



石家莊鐵道大學  
SHIJIAZHUANG TIEDAO UNIVERSITY

在线开放课程

MATLAB在科学研究中的应用

总结与习题

主讲：卞建鹏

# 1、总结



在线开放课程

1. 主成分分析
2. Fisher判别法
3. K近邻法
4. 聚类分析
5. 决策树
6. 神经网络
7. 模糊分析

# 一、填空题

(0701)1. 机器学习主要分为 有监督 学习与 无监督 学习。

(0701)2. 将 $n$ 维特征映射到 $k$ 维空间中( $n < k$ )，计算 $n$ 维数据矩阵的特征中最重要的部分，称为 数据降维。

(0701)3. 通过PCA求得的 最大特征值 对应的 特征向量 即为投射向量。

(0702)4. Fisher线性判别法的基本思想是通过将多维数据投影到某一方向上，使得投影后 类与类 之间尽可能的分开，然后再选择合适的 判别准则。

(0703)5. K近邻法的基本思想是在训练数据集中找到 k 个最近邻的实例，类别由这 $k$ 个近邻中占 最多的实例类别 所决定。

(0703)6. 常用的距离计算方法有欧式距离、绝对距离、切式距离等。其中k近邻与聚类分析主要采用 欧式距离。

(0704)7. 聚类主要分为 K-means聚类 与 层次聚类；K-means聚类基本思想是计算特征矢量到聚类中心的 欧氏距离，并对距离进行比较，确定属于哪一类。

(0705)8. 决策树主要包括根节点、中间节点、分支、叶节点；其中中间节点代表测试的条件，叶节点代表 分类后所获得的分类标记。

(0705)9. ID3算法的属性选择度量就是使用信息增益，选择 最高信息增益 的属性作为决策树当前节点的测试属性。

- (0706)10. 人工神经网络 来源于对人脑实际神经网络的模拟；人工神经元模型的三要素包括 连接权、求和单元、激励函数。
- (0706)11. BP (Back Propagation) 神经网络是一种按误差逆传播算法训练的多层前馈网络，其学习规则是使用 最速下降法。
- (0707)12. 模糊集合的两要素：论域与 隶属函数；隶属度值越接近1，表示x属于高的程度越高。
- (0707) 13. 模糊模式识别原则主要分为 最大隶属度原则与 择近原则。其中条形码与文字识别可应用 择近原则进行判断。

## 二、综合编程题



在线开放课程

已知数据：男性与女性的年龄、身高与体重数据。现给出8组年龄、身高与体重数据，编写MATLAB程序判断其性别？

Gender	Age	Height(cm)	Weight(kg)	Gender	Age	Height(cm)	Weight(kg)
Male	21	163	60	Female	20	153	42
Male	22	164	56	Female	20	156	44
Male	21	165	60	Female	21	156	38
Male	23	168	55	Female	21	157	48
Male	21	169	60	Female	21	158	52
Male	21	170	54	Female	23	158	45
Male	23	170	80	Female	22	159	43
Male	23	170	64	Female	22	160	50
Male	22	171	67	Female	21	160	45
Male	22	172	65	Female	21	160	52
Male	23	172	60	Female	23	160	50
Male	21	172	60	Female	22	161	50
Male	23	173	60	Female	21	161	45

## 二、综合编程题

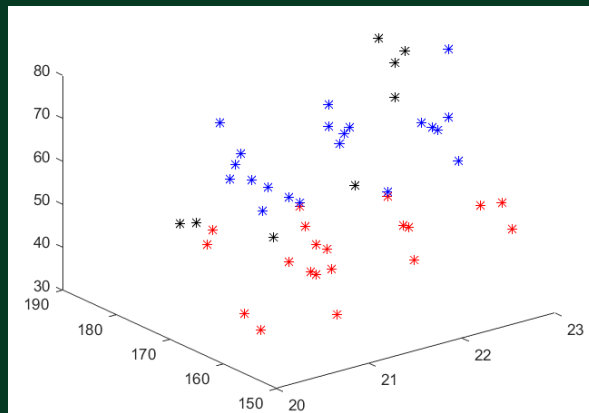


在线开放课程

已知数据：男性与女性的年龄、身高与体重数据。现给出8组年龄、身高与体重数据，编写MATLAB程序判断其性别？

Male	22	173	62	Female	21	162	55
Male	21	174	65	Female	20	162	60
Male	22	175	70	Female	20	163	56
Male	22	175	70	Female	20	163	56
Male	22	175	65	Female	21	163	59
Male	23	175	60	Female	22	164	55
Male	21	175	62	Female	23	164	47
Male	21	176	58	Female	21	165	45
Male	21	178	70	Female	21	165	45
?	23	178	75	?	20	165	60
?	23	180	63	?	20	168	58
?	23	180	71	?	21	168	49
?	23	183	75	?	22	170	54

```
male=xlsread('D:\genderdata.xls','B2:D23')
female=xlsread('D:\genderdata.xls','F2:H23')
test1=xlsread('D:\genderdata.xls','B24:D27');
test2=xlsread('D:\genderdata.xls','F24:H27');
test=[test1;test2];
gender=[male;female;test];
plot3(male(:,1),male(:,2),male(:,3),'b*');
hold on
plot3(female(:,1),female(:,2),female(:,3),'r*');
hold on
plot3(test(:,1),test(:,2),test(:,3),'k*'); hold on
```





```
male=xlsread('D:\genderdata.xls','B2:D23')
```

```
female=xlsread('D:\genderdata.xls','F2:H23')
```

```
test1=xlsread('D:\genderdata.xls','B24:D27');
```

```
test2=xlsread('D:\genderdata.xls','F24:H27');
```

```
test=[test1;test2];
```

```
gender=[male;female;test];
```

```
m=size(gender,1);
```

```
ma1=mean(male);
```

```
ma2=mean(female);
```

```
test =
```

23	178	75
23	180	63
23	180	71
23	183	75
20	165	60
20	168	58
21	168	49
22	170	54

```
ma1 =
```

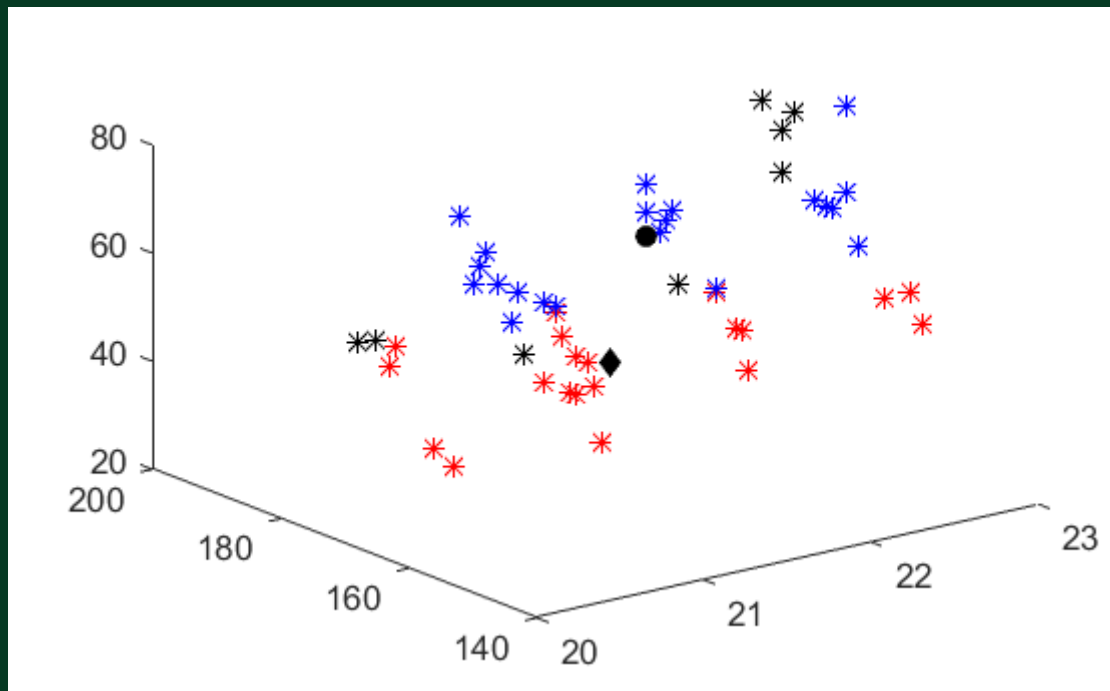
21.8636	171.5909	62.8636
---------	----------	---------

```
ma2 =
```

21.2273	160.4545	49.1818
---------	----------	---------

```
plot3(ma1(1),ma1(2),ma1(3),'ko')
```

```
plot3(ma2(1),ma2(2),ma2(3),'kd')
```



```
d1=(test-repmat(ma1,8,1)).^2
```

```
d2=(test-repmat(ma2,8,1)).^2
```

```
d11=sqrt(sum(d1'))
```

```
d22=sqrt(sum(d2'));
```

```
d=[d11;d22 ];
```

```
clas=zeros(1,8)
```

```
d =
```

```
13.7717    8.4866   11.7560   16.6958    7.4239    6.3264   14.3472    9.0063  
31.2660   24.0023   29.3462   34.3223   11.7983   11.6705    7.5511   10.7204
```

```
for i=1:8
```

```
if d(1,i)<d(2,i)
```

```
    a=1;
```

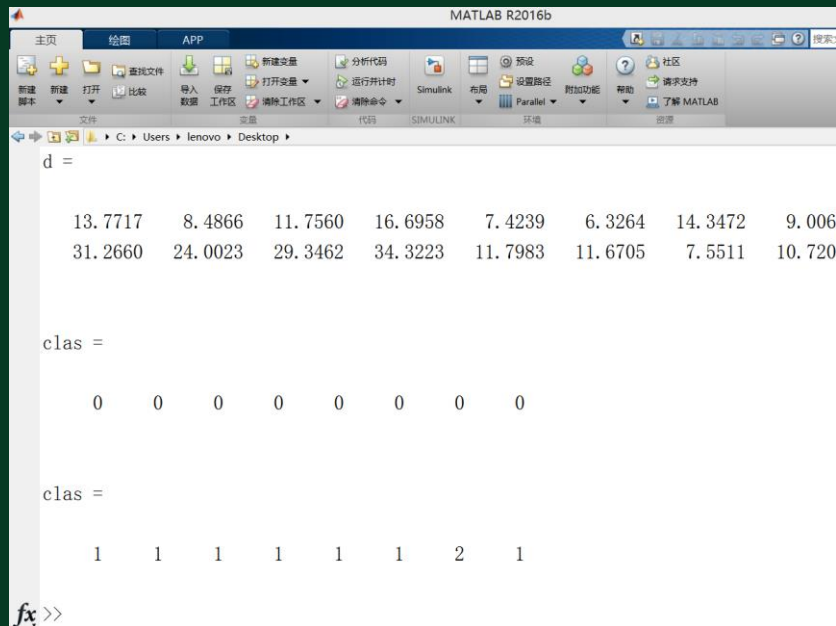
```
else
```

```
    a=2;
```

```
end
```

```
    clas(i)=a;
```

```
end
```



```
MATLAB R2016b
```

```
主页 绘图 APP 搜索
```

```
新建 新建 打开 查找文件 新建变量 分析代码 设置 社区
```

```
脚本 新建 打开 比较 导入 保存 打开变量 运行并计时 Simulink 布局 设置路径 附加功能 帮助 请求支持
```

```
数据 工作区 清除工作区 清除命令 并行 了解 MATLAB
```

```
文件 变量 代码 SIMULINK 环境 资源
```

```
C:\Users\lenovo\Desktop
```

```
d =
```

```
    13.7717    8.4866    11.7560    16.6958    7.4239    6.3264    14.3472    9.006
```

```
    31.2660    24.0023    29.3462    34.3223    11.7983    11.6705    7.5511    10.720
```

```
clas =
```

```
     0     0     0     0     0     0     0     0
```

```
clas =
```

```
     1     1     1     1     1     1     2     1
```

```
fx >>
```



在线开放课程

```
for i=1:8
if clas(i)==1
a='男性';
else
a='女性';
end
fprintf('%s\n',a)
end
```

```
clas =
     1     1     1     1     1     1     2     1

男性
男性
男性
男性
男性
男性
女性
男性
```

### 三、综合编程题

已知数据：电力变压器三种故障对应的5种特征气体，编写MATLAB程序判断测试样本数据对应的故障？

low temperature	H <sub>2</sub>	CH <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>	C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>4</sub>
1	181.00	262.00	210.00	528.00	0.00
2	160.00	130.00	33.00	96.00	0.00
3	98.00	155.00	54.00	300.00	1.50
4	35.35	51.17	8.01	5.47	0.00
5	4.32	193.00	118.00	125.00	0.00
6	170.00	320.00	53.00	520.00	3.20
7	40.26	25.11	18.62	16.02	0.00
8	4.91	10.14	62.09	22.23	0.63
9	27.00	90.00	42.00	63.00	0.20
10	120.00	160.00	55.00	300.00	1.40
high temperature					
1	70.40	69.50	28.90	241.20	10.40
2	135.00	466.00	70.00	502.00	90.00
3	90.00	149.00	32.40	486.00	19.20
4	31.70	40.30	31.70	40.30	31.70
5	20.60	19.89	7.50	61.27	1.52
6	11.30	21.84	11.30	53.10	2.46
7	0.96	15.79	12.30	70.25	0.71
8	164.00	244.00	103.00	497.00	8.30
9	274.00	376.00	55.00	1002.00	17.00
10	18.61	17.97	6.69	55.36	1.36

discharge					
1	980.00	73.00	58.00	12.00	0.00
2	650.00	53.00	34.00	20.00	0.00
3	1565.00	93.00	34.00	47.00	0.00
4	260.00	8.00	2.50	2.00	0.00
5	2587.20	7.88	4.70	1.40	0.00
6	790.00	82.61	8.90	9.17	0.00
7	124.00	20.67	17.81	15.33	0.00
8	9.82	0.98	11.21	2.37	0.30
9	12.60	105.54	13.80	8.32	0.00
10	154.00	50.00	24.90	8.00	0.00
test sample1	34.00	58.00	125.00	543.00	0.20
test sample2	1565.00	93.00	47.00	34.00	0.00

```
low=xlsread('D:\gas data.xlsx','C4:G13')
```

```
high=xlsread('D:\gas data.xlsx','C15:G24')
```

```
discharge=xlsread('D:\gas data.xlsx','C26:G35')
```

```
test1=xlsread('D:\gas data.xlsx','C37:G37');
```

```
test2=xlsread('D:\gas data.xlsx','C38:G38');
```

```
test=[test1;test2];
```

```
temperature=[low;high;discharge;test];
```

```
test =  
  
1.0e+03 *  
  
0.0340    0.0580    0.1250    0.5430    0.0002  
1.5650    0.0930    0.0470    0.0340     0
```

```
[COEFF SCORE latent]=princomp(temperature)
```

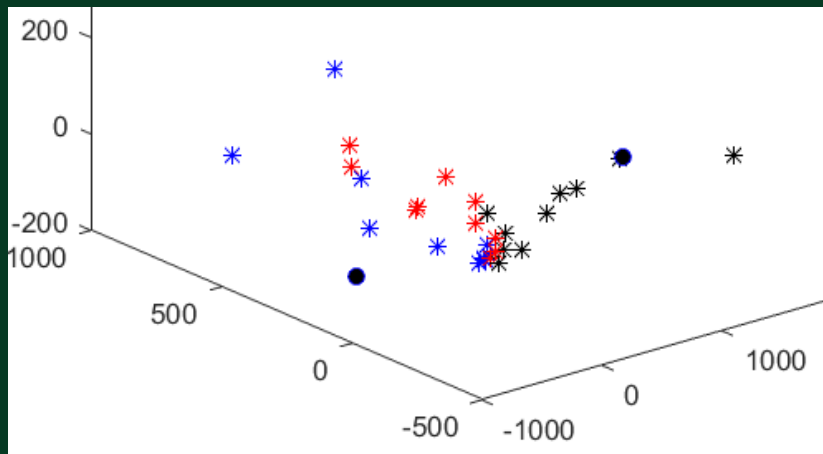
```
pcaD=SCORE(:,1:3)
```

```
plot3(pcaD(1:10,1),pcaD(1:10,2),pcaD(1:10,3),'r*');hold on
```

```
plot3(pcaD(11:20,1),pcaD(11:20,2),pcaD(11:20,3),'b*');hold on
```

```
plot3(pcaD(21:30,1),pcaD(21:30,2),pcaD(21:30,3),'k*');hold on
```

```
plot3(pcaD(31:32,1),pcaD(31:32,2),pcaD(31:32,3),'bo')
```





```
test=[test1;test2]; temperature=[low;high;discharge;test];
```

```
m=size(temperature,1);
```

```
ma1=mean(low);
```

```
ma2=mean(high);
```

```
ma3=mean(discharge);
```

```
ma1 =  
    84.0840    139.6420    65.3720    197.5720     0.6930  
  
ma2 =  
    81.6570    142.0290    35.8790    300.8480    18.2650  
  
ma3 =  
    713.2620    49.4680    20.9820    12.5590     0.0300
```



在线开放课程

```
d1=(test-repmat(ma1,2,1)).^2
```

```
d2=(test-repmat(ma2,2,1)).^2
```

```
d3=(test-repmat(ma3,2,1)).^2
```

```
d11=sqrt(sum(d1'))
```

```
d22=sqrt(sum(d2'))
```

```
d33=sqrt(sum(d3'))
```

```
d=[d11;d22;d33]
```

```
[m1,n1]=sort(d(:,1),1)
```

```
[m2,n2]=sort(d(:,2),1)
```

```
d =  
  
1.0e+03 *  
  
0.3634    1.4908  
0.2761    1.5081  
0.8681    0.8535
```

```
m1 =      n1 =  
  
276.1133    2  
363.3869    1  
868.1343    3
```

```
m2 =  
  
1.0e+03 *  n2 =  
  
0.8535    3  
1.4908    1  
1.5081    2
```

# 总结



在线开放课程

1. 主成分分析
2. Fisher判别法
3. K近邻法
4. 聚类分析
5. 决策树
6. 神经网络
7. 模糊分析