

实 验 报 告

学号：2111078 姓名：李岚琦

学院：人工智能学院 专业：智能科学与技术

实验三：朴素贝叶斯模型

1. 实验目的

学习朴素贝叶斯模型进行分类预测

1. 实验原理

将后验概率最大的类作为输出类

1. 实验步骤与代码结果
2. 生成样本数据和噪声

代码：

%%%%%%%%%%%%%%%%%%% 数据生成 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

n = 2000; % 样本量大小

X = rand(n,2)\*10; % n \* 2的数据矩阵，每一行表示一个数据点，第一列表示x轴坐标，第二列表示y轴坐标

Y = zeros(n,1); % 类别标签

for i=1:n

if 0<X(i,1) && X(i,1)<3 && 0<X(i,2) && X(i,2)<3 % 根据x和y轴坐标确定分类

Y(i) = 1;

end

if 0<X(i,1) && X(i,1)<3 && 3.5<X(i,2) && X(i,2)<6.5

Y(i) = 2;

end

if 0<X(i,1) && X(i,1)<3 && 7<X(i,2) && X(i,2)<10

Y(i) = 3;

end

if 3.5<X(i,1) && X(i,1)<6.5 && 0<X(i,2) && X(i,2)<3

Y(i) = 4;

end

if 3.5<X(i,1) && X(i,1)<6.5 && 3.5<X(i,2) && X(i,2)<6.5

Y(i) = 5;

end

if 3.5<X(i,1) && X(i,1)<6.5 && 7<X(i,2) && X(i,2)<10

Y(i) = 6;

end

if 7<X(i,1) && X(i,1)<10 && 0<X(i,2) && X(i,2)<3

Y(i) = 7;

end

if 7<X(i,1) && X(i,1)<10 && 3.5<X(i,2) && X(i,2)<6.5

Y(i) = 8;

end

if 7<X(i,1) && X(i,1)<10 && 7<X(i,2) && X(i,2)<10

Y(i) = 9;

end

end

X = X(Y>0,:); % 注意X是在[0,10]\*[0,10]范围内均匀生成的，而我们只标出了一部分X，类别之间的白色间隔中的点没有标，因此需要将这些点去掉

Y = Y(Y>0,:); % X(Y>0,:)表示只取X中对应的Y大于0的行，这是因为白色间隔中的点的Y都为0

nn = length(Y); % 去除掉白色间隔剩下的点

n = 2000;

X(nn+1:n,:) = rand(n-nn,2)\*10; % 增加n-nn个噪声点

Y(nn+1:n) = ceil( rand(n-nn,1)\*9 ); % 噪声点的标签随机选取。rand(n-nn,1)\*9表示生产[0,9]的均匀分布，ceil表示上取整，故结果为1,2,...,9

figure(1)

set (gcf,'Position',[1,1,700,600], 'color','w')

set(gca,'Fontsize',18)

plot(X(Y==1,1),X(Y==1,2),'ro','LineWidth',1,'MarkerSize',10); % 画第一类数据点 X(Y==1,1)表示类别为1（Y==1）的点的第一维度坐标，X(Y==1,2)表示类别为1的点的第二维度坐标

hold on;

plot(X(Y==2,1),X(Y==2,2),'ko','LineWidth',1,'MarkerSize',10); % 画第二类数据点

hold on;

plot(X(Y==3,1),X(Y==3,2),'bo','LineWidth',1,'MarkerSize',10); % 画第三类数据点

hold on;

plot(X(Y==4,1),X(Y==4,2),'g\*','LineWidth',1,'MarkerSize',10); % 画第四类数据点

hold on;

plot(X(Y==5,1),X(Y==5,2),'m\*','LineWidth',1,'MarkerSize',10); % 画第五类数据点

hold on;

plot(X(Y==6,1),X(Y==6,2),'c\*','LineWidth',1,'MarkerSize',10); % 画第六类数据点

hold on;

plot(X(Y==7,1),X(Y==7,2),'b+','LineWidth',1,'MarkerSize',10); % 画第七类数据点

hold on;

plot(X(Y==8,1),X(Y==8,2),'r+','LineWidth',1,'MarkerSize',10); % 画第八类数据点

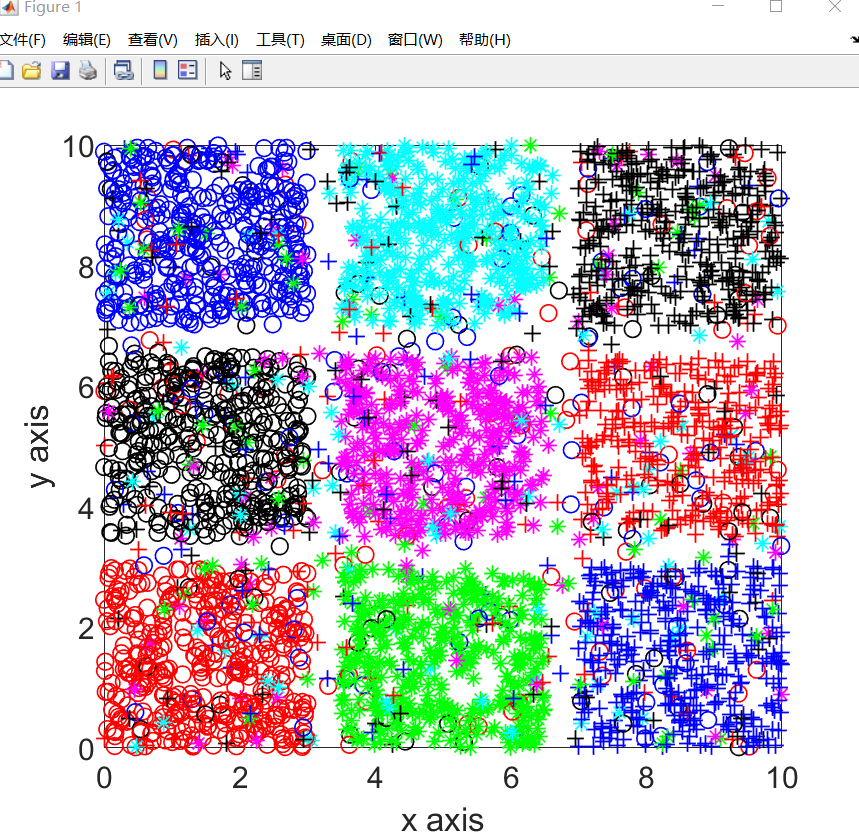
hold on;

plot(X(Y==9,1),X(Y==9,2),'k+','LineWidth',1,'MarkerSize',10); % 画第九类数据点

hold on;

xlabel('x axis');

ylabel('y axis');

运行结果：

1. 生成测试数据

代码：

%%%%%%%%%%%%%%%%%%% 生成测试数据 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%% 生成测试数据:与训练数据同分布 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

m = 100; % 测试样本量大小

Xt = rand(m,2)\*10;

Yt = zeros(m,1);

for i=1:m

if 0<Xt(i,1) && Xt(i,1)<3 && 0<Xt(i,2) && Xt(i,2)<3

Yt(i) = 1;

end

if 0<Xt(i,1) && Xt(i,1)<3 && 3.5<Xt(i,2) && Xt(i,2)<6.5

Yt(i) = 2;

end

if 0<Xt(i,1) && Xt(i,1)<3 && 7<Xt(i,2) && Xt(i,2)<10

Yt(i) = 3;

end

if 3.5<Xt(i,1) && Xt(i,1)<6.5 && 0<Xt(i,2) && Xt(i,2)<3

Yt(i) = 4;

end

if 3.5<Xt(i,1) && Xt(i,1)<6.5 && 3.5<Xt(i,2) && Xt(i,2)<6.5

Yt(i) = 5;

end

if 3.5<Xt(i,1) && Xt(i,1)<6.5 && 7<Xt(i,2) && Xt(i,2)<10

Yt(i) = 6;

end

if 7<Xt(i,1) && Xt(i,1)<10 && 0<Xt(i,2) && Xt(i,2)<3

Yt(i) = 7;

end

if 7<Xt(i,1) && Xt(i,1)<10 && 3.5<Xt(i,2) && Xt(i,2)<6.5

Yt(i) = 8;

end

if 7<Xt(i,1) && Xt(i,1)<10 && 7<Xt(i,2) && Xt(i,2)<10

Yt(i) = 9;

end

end

Xt = Xt(Yt>0,:);

Yt = Yt(Yt>0,:);

m = length(Yt);

Ym = zeros(m,1); % 记录模型输出结果

figure(2)

set (gcf,'Position',[1,1,700,600], 'color','w')

set(gca,'Fontsize',18)

plot(X(Y==1,1),X(Y==1,2),'ro','LineWidth',1,'MarkerSize',10); % 画第一类数据点

hold on;

plot(X(Y==2,1),X(Y==2,2),'ko','LineWidth',1,'MarkerSize',10); % 画第二类数据点

hold on;

plot(X(Y==3,1),X(Y==3,2),'bo','LineWidth',1,'MarkerSize',10); % 画第三类数据点

hold on;

plot(X(Y==4,1),X(Y==4,2),'g\*','LineWidth',1,'MarkerSize',10); % 画第四类数据点

hold on;

plot(X(Y==5,1),X(Y==5,2),'b\*','LineWidth',1,'MarkerSize',10); % 画第五类数据点

hold on;

plot(X(Y==6,1),X(Y==6,2),'c\*','LineWidth',1,'MarkerSize',10); % 画第六类数据点

hold on;

plot(X(Y==7,1),X(Y==7,2),'b+','LineWidth',1,'MarkerSize',10); % 画第七类数据点

hold on;

plot(X(Y==8,1),X(Y==8,2),'r+','LineWidth',1,'MarkerSize',10); % 画第八类数据点

hold on;

plot(X(Y==9,1),X(Y==9,2),'k+','LineWidth',1,'MarkerSize',10); % 画第九类数据点

hold on;

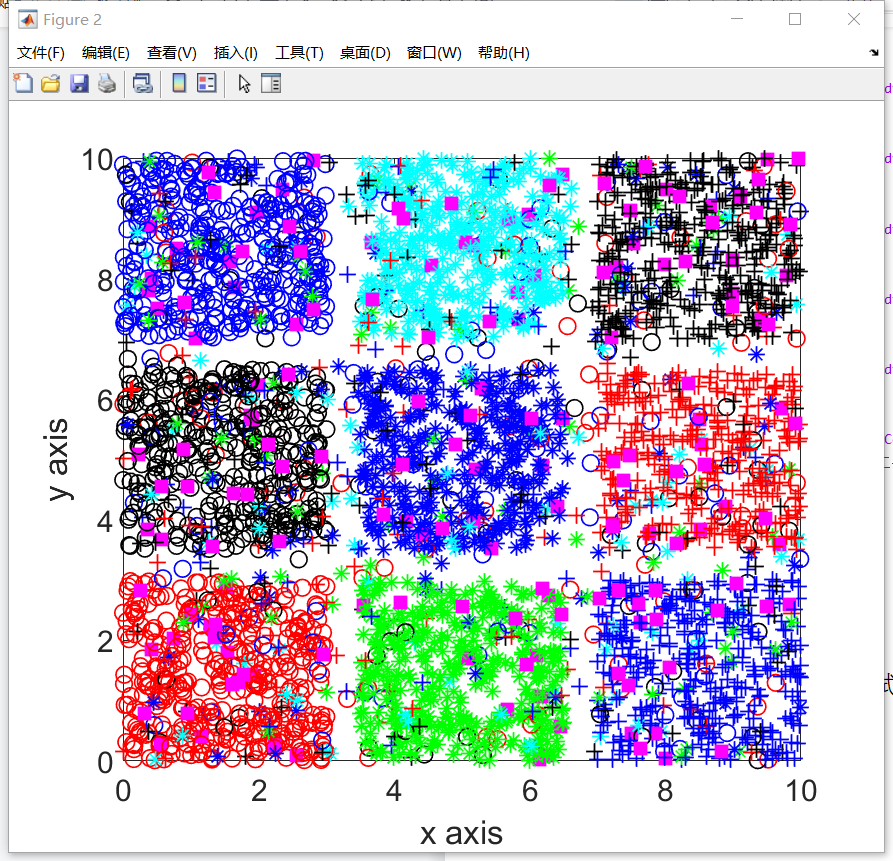
plot(Xt(:,1),Xt(:,2),'ms','MarkerFaceColor','m','LineWidth',1,'MarkerSize',10); % 画测试数据点 Xt(:,2)表示Xt的第二列，包括所有行

hold on;

xlabel('x axis');

ylabel('y axis');

运行结果：



1. 利用朴素贝叶斯计算各个测试点的类型概率

代码：

%%%%%%%%%%%%%%%%%% 贝叶斯算法：学生实现 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%% 给出模型的预测输出，并与测试数据的真实输出比较，计算错误率 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

GaiXian=ones(m,9)\*1/9;

p1=1/9;%%各个先验概率

p2=1/9;

p3=1/9;

p4=1/9;

p5=1/9;

p6=1/9;

p7=1/9;

p8=1/9;

p9=1/9;

%%求各个测试点的条件概率

tmp=zeros(m,9);

for i=1:1:m

for j=1:1:9

tmp(i,j)=1/9\*tiaojian1(Xt(i,:),j)\*tiaojian2(Xt(i,:),j); %%统计联合概率

end

for j=1:1:9

if tmp(i,j)==max(tmp(i,:))

Ym(i,1)=j;

end

end

end

%将测试结果存放在Ym中

function[re]=tiaojian1(x,y)%%x第一个特征的条件概率

re=0;

if (y==1||y==2||y==3)&&0<x(1,1)&&x(1,1)<3

re=1;

end

if (y==4||y==5||y==6)&&3.5<x(1,1)&&x(1,1)<6.5

re=1;

end

if (y==7||y==8||y==9)&&7<x(1,1)&&x(1,1)<10

re=1;

end

end

function[re]=tiaojian2(x,y)%%x第二个特征的条件概率

re=0;

if (y==1||y==4||y==7)&&0<x(1,2)&&x(1,2)<3

re=1;

end

if (y==2||y==5||y==8)&&3.5<x(1,2)&&x(1,2)<6.5

re=1;

end

if (y==3||y==6||y==9)&&7<x(1,2)&&x(1,2)<10

re=1;

end

end

1. 绘制预测结果

代码：%%%%%%%%%%%%%%%%%% 结果可视化 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

figure(3)

set (gcf,'Position',[1,1,700,600], 'color','w')

set(gca,'Fontsize',18)

plot(X(Y==1,1),X(Y==1,2),'ro','LineWidth',1,'MarkerSize',10); % 画第一类数据点

hold on;

plot(X(Y==2,1),X(Y==2,2),'ko','LineWidth',1,'MarkerSize',10); % 画第二类数据点

hold on;

plot(X(Y==3,1),X(Y==3,2),'bo','LineWidth',1,'MarkerSize',10); % 画第三类数据点

hold on;

plot(X(Y==4,1),X(Y==4,2),'g\*','LineWidth',1,'MarkerSize',10); % 画第四类数据点

hold on;

plot(X(Y==5,1),X(Y==5,2),'b\*','LineWidth',1,'MarkerSize',10); % 画第五类数据点

hold on;

plot(X(Y==6,1),X(Y==6,2),'c\*','LineWidth',1,'MarkerSize',10); % 画第六类数据点

hold on;

plot(X(Y==7,1),X(Y==7,2),'b+','LineWidth',1,'MarkerSize',10); % 画第七类数据点

hold on;

plot(X(Y==8,1),X(Y==8,2),'r+','LineWidth',1,'MarkerSize',10); % 画第八类数据点

hold on;

plot(X(Y==9,1),X(Y==9,2),'k+','LineWidth',1,'MarkerSize',10); % 画第九类数据点

hold on;

plot(Xt(Ym==1,1),Xt(Ym==1,2),'ro','MarkerFaceColor','r','LineWidth',1,'MarkerSize',10); % 画第一类数据点

hold on;

plot(Xt(Ym==2,1),Xt(Ym==2,2),'ko','MarkerFaceColor','k','LineWidth',1,'MarkerSize',10); % 画第二类数据点

hold on;

plot(Xt(Ym==3,1),Xt(Ym==3,2),'bo','MarkerFaceColor','b','LineWidth',1,'MarkerSize',10); % 画第三类数据点

hold on;

plot(Xt(Ym==4,1),Xt(Ym==4,2),'go','MarkerFaceColor','g','LineWidth',1,'MarkerSize',10); % 画第四类数据点

hold on;

plot(Xt(Ym==5,1),Xt(Ym==5,2),'bo','MarkerFaceColor','b','LineWidth',1,'MarkerSize',10); % 画第五类数据点

hold on;

plot(Xt(Ym==6,1),Xt(Ym==6,2),'co','MarkerFaceColor','c','LineWidth',1,'MarkerSize',10); % 画第六类数据点

hold on;

plot(Xt(Ym==7,1),Xt(Ym==7,2),'bo','MarkerFaceColor','b','LineWidth',1,'MarkerSize',10); % 画第七类数据点

hold on;

plot(Xt(Ym==8,1),Xt(Ym==8,2),'ro','MarkerFaceColor','r','LineWidth',1,'MarkerSize',10); % 画第八类数据点

hold on;

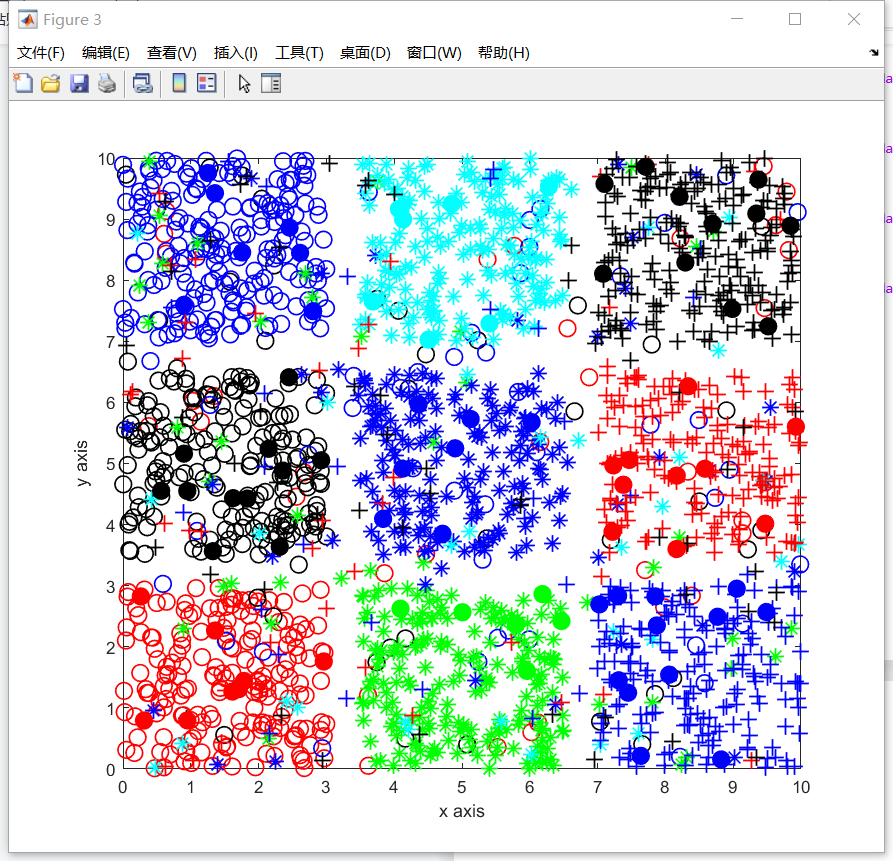
plot(Xt(Ym==9,1),Xt(Ym==9,2),'ko','MarkerFaceColor','k','LineWidth',1,'MarkerSize',10); % 画第九类数据点

hold on;

xlabel('x axis');

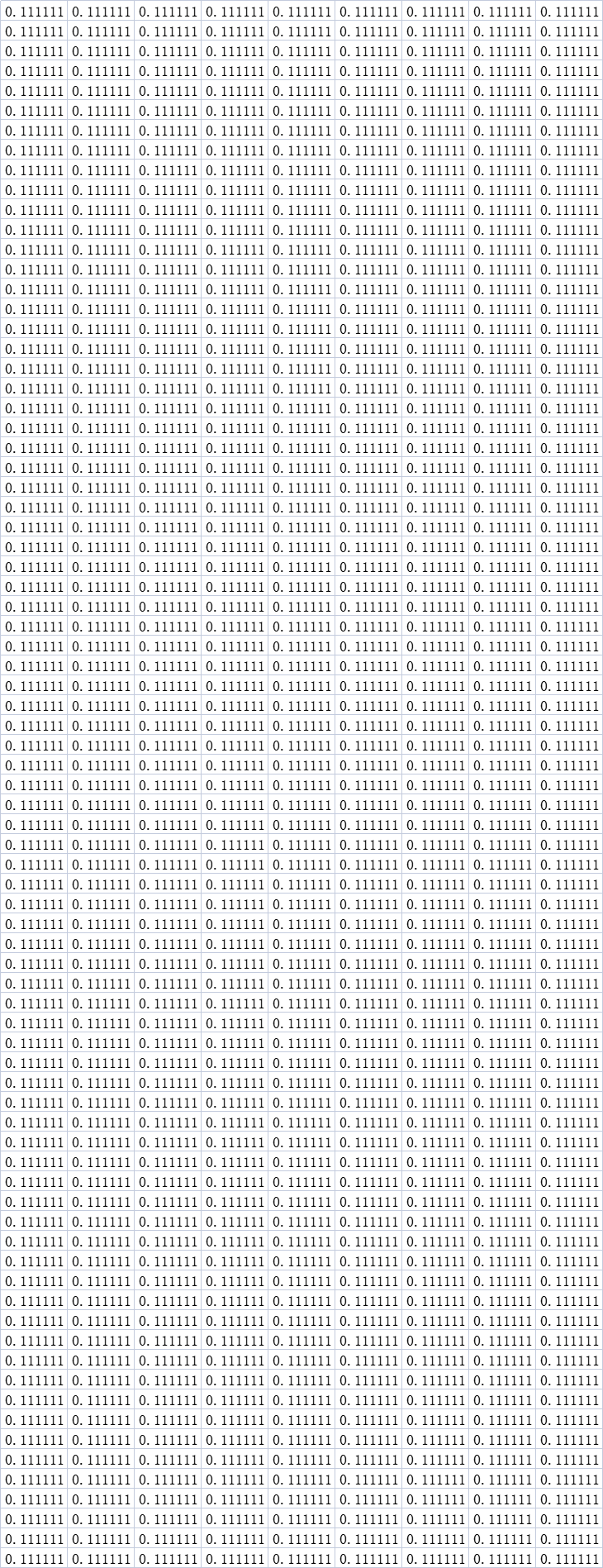
ylabel('y axis');

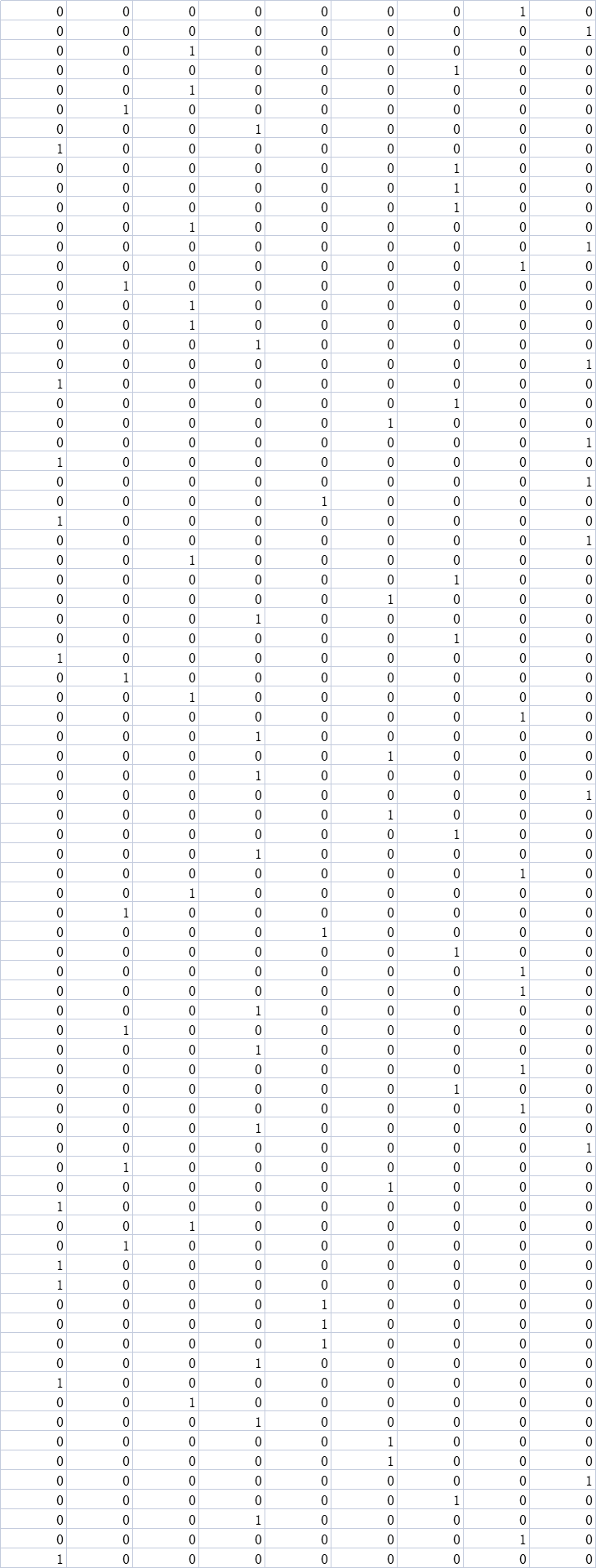
运行结果：



1. 先验概率和条件概率的表格

m行9列的表格，第i行第j列表示第i个测试数据对于类型j的概率



条件概率：

1. 附加题

增加噪声点，观察噪声比对算法的影响，报告观察结果；

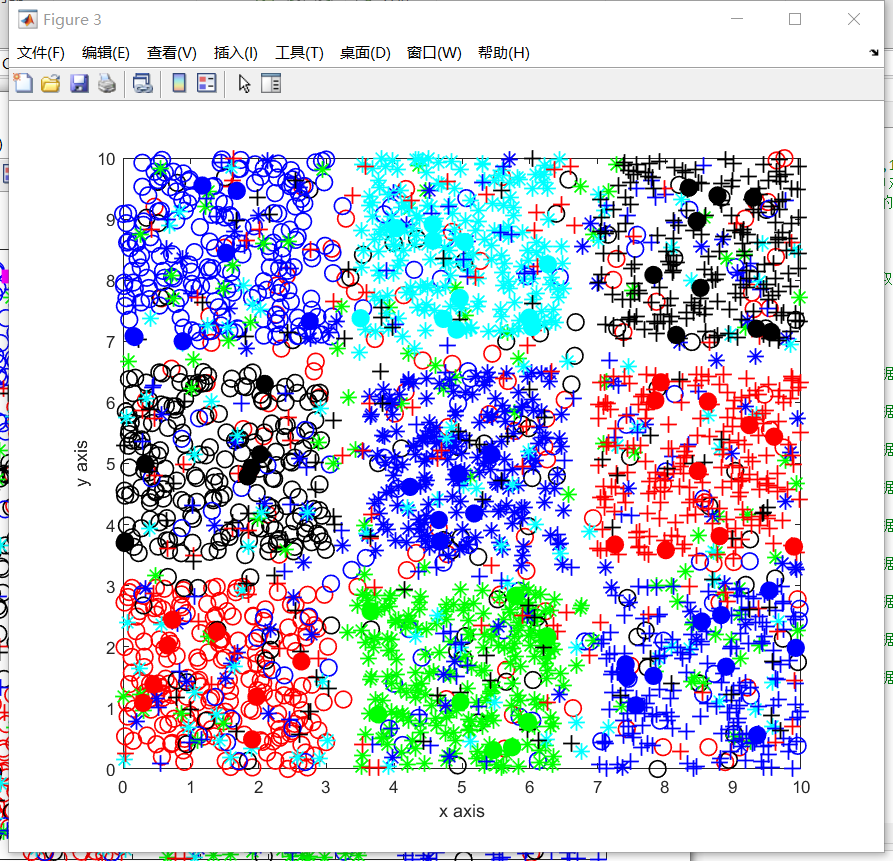
将： n=2000

改为：

N=2500

也即增加噪声：

运行结果：



观测发现噪声对于预测无明显影响

与K近邻模型的实验报告相对比，发现K近邻算法对噪声更敏感