530 陈斯杰 电子信息工程 第13次作业

1.

设D_1为绝对值距离矩阵，D_2为平方和距离矩阵，利用MATLAB计算得到：

$$D_1=\begin{bmatrix} 0 & 12 & 5 & 22 & 9 \\ 12 & 0 & 13 & 28 & 13 \\ 5 & 13 & 0 & 21 & 16 \\ 22 & 28 & 21 & 0 & 19 \\ 9 & 13 & 6 & 19 & 0 \end{bmatrix}, D_2=\begin{bmatrix} 0 & 7.0711 & 2.6458 & 16.4317 & 5.0000 \\ 7.0711 & 0 & 8.5440 & 22.4499 & 8.6603 \\ 2.6458 & 8.5440 & 0 & 14.7309 & 5.0990 \\ 16.4317 & 22.4499 & 14.7309 & 0 & 14.3875 \\ 5.0000 & 8.6603 & 5.0990 & 14.3875 & 0 \end{bmatrix}$$

附MATLAB代码如下:

```
%q1
clear all;
clc;
a=[2 4 6 32;
   5 2 5 38;
   3 3 7 30;
   1 2 3 16;
   4 3 2 30];
D1=zeros(5,5);
D2=zeros(5,5);
for i=1:5
    for j=1:5
        for k=1:4
            D1(i,j)=D1(i,j)+abs(a(i,k)-a(j,k));
            D2(i,j)=D2(i,j)+(a(i,k)-a(j,k))^2;
        end
        D2(i,j)=sqrt(D2(i,j));
    end
end
D1,D2
```

(1) 设样品间距离取为欧氏距离，类间的距离采用最短距离法，计算如下：

在matlab中运行如下代码：

```
clc;clear;
x=[1;2;3;6;9;11];
D = pdist(x,'cityblock');
squareform(D)
```

得$D_{(0)}$计算结果为：

	G_1	G_2	G_3	G_4	G_5	G_6
G_1	0					
G_2	1	0				
G_3	2	1	0			
G_4	5	4	3	0		
G_5	8	7	6	3	0	
G_6	10	9	8	5	2	0

$D_{(0)}$中最小元素是$D_{(12)} = D_{(23)} = 1$,将样品G_1、G_2、G_3合并成G_7，构成$D_{(1)}$:

	G_7	G_4	G_5	G_6
G_7	0			
G_4	3	0		
G_5	6	3	0	
G_6	8	5	2	0

$D_{(1)}$中最小元素是$D_{(56)} = 2$,将样品G_5、G_6合并成G_8，构成$D_{(2)}$:

	G_7	G_4	G_8
G_7	0		
G_4	3	0	
G_8	6	3	0

$D_{(2)}$中最小元素是$D_{(74)} = D_{(48)} = 3$,将样品G_7、G_4、G_8合并成G_9，这样六个样品就聚为一类. 运行下列代码可得聚类图：

```
Z1 = linkage(D,'single');
[H,T] = dendrogram(Z1,'colorthreshold','default');
set(H,'LineWidth',2)
grid on;title('最短距离法聚类图')
```

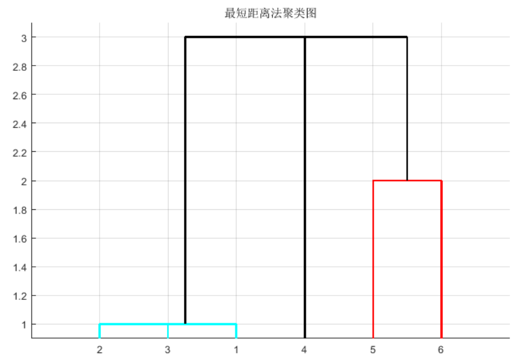


图 1: 聚类图

(2) 重心法，计算如下：由最小距离法中$D_{(0)}$得$D^2_{(2)}$:

	G_1	G_2	G_3	G_4	G_5	G_6
G_1	0					
G_2	1	0				
G_3	4	1	0			
G_4	25	16	9	0		
G_5	64	49	36	9	0	
G_6	100	81	64	25	4	0

$D^2_{(0)}$中最小元素是$D_{(12)^2} = D_{(23)^2} = 1$,将样品$G_1$、$G_2$、$G_3$合并成$G_7$，构成$D^2_{(1)}$,其中$p = 1, q = 2, s = 3$ 和$r = 7$类，$n_r = n_p + n_q + n_s$,

$$\bar{X}_r^2 = \frac{1}{n_r}(n_p)\bar{X}_p + (n_q)\bar{X}_q + (n_s)\bar{X}_s = \frac{1}{3}(1*1+1*2+1*3) = 2$$

$$D^2_{kr} = \frac{1}{3}D^2_{k1} + \frac{1}{3}D^2_{k2}\frac{1}{3}D^2_{k3} - \frac{1}{9}D^2_{12} - \frac{1}{9}D^2_{13} - \frac{1}{9}D^2_{23}, k = 4,5,6$$

所以得$D^2_{(1)}$:

	G_7	G_4	G_5	G_6
G_7	0			
G_4	16	0		
G_5	49	9	0	
G_6	121.5	25	4	0

$D_{(1)}$中最小元素是$D_{(56)} = 4$,将样品G_5、G_6合并成G_8，同$D_{(1)}$计算得$D_{(2)}$:

	G_7	G_4	G_8
G_7	0		
G_4	16	0	
G_8	83.25	15	0

$D_{(2)}$中最小元素是$D_{(48)} = 15$,将样品G_4、G_8合并成G_9，最后G_7、G_9并为一类，这样六个样品就聚为一类.

3.我们采用欧式距离来计算该五个向量之间的距离，经由R运行如下代码：

```
x_3 <- matrix(c(1,1,6,8,8,1,2,3,2,0),nrow = 5,ncol
    = 2,byrow = FALSE)
dist(x_3,p=2)

#输出结果为:
  1         2         3         4
2 1.000000
3 5.385165  5.099020
4 7.071068  7.000000  2.236068
5 7.071068  7.280110  3.605551  2.000000
```

整理得到一下表格：

	1	2	3	4	5
1	0	$\boxed{1}$	5.382	7.071	7.071
2		0	5.099	7.000	7.280
3			0	2.236	3.606
4				0	2.000
5					0

Step1：选取上三角中最小的数字，合并$\{1\}\{2\}$;

	3	4	5	$6(\{1,2\})$
3	0	2.236	3.606	5.099
4		0	$\boxed{2}$	7
5			0	7.071
6				0

Step2：合并$\{4\}\{5\}$;

	3	$6(\{1,2\})$	$7(\{4,5\})$
3	0	5.099	$\boxed{2.236}$
6		0	7
7			0

Step3：合并$\{3\}\{7\}$;

	$6(\{1,2\})$	$8(\{3,4,5\})$
3	0	$\boxed{5.099}$
8		0

所以按两类分应分为：$\{1,2\},\{3,4,5\}$;

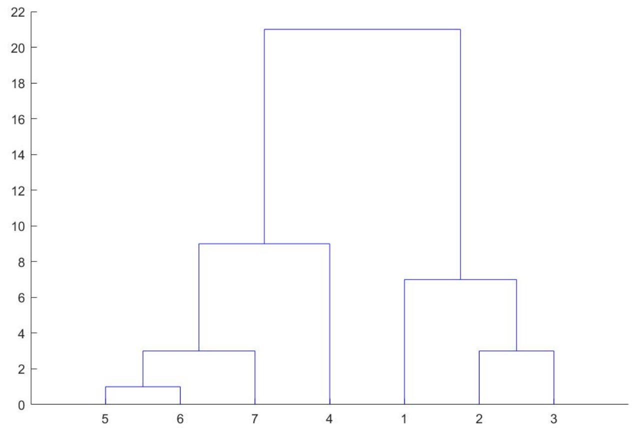
按3类分应分为：$\{1,2\},\{3\},\{4,5\}$;

按4类分应分为：$\{1,2\},\{3\},\{4\},\{5\}$;

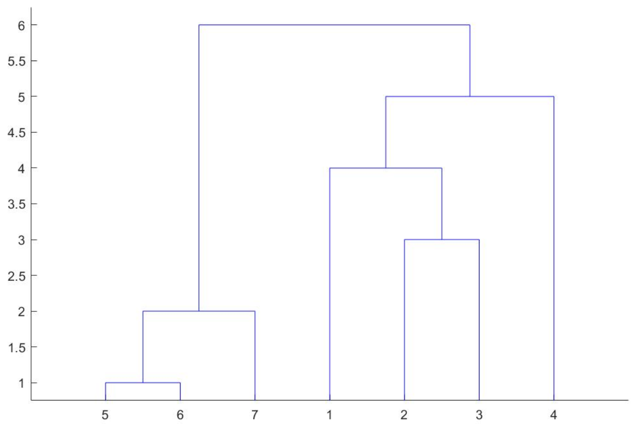
按5类分应分为：$\{1\},\{2\},\{3\},\{4\},\{5\}$;

4.

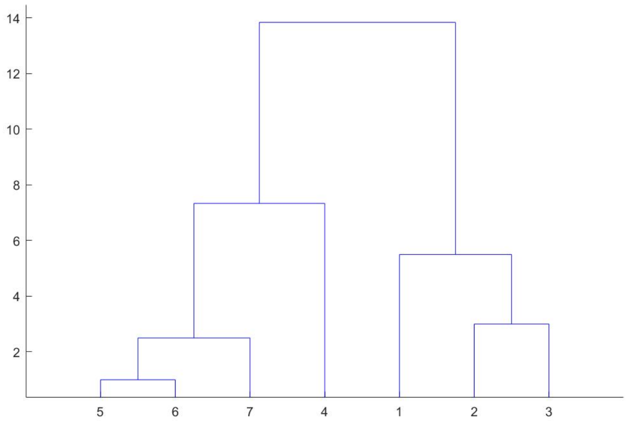
最小距离法：



最大距离法：



重心法：



5.

使用clusterdata(x,1)函数得到的结果为：

3 2 3 1 2 1 1 2 3

因此作出以下预报：

1968年约为240，1969年约为452.

解：若使用系统聚类方法,样品间距离分别取欧氏距离和马氏距离,类间距离取类平方距离,matlab代码如下：

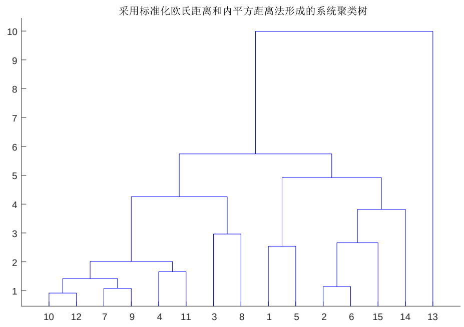
```
clear
clc
n = 15;c = 8;
X = [11.09  0.21  0.05  96.98  70.53  1.86
    -44.04  81.99  ;
        11.96  0.59  0.74  51.78  90.73  4.95
            7.02  16.11  ;
        0  0.03  0.03  181.99  100  -2.98  103.33
             21.18;
```

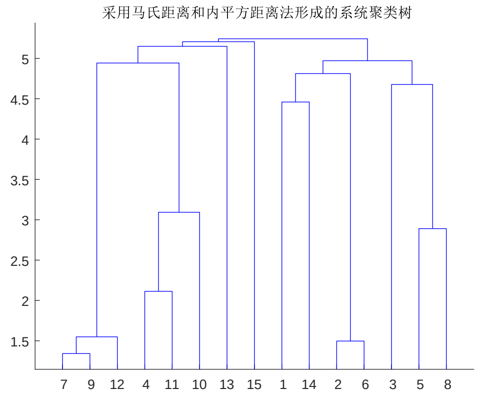
```
11.58 0.13 0.17 46.07 92.18 1.14 6.55 −56.32;
−6.19 −0.09 0.03 43.3 82.24 1.52 −1713.5 −3.36;
10 0.47 0.48 68.4 86 4.7 −11.56 0.85 ;
10.49 0.11 0.35 82.98 99.87 1.02 100.23 30.32;
11.12 −1.69 0.12 132.14 100 −0.66 −4454.39 −62.75;
3.41 0.04 0.2 67.86 98.51 1.25 −11.25 −11.43;
1.16 0.01 0.54 43.7 100 1.03 −87.18 −7.41;
30.22 0.16 0.4 87.36 94.88 0.53 729.41 −9.97;
8.19 0.22 0.38 30.31 100 2.73 −12.31 −2.77;
95.79 −5.2 0.5 252.34 99.34 −5.42 −9816.52 −46.82;
16.55 0.35 0.93 72.31 84.05 2.14 115.95 123.41;
−24.18 −1.16 0.79 56.26 97.8 4.81
```

```
              -533.89   -27.74];
      X2 = zscore(X);%标准化数据
      Y2 = pdist(X2,'seuclidean');%计算距离
      Z2 = linkage(Y2,'ward');%定义变量之间的连
         接
      C2 = cophenet(Z2,Y2);%评价聚类信息
      %创建聚类，并作出谱系图
      T = cluster(Z2,15);
      H = dendrogram(Z2);
      title('采用标准化欧氏距离和内平方距离法形成的
         系统聚类树');
      figure(2)
      X2 = zscore(X);%标准化数据
      Y2 = pdist(X2,'mahalanobis');%计算距离
      Z2 = linkage(Y2,'ward');%定义变量之间的连
         接
      C2 = cophenet(Z2,Y2);%评价聚类信息
      %创建聚类，并作出谱系图
      T = cluster(Z2,15);
      H = dendrogram(Z2);
      title('采用马氏距离和内平方距离法形成的系统聚
         类树');
```

运行结果为:

如果使用K-means方法，样品间距离取欧氏距离,分别分为3,4,5,6类,matlab代





码如下：

```
clear
 clc
X = [11.09 0.21 0.05 96.98 70.53 1.86 -44.04 81.99 ;
    11.96 0.59 0.74 51.78 90.73 4.95 7.02 16.11 ;
    0 0.03 0.03 181.99 100 -2.98 103.33 21.18;
    11.58 0.13 0.17 46.07 92.18 1.14 6.55 -56.32;
    -6.19 -0.09 0.03 43.3 82.24 1.52 -1713.5 -3.36;
    10 0.47 0.48 68.4 86 4.7 -11.56 0.85 ;
    10.49 0.11 0.35 82.98 99.87 1.02 100.23 30.32;
```

```
    11.12 -1.69 0.12 132.14 100 -0.66 -4454.39 -62.75;
    3.41 0.04 0.2 67.86 98.51 1.25 -11.25 -11.43;
    1.16 0.01 0.54 43.7 100 1.03 -87.18 -7.41;
    30.22 0.16 0.4 87.36 94.88 0.53 729.41 -9.97;
    8.19 0.22 0.38 30.31 100 2.73 -12.31 -2.77;
    95.79 -5.2 0.5 252.34 99.34 -5.42 -9816.52 -46.82;
    16.55 0.35 0.93 72.31 84.05 2.14 115.95 123.41;
    -24.18 -1.16 0.79 56.26 97.8 4.81 -533.89 -27.74];
opts = statset('Display','final');
[Idx,Ctrs,SumD,D] = kmeans(X,3,'Replicates',5,'Options',opts);
[Idx,Ctrs,SumD,D] = kmeans(X,4,'Replicates',5,'Options',opts);
[Idx,Ctrs,SumD,D] = kmeans(X,5,'Replicates',5,'Options',opts);
```

类别	编号
1	13
2	8
3	1,2,3,4,5,6,7,9,10,11,12,14,15

类别	编号
1	13
2	8
3	1,2,3,4,6,7,9,10,11,12,14,15
4	5

类别	编号
1	13
2	8
3	1,2,3,4,6,7,9,10,12,14,15
4	5
5	11

类别	编号
1	13
2	8
3	1,2,3,4,6,7,9,10,12,14
4	5
5	11
6	15

7.

将北京 河南排序为1-16，将食品-娱乐教育文化排序为1-6

(1).K-means,'Distance':'cityblock':

```
%q7_1
clear all;
clc;
name=['北京 ';'天津 ';'河北 ';'山西 ';'内蒙古';'辽宁 ';'吉林 ';'黑龙江';'上海 ' ...
    ;'江苏 ';'浙江 ';'安徽 ';'福建 ';'江西 ';'山东 ';'河南 '];
X=[190.33 43.77 9.73 60.54 49.01 9.04;
    135.2 36.4 10.47 44.16 36.49 3.94;
    95.21 22.83 9.3 22.44 22.81 2.8;
    104.78 25.11 6.4 9.89 18.17 3.25;
    128.41 27.63 8.94 12.58 23.99 2.27;
    145.68 32.83 17.79 27.29 39.09 3.47;
    159.37 33.38 18.37 11.81 25.29 5.22;
    116.22 29.57 13.24 13.76 21.75 6.04;
    221.11 38.64 12.53 115.65 50.82 5.89;
    144.98 29.12 11.67 42.6 27.3 5.74;
    169.92 32.75 12.72 47.12 34.35 5;
    135.11 23.09 15.62 23.54 18.18 6.39;
    144.92 21.26 16.96 19.52 21.75 6.73;
    140.54 21.5 17.64 19.19 15.97 4.94;
    115.84 30.26 12.2 33.6 33.77 3.85;
    101.18 23.26 8.46 20.2 20.5 4.3];%元素X(i)的指标为X(i,j)
[Idx,C,sumD,D]=kmeans(X,4,'Distance','cityblock');
g1=name(Idx==1,1:3)
g2=name(Idx==2,1:3)
g3=name(Idx==3,1:3)
g4=name(Idx==4,1:3)
```

利用MATLAB的kmeans函数，并且选择参数K=4，cityblock距离算法得到的

分类结果为：

【北京，上海】;【天津，辽宁，吉林，江苏，浙江，安徽，福建，江西】;【河北，山西，河南】;【内蒙古，黑龙江，山东】

(2).动态聚类法,'metric':'seuclidean':

```
%q7_2
clear all;
clc;
name=['北京 ';'天津 ';'河北 ';'山西 ';'内蒙古';'辽宁 ';'吉林 ';'黑龙江';'上海 ' ...
    ;'江苏 ';'浙江 ';'安徽 ';'福建 ';'江西 ';'山东 ';'河南 '];
X=[190.33 43.77 9.73 60.54 49.01 9.04;
    135.2 36.4 10.47 44.16 36.49 3.94;
    95.21 22.83 9.3 22.44 22.81 2.8;
    104.78 25.11 6.4 9.89 18.17 3.25;
    128.41 27.63 8.94 12.58 23.99 2.27;
    145.68 32.83 17.79 27.29 39.09 3.47;
    159.37 33.38 18.37 11.81 25.29 5.22;
    116.22 29.57 13.24 13.76 21.75 6.04;
    221.11 38.64 12.53 115.65 50.82 5.89;
    144.98 29.12 11.67 42.6 27.3 5.74;
    169.92 32.75 12.72 47.12 34.35 5;
    135.11 23.09 15.62 23.54 18.18 6.39;
    144.92 21.26 16.96 19.52 21.75 6.73;
    140.54 21.5 17.64 19.19 15.97 4.94;
    115.84 30.26 12.2 33.6 33.77 3.85;
    101.18 23.26 8.46 20.2 20.5 4.3];%元素X(i)的指标为X(i,j)
Dis=squareform(pdist(zscore(X),'seuclidean'));%标准化欧氏距离
Z=linkage(Dis);
T=cluster(Z,4);
g1=name(T==1,1:3)
g2=name(T==2,1:3)
```

```
g3=name(T==3,1:3)
g4=name(T==4,1:3)
```

动态聚类且分为4类，距离算法使用标准化欧氏距离得到的分类结果为：【上海】;【北京】;【天津，辽宁，吉林，黑龙江，江苏，浙江，安徽，福建，江西，山东】;【河北，山西，内蒙古，河南】

13.

利用MATLAB进行判别分析，选用diagQuadratic模型：

```
%q_7
clear all;
clc;
A=[76 99 5374;
    79.5 99 5359;
    78 99 5372;
    72.1 95.9 5242;
    73.8 77.7 5370];
B=[71.2 93 4250;
    75.3 94.9 3412;
    70 91.2 3390;
    72.8 99 2300;
    62.9 80.6 3799];
X=[68.5 79.3 1950;
    69.9 96.9 2840;
    77.6 93.8 5233;
    69.3 90.3 5158];
T=[A;B];G=[1;1;1;1;1;2;2;2;2;2];
CLASS=classify(X,T,G,'diagQuadratic')
```

可以得到结果：希腊为高发展水平国家，中国、罗马尼亚、哥伦比亚为中等发展国家

(1) 距离准则：

$$d^2(x,G_1)=\frac{(x-\mu_1)^2}{\sigma_1^2}=1$$

$$d^2(x,G_2)=\frac{(x-\mu_2)^2}{\sigma_2^2}=1.5625$$

$$d^2(x,G_3)=\frac{(x-\mu_3)^2}{\sigma_3^2}=0.25$$

所以$d^2(x,G_3)$最小，根据距离准则，$x=2.5$应判归为第三个总体.

(2) $Bayes$准则：由题干可知，三个总体的密度函数依次如下：

$$f_1(x)=\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}=\frac{1}{0.5\sqrt{2\pi}}e^{-\frac{(x-2)^2}{0.5}}$$

$$f_2(x)=\frac{1}{2\sqrt{2\pi}}e^{-\frac{x^2}{8}}$$

$$f_1(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-3)^2}{2}}$$

因为$L(j|i)=\begin{cases}1, & \text{i != j;}\\ 0, & \text{i = j.}\end{cases}$ 所以使错判损失最小与使后验概率最大是等价的；且三个总体的组内协差阵和先验概率均相等，所以$D_t^2(X)=d_t^2(X)$

计算三个总体的后验概率结果如下：

$$P(G_1|x)=\frac{e^{-0.5d_1^2(x)}}{e^{-0.5d_1^2(x)}+e^{-0.5d_2^2(x)}+e^{-0.5d_3^2(x)}}=0.3242$$

$$P(G_2|x)=\frac{e^{-0.5d_2^2(x)}}{e^{-0.5d_1^2(x)}+e^{-0.5d_2^2(x)}+e^{-0.5d_3^2(x)}}=0.1577$$

$$P(G_3|x)=\frac{e^{-0.5d_3^2(x)}}{e^{-0.5d_1^2(x)}+e^{-0.5d_2^2(x)}+e^{-0.5d_3^2(x)}}=0.5181$$

所以$P(G_3|x)$最大，根据$Bayes$准则，$x=2.5$应判归为第三个总体.

9.应用R语言对问题进行分析，先取样本$x_{(1)}$进行分析：

先求出该样本到两个类均值之间的距离（马氏距离和广义平方距离）。

```r
distma <-function(a,b,s)
{
  return (t(a-b)%*%solve(s)%*%(a-b))
}
gensqr_1 <- function(a,b,s)
{
  return (t(a-b)%*%solve(s)%*%(a-b)+log(det(s)))
}
q1 <- 0.5
q2 <- 0.5
L1_2 <- 75
L2_1 <- 10
x_9_1 <- matrix(c(20,20),nrow = 2)
x_9_2 <-  matrix(c(15,20),nrow = 2)
g1 <- matrix(c(10,15),nrow = 2)#类的均值G1
g2 <- matrix(c(20,25),nrow = 2)#类的均值G2
s1 <- matrix(c(18,12,12,32),nrow = 2)#类的协方差矩阵G1
s2 <- matrix(c(20,-7,-7,5),nrow = 2)#类的协方差矩阵G2
d11 <- distma(x_9_1,g1,s1)#x(1)到类的马氏距离G1
d12 <- distma(x_9_1,g2,s2)#x(1)到类的马氏距离G2

D11 <- gensqr_1(x_9_1,g1,s1)#x(1)到类的广义平方距离G1
```

```
D12 <- gensqr_1(x_9_1,g2,s2)#x(1)到类的广义平方距离G2

#输出结果为:
> D11
           [,1]
[1,] 11.73972
> D12
           [,1]
[1,] 13.73575
```

进而我们再计算两个后验概率$P(1|x_{(1)}), P(2|x_{(1)})$,再求出两个误选损失。

```
P1_1 <- exp(-0.5*D11)/(exp(-0.5*D11)+exp(-0.5*D12))#后验概率P(1|x1)
P2_1 <- exp(-0.5*D12)/(exp(-0.5*D11)+exp(-0.5*D12))#后验概率P(2|x1)
loss1_1 <- P2_1*L1_2#判别x(1)属于的平均损失G1
loss1_2 <- P1_1*L2_1#判别x(1)属于的平均损失G2

#输出结果为:
> P1_1
           [,1]
[1,] 0.7306677
```

```
> P1_2
            [,1]
[1,] 0.2693323
> loss1_1
           [,1]
[1,] 20.19993
> loss1_2
           [,1]
[1,] 7.306677
```

按照标准贝叶斯判别法，应将$x_{(1)}$归为G_1类。但是按照平均损失最小原则应将$x_{(1)}$归为G_2类。

同理，对于$x_{(2)}$,我们通过求得$loss2_1, loss2_2$，

```
d21 <- distma(x_9_2,g1,s1)#x(2)到类的马氏距离G1
d22 <- distma(x_9_2,g2,s2)#x(2)到类的马氏距离G2

D21 <- gensqr_1(x_9_2,g1,s1)#x(2)到类的广义平方距离G1
D22 <- gensqr_1(x_9_2,g2,s2)#x(2)到类的广义平方距离G2

P1_2 <- exp(-0.5*D21)/(exp(-0.5*D21)+exp(-0.5*D22))#后验概率P(1|x2)
P2_2 <- exp(-0.5*D22)/(exp(-0.5*D21)+exp(-0.5*D22))#后验概
```

```
    率P(2|x2)
loss2_1 <- P2_2*L1_2#判别x(2)属于的平均损失G1
loss2_2 <- P1_2*L2_1#判别x(2)属于的平均损失G2

#输出结果为:
> P1_2
             [,1]
[1,] 0.9995643
> P2_2
                 [,1]
[1,] 0.0004356617
> loss2_1
              [,1]
[1,] 0.03267463
> loss2_2
            [,1]
[1,] 9.995643
```

所以，按照标准贝叶斯判别法或者平均损失最小原则，都应将$x_{(2)}$归为G_1类。

10.假设对式样5，包装5，耐久性5做Fisher判别。

由matlab

```
X=[9 8 7;7 6 6;10 7 8;8 4 5;9 9 7;8 6 7;7 5 6;4 4
    4;3 6 6;6 3 3;2 4 5;1 2 2];
```

```
x1=X(1:7,:);
x2=X(8:12,:);
sample=[5 5 5];
y=fisher(x1,x2,sample)

function y=fisher(x1,x2,sample)
%函数Fisher
%x1,x2,分别为两类训练样本及待测数据集，其中行为样本数，列为特征数sample
r1=size(x1,1);r2=size(x2,1);
r3=size(sample,1);
a1=mean(x1)';a2=mean(x2)';
s1=cov(x1)*(r1-1);s2=cov(x2)*(r2-1);
sw=s1+s2;%求出协方差矩阵
w=inv(sw)*(a1-a2)*(r1+r2-2);
y1=mean(w'*a1);
y2=mean(w'*a2);
y0=(r1*y1+r2*y2)/(r1+r2);
for i=1:r3
   y(i)=w'*sample(i,:)';
     if y(i)>y0
         y(i)=0;
     else
         y(i)=1;
```

```
    end
end
```

可知，式样5，包装5，耐久性5的产品不值得购买。

12、使用Fisher判别法，由Matlab可求出

由matlab

```
X=[-1.9 3.2;-6.9 10.4;5.2 2.0;5.0 2.5;7.3
    0;6.8,12.7;0.9 -15.4;-12.5 -2.5;1.5 1.3;3.8 6.8
      0.2 0.2;-0.1 7.5;0.4 14.6;2.7 8.3;2.1 0.8;-4.6
          4.3;-1.7 10.9;-2.6 13.1;2.6 12.8;-2.8 10]
x1=X(1:10,:);
x2=X(10:20,:);
sample=[8.2 2];
y=fisher(x1,x2,sample)

function y=fisher(x1,x2,sample)
%函数Fisher
%x1,x2,分别为两类训练样本及待测数据集，其中行为样本数，列
    为特征数sample
r1=size(x1,1);r2=size(x2,1);
r3=size(sample,1);
a1=mean(x1)';a2=mean(x2)';
```

```
s1=cov(x1)*(r1-1);s2=cov(x2)*(r2-1);
sw=s1+s2;%求出协方差矩阵
w=inv(sw)*(a1-a2)*(r1+r2-2);
y1=mean(w'*a1);
y2=mean(w'*a2);
y0=(r1*y1+r2*y2)/(r1+r2);
for i=1:r3
   y(i)=w'*sample(i,:)';
     if y(i)>y0
         y(i)=0;
     else
         y(i)=1;
     end
end
```

预报为明天是雨天。

11.

(1)使用函数[class,err]=classify(train,train,group)得到的结果为[1 1 1 2 1 2 2 3 3 3]

(2)使用函数[class,err]=classify(new,train,group)得到的结果为[2],即预测新品销售情况为平销.

解：若使用距离判别法，matlab代码如下:

```
     clear
```

```
clc
sample = [53 1 9 18 50 11.20 2.02 3.58];
training = [23 1 7 2 31 6.60 0.34 1.71;
 34 1 17 3 59 8.00 1.81 2.91;
 42 2 7 23 41 4.60 0.94 0.94;
 39 1 19 5 48 13.10 1.93 4.36;
 35 1 9 1 34 5.00 0.40 1.30;
 37 1 1 3 24 15.10 1.80 1.82
 29 1 13 1 42 7.40 1.46 1.65
 32 2 11 6 75 23.30 7.76 9.72
 28 2 2 3 23 6.40 0.19 1.29
 26 1 4 3 27 10.50 2.47 0.36];
for i = 1:8
    mean1(i) = mean(training(1:5,i));
    mean2(i) = mean(training(6:10,i));
end
%training(1:5,:)
cov1 = cov(training(1:5,:));
cov2 = cov(training(6:10,:));
%inv(cov1)
d1 = (sample - mean1)*cov1*(sample -
   mean1)'
d2 = (sample - mean2)*cov2*(sample -
```

```
mean2)'
```

判别公式为:

$$若d_1 < d_2,属于1类，即已履行还贷责任；$$

$$若d_1 > d_2,属于1类，即未履行还贷责任；$$

本题求出的

$$d_1 = 6.0551 * 10^4,$$

$$d_2 = 9.4369 * 10^4$$

$$d_1 < d_2$$

因此该客户已履行还贷责任。

若使用Bayes判别法，使用matlab代码为：

```
clear
 clc
 sample = [53 1 9 18 50 11.20 2.02 3.58];
 training = [23 1 7 2 31 6.60 0.34 1.71;
  34 1 17 3 59 8.00 1.81 2.91;
  42 2 7 23 41 4.60 0.94 0.94;
  39 1 19 5 48 13.10 1.93 4.36;
  35 1 9 1 34 5.00 0.40 1.30;
  37 1 1 3 24 15.10 1.80 1.82
  29 1 13 1 42 7.40 1.46 1.65
  32 2 11 6 75 23.30 7.76 9.72
  28 2 2 3 23 6.40 0.19 1.29
```

```
  26  1  4  3  27  10.50  2.47  0.36];
 groups = [1  1  1  1  1  2  2  2  2  2];
 nb=NaiveBayes.fit(training,groups);
 cpre=predict(nb,sample)
```

求出结果，cpre=1，即该客户已履行还贷责任。

若使用Fisher判别法，matlab代码如下：

```
Fisher.m
 function  y=fisher(x1,x2,sample)
  %函数Fisher
  %x1,x2,分别为两类训练样本及待测数据集，其中行为样本数，列为特征数sample
   r1=size(x1,1);r2=size(x2,1);
   r3=size(sample,1);
   a1=mean(x1)';a2=mean(x2)';
   s1=cov(x1)*(r1-1);s2=cov(x2)*(r2-1);
   sw=s1+s2;%求出协方差矩阵
   w=inv(sw)*(a1-a2)*(r1+r2-2);
   y1=mean(w'*a1);
   y2=mean(w'*a2);
   y0=(r1*y1+r2*y2)/(r1+r2);
   for  i=1:r3
      y(i)=w'*sample(i,:)';
        if  y(i)>y0
```

```
                y(i)=0;
            else
                y(i)=1;
            end
    end
m12:
clear
    clc
    sample = [53  1  9  18  50  11.20  2.02  3.58];
    training = [23  1  7  2  31  6.60  0.34  1.71;
      34  1  17  3  59  8.00  1.81  2.91;
      42  2  7  23  41  4.60  0.94  0.94;
      39  1  19  5  48  13.10  1.93  4.36;
      35  1  9  1  34  5.00  0.40  1.30;
      37  1  1  3  24  15.10  1.80  1.82
      29  1  13  1  42  7.40  1.46  1.65
      32  2  11  6  75  23.30  7.76  9.72
      28  2  2  3  23  6.40  0.19  1.29
      26  1  4  3  27  10.50  2.47  0.36];
x1 = training(1:5,:);
x2 = training(6:10,:);
fisher(x1,x2,sample);
```

最后结果为，该客户已履行还贷责任.