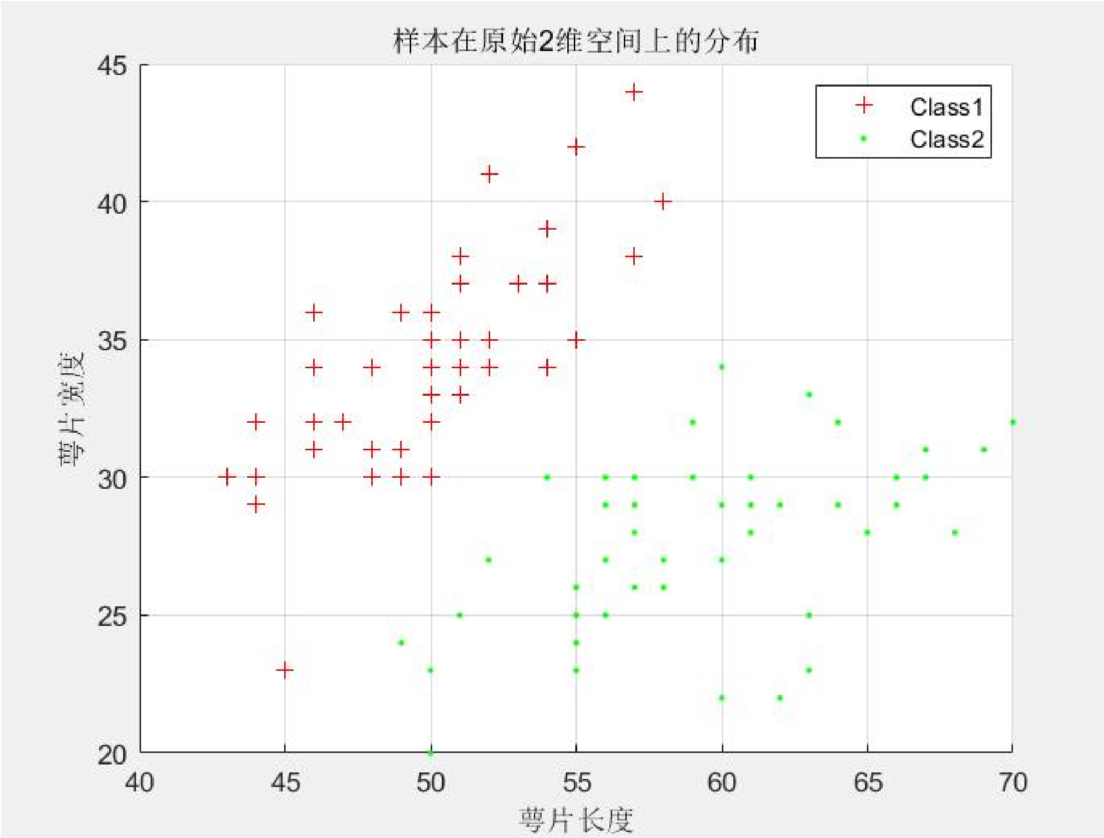
具体来说，程序的主要步骤如下：

1. 加载数据集 iris.dat，并将数据集分为三个类别 setosa、versicolor 和 virginica。

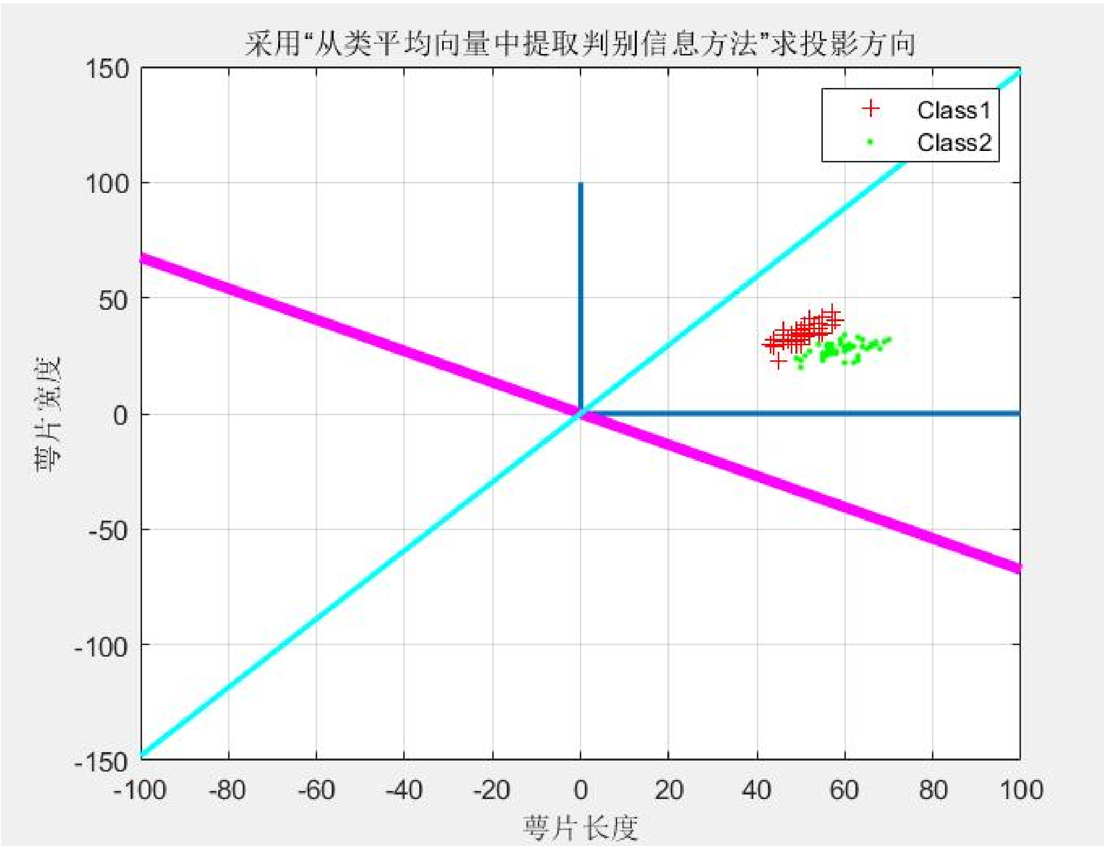
2. 对数据集进行预处理，计算协方差矩阵 sigma1 和 sigma2，均值向量 mu1 和 mu2，以及样本数 n1 和 n2 等统计量。

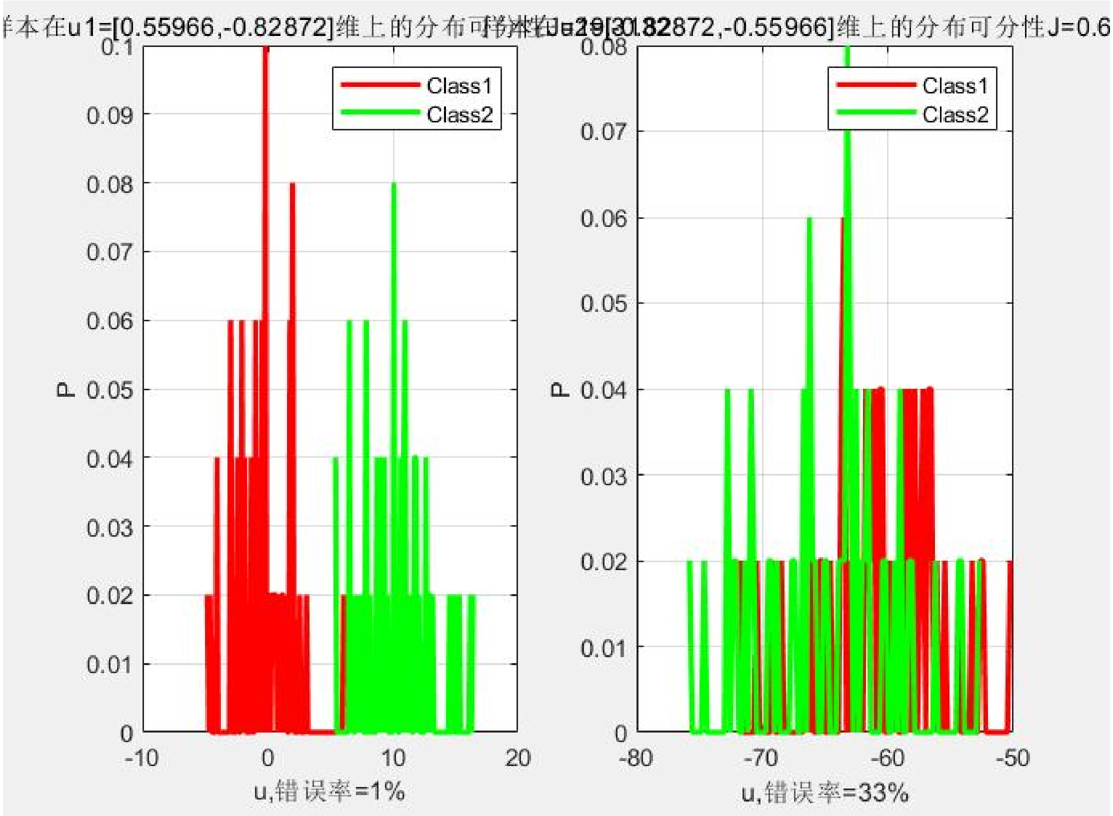
3. 计算类内散度矩阵 Sw 和类间散度矩阵 Sb，并通过求解广义特征值问题，得到投影方向 U。

4. 采用“从类平均向量中提取判别信息方法”求投影方向，绘制样本在原始 2 维空间和新的投影空间上的分布情况，并计算可分性指标 J1 和错误率。

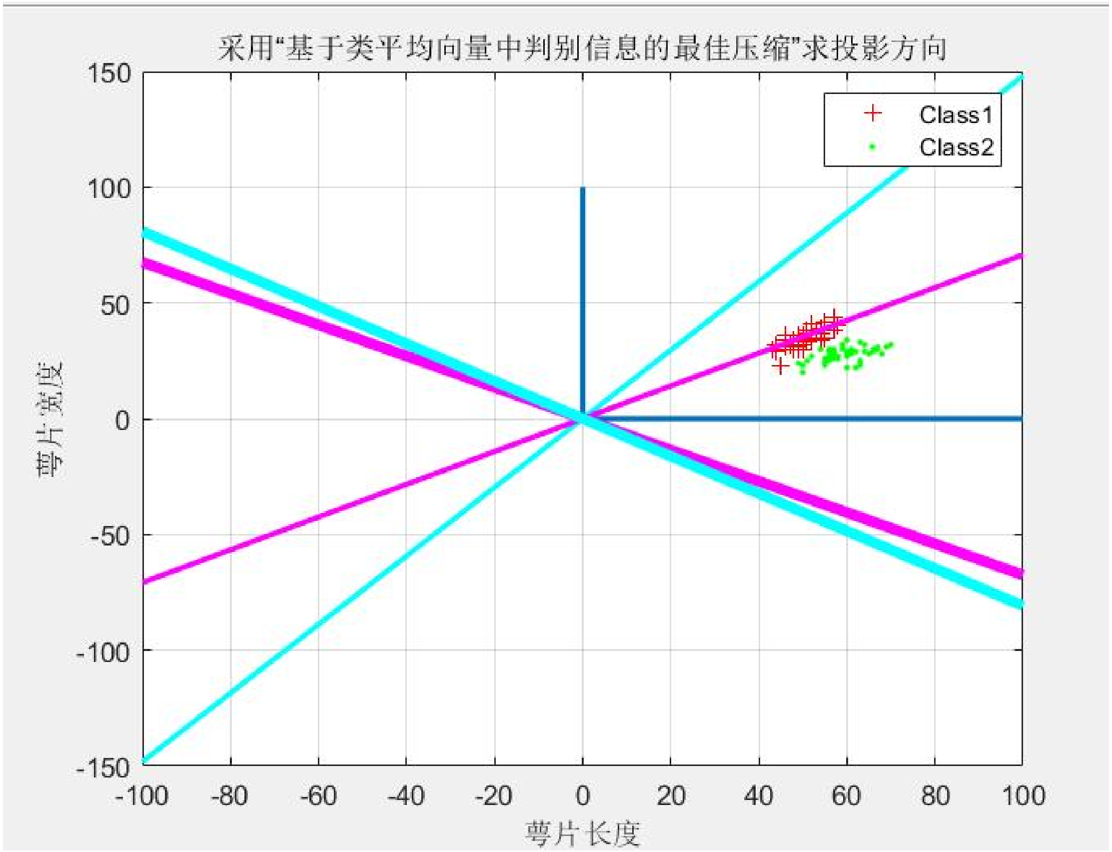


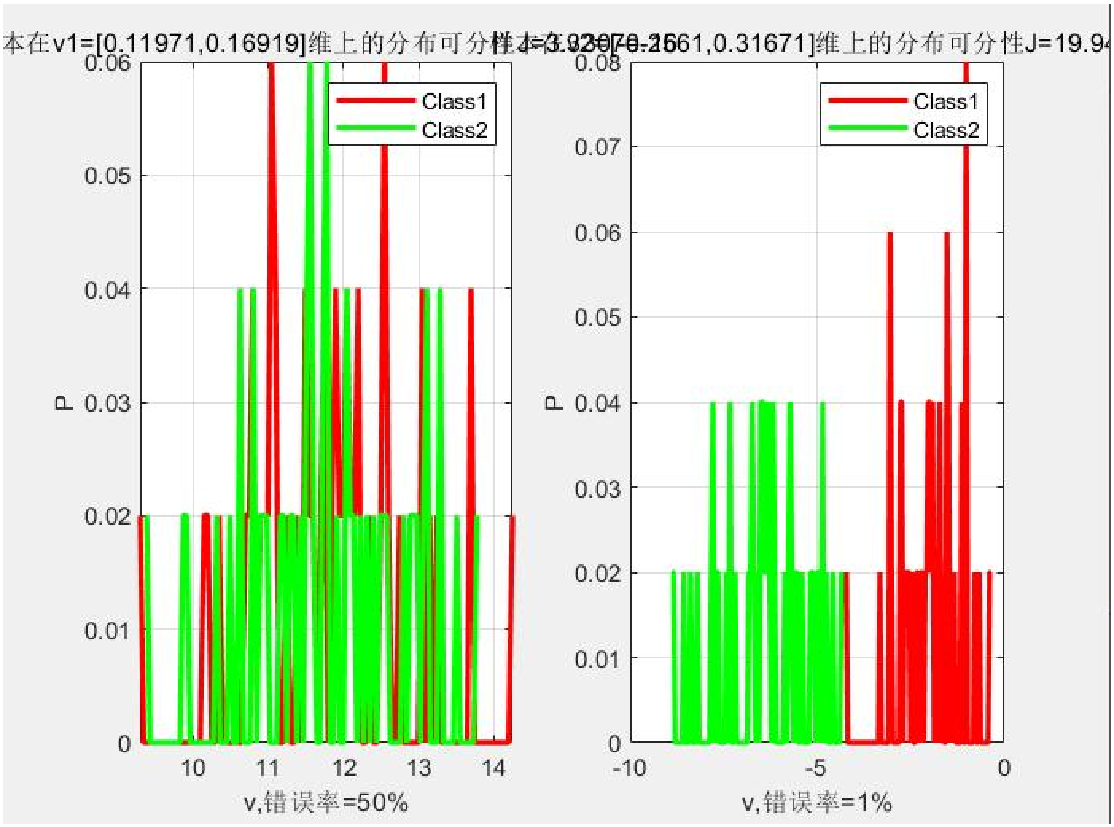

5. 采用“基于类平均向量中判别信息的最佳压缩”求投影方向，绘制样本在原始 2 维空间和新的投影空间上的分布情况，并计算可分性指标 J2 和错误率。






6. 最后，程序通过绘图展示了样本在原始 2 维空间和新的投影空间上的分布情况，并计算了可分性指标 J 和错误率，用于评估特征提取效果。






## 下面逐部解释各部分的作用：

1. `clc`、`close all`、`clear all`：清空命令窗口、关闭所有图形窗口、清除工作区变量。

2. `load iris.dat`：读入 iris 数据集。

3.`setosa=iris((iris(:,5)==1),:);versicolor=iris((iris(:,5)==2),:);virginica=iris((iris(:,5)==3),:);`：按照鸢尾花数据集中的类别将数据分为三类。

4. `obsv_n=size(iris,1); fata=iris(:,1:4);`：获取样本数和前四列的特征值。

5. `X=setosa(:,1:2);Y=versicolor(:,1:2);`：选取第一和第二列特征值作为两个类别的样本。

6.`figure(1);scatter(X(:,1),X(:,2),'r+');holdon;gridon;scatter(Y(:,1),Y(:,2),'g.');xlabel('萼片长度');ylabel('萼片宽度');h=legend('Class1','Class2');title('样本在原始 2 维空间上的分布');`：绘制数据点散点图，观察数据在原始二维空间的分布。

7.`sigma1=cov(X);sigma2=cov(Y);mu1=mean(X);mu2=mean(Y);P1=0.5;P2=0.5;n1=50;n2=50;Sw=sigma1*P1+P2*sigma2;Sb=(mu1-mu2)'*(mu1-mu2);`：计算两个类别的协方差矩阵、均值向量、先验概率和样本数，并计算类间散度矩阵和类内散度矩阵。

8.`[U,D]=eig(Sw);forii=1:2lambda(ii)=D(ii,ii);J1(ii)=U(:,ii)'*Sb*U(:,ii)/lambda(ii);end`：通过求解 Sw 的特征向量和特征值，计算投影方向，并计算每个方向的可分性指标 J1。

9. `figure(2);plot(X(:,1),X(:,2),'r+');hold on; grid on;plot(Y(:,1),Y(:,2),'g.');xlabel('萼片长度');ylabel('萼片宽度');line([0,100],[0,0],'LineWidth',2);line([0,0],[0,100],'LineWidth',2);u1=U(:,1);u10=100;u11=u1(1)/u1(2)*u10;line([-u10,u10],[-u11,u11],'Color','m','LineWidth',4);holdon;u2=U(:,

2);u21=100;u20=u2(1)/u2(2)*u21;line([-u21,u21],[-u20,u20],'Color','c','LineWidth',2);hold on;h=legend('Class1','Class2');title('采用“从类平均向量中提取判别信息方法”求投影方向');`：绘制数据点散点图和投影方向，观察投影后数据的分布情况。

10.`B=U*D^(-0.5);SB1=B'*Sb*B;[V,D1]=eig(SB1);W=B*V;forii=1:2v(:,ii)=W(:,ii);J2(ii)=D1(ii,ii);end`：通过 KL 变换求解投影矩阵 W，并计算每个方向的可分性指标 J2。

11. `figure(3);plot(X(:,1),X(:,2),'r+');hold on; grid on;plot(Y(:,1),Y(:,2),'g.');xlabel('萼片长度');ylabel('萼片宽度');line([0,100],[0,0],'LineWidth',2);line([0,0],[0,100],'LineWidth',2);u1=U(:,1);u10=100;u11=u1(1)/u1(2)*u10;line([-u10,u10],[-u11,u11],'Color','m','LineWidth',4);holdon;u2=U(:,2);u21=100;u20=u2(1)/u2(2)*u21;line([-u21,u21],[-u20,u20],'Color','c','LineWidth',2);holdon;v1=W(:,1);v10=100;v11=v1(1)/v1(2)*v10;line([-v10,v10],[-v11,v11],'Color','m','LineWidth',2);holdon;v2=W(:,2);v21=100;v20=v2(1)/v2(2)*v21;line([-v21,v21],[-v20,v20],'Color','c','LineWidth',4);hold on;h=legend('Class1','Class2');title('采用“基于类平均向量中判别信息的最佳压缩”求投影方向');`：绘制数据点散点图和投影方向，观察投影后数据的分布情况。

12.`CC1(1,:)='u1';CC1(2,:)='u2';forK0=1:2u1=U(:,K0);XX1=X*u1;YY1=Y*u1;[p1,x1]=hist(XX1,100);[p2,x2]=hist(YY1,100);T0=(mean(YY1)+mean(XX1))/2;Sym=sign(mean(XX1)-mean(YY1));Ne1=0;Ne1=0;fork=1:n1ifSym*XX1(k)>=Sym*T0elseNe1=Ne1+1;endendNe2=0;fork=1:n2ifSym*YY1(k)<Sym*T0elseNe2=Ne2+1;endender1=Ne1/n1;er2=Ne2/n2;er=(er1*n1/(n1+n2)+er2*n2/(n1+n2))*100;figure(4);subplot(1,2,K0);plot(x1,p1/n1,'r','LineWidth',2);hold on;plot(x2,p2/n2,'g','LineWidth',2);grid on;title(['样本在',CC1(K0,:),'=[',num2str(u1(1)),',',num2str(u1(2)),']','维上的分布',' 可分性J=',num2str(J1(K0))]);xlabel([CC1(K0,:),' 错误率=',num2str(er),'%']);ylabel('P');h=legend('Class1','Class2');end`：计算每个投影方向的分类效果，并绘制数据在该方向上的分布图。

13.`CC2(1,:)='v1';CC2(2,:)='v2';forK0=1:2v1=W(:,K0);XXX1=X*v1;YYY1=Y*v1;[pp1,xx1]=hist(XXX1,100);[pp2,xx2]=hist(YYY1,100);T0=(mean(YYY1)+mean(XXX1))/2;Sym=sign(mean(XXX1)-mean(YYY1));Ne1=0;Ne1=0;fork=1:n1ifSym*XXX1(k)>=Sym*T0elseNe1=Ne1+1;endendNe2=0;fork=1:n2ifSym*YYY1(k)<Sym*T0elseNe2=Ne2+1;endender1=Ne1/n1;er2=Ne2/n2;er=(er1*n1/(n1+n2)+er2*n2/(n1+n2))*100;figure(5);subplot(1,2,K0);plot(xx1,pp1/n1,'r','LineWidth',2);holdon;plot(xx2,pp2/n2,'g','LineWidth',2);grid on;title(['样本在',CC2(K0,:),'=[',num2str(v1(1)),',',num2str(v1(2)),']','维上的分布',' 可分性 J=',num2str(J2(K0))]);xlabel([CC2(K0,:),' 错误率=',num2str(er),'%']);ylabel('P');h=legend('Class1','Class2');end`：计算每个投影方向的分类效果，并绘制数据在该方向上的分布图。

## 5、源程序

```
    clc
close all;clear all;
load iris.dat;
setosa = iris((iris(:,5)==1),:);
versicolor=iris((iris(:,5)==2),:);
```

```
virginica=iris((iris(:,5)==3),:);
obsv_n=size(iris,1);
fata=iris(:,1:4);
X=setosa(:,1:2);
Y=versicolor(:,1:2);
X
Y
figure(1);
scatter(X(:,1),X(:,2),'r+');hold on; grid on;
scatter(Y(:,1),Y(:,2),'g.');
xlabel('萼片长度');ylabel('萼片宽度');
h=legend('Class1','Class2');
title('样本在原始 2 维空间上的分布');
sigma1=cov(X);sigma2=cov(Y);
mu1=mean(X);mu2=mean(Y);
P1=0.5;P2=0.5;n1=50;n2=50;
Sw=sigma1*P1+P2*sigma2;
Sb=(mu1-mu2)'*(mu1-mu2);
Sw
Sb
[U,D]=eig(Sw);
U
D
for ii=1:2
lambda(ii)=D(ii,ii);
J1(ii)=U(:,ii)'*Sb*U(:,ii)/lambda(ii);
end
J1
figure(2);
plot(X(:,1),X(:,2),'r+');hold on; grid on;
plot(Y(:,1),Y(:,2),'g.');
xlabel('萼片长度');ylabel('萼片宽度');
line([0,100],[0,0],'LineWidth',2);
line([0,0],[0,100],'LineWidth',2);
u1=U(:,1);
u10=100;u11=u1(1)/u1(2)*u10;
line([-u10,u10],[-u11,u11],'Color','m','LineWidth',4);hold on;
u2=U(:,2);
u21=100;u20=u2(1)/u2(2)*u21;
line([-u21,u21],[-u20,u20],'Color','c','LineWidth',2);hold on;
h=legend('Class1','Class2');
title('采用“从类平均向量中提取判别信息方法”求投影方向');
B=U*D^(-0.5);
SB1=B'*Sb*B;
```

```
[V,D1]=eig(SB1);
W=B*V;
W
for ii=1:2
v(:,ii)=W(:,ii);
J2(ii)=D1(ii,ii);
end
J2
figure(3);
plot(X(:,1),X(:,2),'r+');hold on; grid on;
plot(Y(:,1),Y(:,2),'g.');
xlabel('萼片长度');ylabel('萼片宽度');
line([0,100],[0,0],'LineWidth',2);
line([0,0],[0,100],'LineWidth',2);
u1=U(:,1);
u10=100;u11=u1(1)/u1(2)*u10;
line([-u10,u10],[-u11,u11],'Color','m','LineWidth',4);hold on;
u2=U(:,2);
u21=100;u20=u2(1)/u2(2)*u21;
line([-u21,u21],[-u20,u20],'Color','c','LineWidth',2);hold on;
v1=W(:,1);
v10=100;v11=v1(1)/v1(2)*v10;
line([-v10,v10],[-v11,v11],'Color','m','LineWidth',2);hold on;
v2=W(:,2);
v21=100;v20=v2(1)/v2(2)*v21;
line([-v21,v21],[-v20,v20],'Color','c','LineWidth',4);hold on;
h=legend('Class1','Class2');
title('采用“基于类平均向量中判别信息的最佳压缩”求投影方向');
CC1(1,:)='u1';
CC1(2,:)='u2';
for K0=1:2
u1=U(:,K0);
XX1=X*u1;
YY1=Y*u1;
[p1,x1]=hist(XX1,100); [p2,x2]=hist(YY1,100);
T0=(mean(YY1)+mean(XX1))/2;
Sym=sign(mean(XX1)-mean(YY1));
Ne1=0;
Ne1=0;
for k=1:n1
if Sym*XX1(k)>=Sym*T0
else
Ne1=Ne1+1;
end
```

```
end
Ne2=0;
for k=1:n2
if Sym*YY1(k)<Sym*T0
else
Ne2=Ne2+1;
end
end
er1=Ne1/n1;
er2=Ne2/n2;
er=(er1*n1/(n1+n2)+er2*n2/(n1+n2))*100;
figure(4);
subplot(1,2,K0);
plot(x1,p1/n1,'r','LineWidth',2);hold on;
plot(x2,p2/n2,'g','LineWidth',2);grid on;
title(['样本在',CC1(K0,:),'=[',num2str(u1(1)),',',num2str(u1(2)),']','维
上的分布','可分性 J=',num2str(J1(K0))]);
xlabel([CC1(K0),',错误率=',num2str(er),'%']);ylabel('P');
h=legend('Class1','Class2');
end
CC2(1,:)='v1';
CC2(2,:)='v2';
for K0=1:2
v1=W(:,K0);
XXX1=X*v1;
YYY1=Y*v1;
[pp1,xx1]=hist(XXX1,100); [pp2,xx2]=hist(YYY1,100);
T0=(mean(YYY1)+mean(XXX1))/2;
Sym=sign(mean(XXX1)-mean(YYY1));
Ne1=0;
Ne1=0;
for k=1:n1
if Sym*XXX1(k)>=Sym*T0
else
Ne1=Ne1+1;
end
end
Ne2=0;
for k=1:n2
if Sym*YYY1(k)<Sym*T0
else
Ne2=Ne2+1;
end
end
```

```
er1=Ne1/n1;
er2=Ne2/n2;
er=(er1*n1/(n1+n2)+er2*n2/(n1+n2))*100;
figure(5);
subplot(1,2,K0);
plot(xx1,pp1/n1,'r','LineWidth',2);hold on;
plot(xx2,pp2/n2,'g','LineWidth',2);grid on;
title(['样本在',CC2(K0,:),'=[',num2str(v1(1)),',',num2str(v1(2)),']','维上的分布','可分性 J=',num2str(J2(K0))]);
xlabel([CC2(K0),',错误率=',num2str(er),'%']);ylabel('P');
h=legend('Class1','Class2');
end
```