第二问

*%% 数据导入及处理*

clc, clear all, close all

A=xlsread('理化指标.xls','葡萄酒指标汇总', 'C3:J29');*% 红葡萄酒*

*%A=xlsread('理化指标.xls','葡萄酒指标汇总', 'C33:J60');% 白葡萄酒*

*% 红葡萄酒质量评分数据*

 B=[65.4    77.15   77.5    69.9    72.7    69.25   68.4    69.15   79.85   71.5    65.85   61.1    71.7    72.8    62.2    72.4    76.9    62.65   75.6    77.51111111 74.65   74.4    81.35   74.75   68.7    72.9    72.25];

*% 白葡萄酒质量评分数据*

*% B=[79.95  75  76.90555556 78.15   76.25   71.95   75.85   71.33333333 76.65   77.05   71.85   67.85   69.9    74.55   75.4    70.65   79.55   74.9    74.3    77.2    77.8    75.2    76.65   74.7    78.3    77.8    70.9    72.2];*

*%  数据标准化处理*

a=size(A,1);

b=size(A,2);

**for** i=1**:**b

    SA(**:**,i)=(A(**:**,i)**-**mean(A(**:**,i)))**/**std(A(**:**,i));

**end**

*%% 计算相关系数矩阵的特征值和特征向量*

CM=corrcoef(SA);  *% 计算相关系数矩阵(correlation matrix)*

[V, D]=eig(CM);  *% 计算特征值和特征向量*

**for** j=1**:**b

    DS(j,1)=D(b**+**1**-**j, b**+**1**-**j); *% 对特征值按降序进行排序*

**end**

**for** i=1**:**b

    DS(i,2)=DS(i,1)**/**sum(DS(**:**,1)); *%贡献率*

    DS(i,3)=sum(DS(1**:**i,1))**/**sum(DS(**:**,1)); *%累积贡献率*

**end**

*%% 选择主成分及对应的特征向量*

T=0.8;  *% 主成分信息保留率.*

**for** K=1**:**b

**if** DS(K,3)**>**=T

        Com\_num=K;

**break**;

**end**

**end**

*% 提取主成分对应的特征向量*

**for** j=1**:**Com\_num

    PV(**:**,j)=V(**:**,b**+**1**-**j);

**end**

*%%  计算各评价对象的主成分得分*

new\_score=SA**\***PV;

**for** i=1**:**a

    total\_score(i,1)=sum(new\_score(i,**:**));

    total\_score(i,2)=i;

**end**

result\_report=[new\_score, total\_score]; *% 将各主成分得分与总分放在同一个矩阵中*

result\_report=sortrows(result\_report,(K**+**2)); *% 按总分降序排序*

*%% 用k-Means聚类法确定最佳的聚类数*

A=result\_report(**:**,(K**+**1));

A=[A,**B'**];

X=A;

numC=7;

**for** i=1**:**numC

    kidx = kmeans(X,i);

    silh = silhouette(X,kidx); *%计算轮廓值*

    silh\_m(i) = mean(silh);    *%计算平均轮廓值*

**end**

figure

plot(1**:**numC,silh\_m,'ko-', 'linewidth',2)

set(gca,'linewidth',2);

xlabel('类别数')

ylabel('平均轮廓值')

title(' 不同类别对应的平均轮廓值')

*% 绘制2至5类时的轮廓值分布图*

figure

set(gca,'linewidth',2);

**for** i=2**:**5

    kidx = kmeans(X,i);

    subplot(2,2,i**-**1);

    [~,h] = silhouette(X,kidx);

    set(gca,'linewidth',2);

    title([num2str(i), '类时的轮廓值 ' ])

    snapnow

    xlabel('轮廓值');

    ylabel('类别数');

**end**

*%% Kmeans聚类及结果报告*

figure

[idx,ctr]=kmeans(A,3,'dist','sqEuclidean');

[m,n]=size(A);

t1=ones(1,n)**\***30;

c1=find(idx**==**1); c2=find(idx**==**2); c3=find(idx**==**3);

h=plot(A(idx**==**1),A(idx**==**1,2),'k--\*', A(idx**==**2),A(idx**==**2,2),'k--s', A(idx**==**3),A(idx**==**3,2),'k--d');

xlabel('主成分得分','fontsize',12);

ylabel('质量评分依据','fontsize',12);

title('红葡萄酒理化指标聚类图','fontsize',12)

set(h, 'MarkerSize',8, 'MarkerFaceColor','k');

set(gca,'linewidth',2) ;

disp('主成分得分(最后1列为样本编号，倒数第2列为总分，前面为各主成分得分)')

result\_report

disp('分类结果：');

disp(['第1类:' ,'中心点：',num2str(ctr(1)),'  ','该类样品编号：', num2str(**c1'**)]);

disp(['第2类:' ,'中心点：',num2str(ctr(2)),'  ','该类样品编号：', num2str(**c2'**)]);

disp(['第3类:' ,'中心点：',num2str(ctr(3)),'  ','该类样品编号：', num2str(**c3'**)]);

第三问

%% 数据导入及处理

*%% 数据导入及处理*

clc, clear all, close all

*%A=xlsread('理化指标.xls','葡萄酒指标汇总', 'B3:J29');% 红葡萄酒*

*%B=xlsread('理化指标.xls','酿酒葡萄指标汇总', 'C3:AF29');% 红葡萄酒*

A=xlsread('理化指标.xls','葡萄酒指标汇总', 'B33:J60');*% 白葡萄酒*

B=xlsread('理化指标.xls','酿酒葡萄指标汇总', 'C34:AF61');*% 白葡萄酒*

B=B(**:**,all(~isnan(B),1));

*%%逐步回归法*

stepwise(B,A(**:**,3)*%可更改因变量*

第四问

%% 数据导入及处理

*%% 数据导入及处理*

clc, clear all, close all

A=xlsread('理化指标.xls','葡萄酒指标汇总', 'C3:J29');*% 红葡萄酒*

*%A=xlsread('理化指标.xls','酿酒葡萄指标汇总', 'C3:AF29');% 红葡萄酒*

*%A=xlsread('理化指标.xls','酿酒葡萄指标汇总', 'C34:AF61');% 白葡萄酒*

*%A=xlsread('理化指标.xls','葡萄酒指标汇总', 'C33:J60');% 白葡萄酒*

*% 红葡萄酒质量评分数据*

 B=[65.4    77.15   77.5    69.9    72.7    69.25   68.4    69.15   79.85   71.5    65.85   61.1    71.7    72.8    62.2    72.4    76.9    62.65   75.6    77.51111111 74.65   74.4    81.35   74.75   68.7    72.9    72.25];

*% 白葡萄酒质量评分数据*

*% B=[79.95  75  76.90555556 78.15   76.25   71.95   75.85   71.33333333 76.65   77.05   71.85   67.85   69.9    74.55   75.4    70.65   79.55   74.9    74.3    77.2    77.8    75.2    76.65   74.7    78.3    77.8    70.9    72.2];*

*%  数据标准化处理*

a=size(A,1);

b=size(A,2);

**for** i=1**:**b

    SA(**:**,i)=(A(**:**,i)**-**mean(A(**:**,i)))**/**std(A(**:**,i));

**end**

*%% 计算相关系数矩阵的特征值和特征向量*

CM=corrcoef(SA);  *% 计算相关系数矩阵(correlation matrix)*

[V, D]=eig(CM);  *% 计算特征值和特征向量*

**for** j=1**:**b

    DS(j,1)=D(b**+**1**-**j, b**+**1**-**j); *% 对特征值按降序进行排序*

**end**

**for** i=1**:**b

    DS(i,2)=DS(i,1)**/**sum(DS(**:**,1)); *%贡献率*

    DS(i,3)=sum(DS(1**:**i,1))**/**sum(DS(**:**,1)); *%累积贡献率*

**end**

*%% 选择主成分及对应的特征向量*

T=0.8;  *% 主成分信息保留率.*

**for** K=1**:**b

**if** DS(K,3)**>**=T

        Com\_num=K;

**break**;

**end**

**end**

*% 提取主成分对应的特征向量*

**for** j=1**:**Com\_num

    PV(**:**,j)=V(**:**,b**+**1**-**j);

**end**

*%%  计算各评价对象的主成分得分*

new\_score=SA**\***PV;

**for** i=1**:**a

    total\_score(i,1)=sum(new\_score(i,**:**));

    total\_score(i,2)=i;

**end**

result\_report=[new\_score, total\_score]; *% 将各主成分得分与总分放在同一个矩阵中*

result\_report=sortrows(result\_report,(K**+**2)); *% 按总分降序排序*

*%%  分析是否相关*

data=[total\_score(**:**,1),**B'**];

corr(data)