

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

实验五：聚类

1. 实验目的

学习K均值聚类算法

1. 实验原理

K均值聚类算法：

• 初始化：随机选择k个样样本点作为初始类中心

• 对每个样本，计算其到每个聚类中心的距离，将每个样本指派到与其距离最近的类中心

的类中

• 计算新的类中心：对每一个类，计算当前类中的样本的均值，作为新的类中心

• 迭代，指派样本-计算类中心-指派样本-计算类中心-…，直到算法收敛

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)\*9; % 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

alpha = 0;

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+alpha<X(i,2) && X(i,2)<6+alpha

Y(i) = 2;

end

if 0<X(i,1) && X(i,1)<3 && 6+2\*alpha<X(i,2) && X(i,2)<9+2\*alpha

Y(i) = 3;

end

if 3+alpha<X(i,1) && X(i,1)<6+alpha && 0<X(i,2) && X(i,2)<3

Y(i) = 4;

end

if 3+alpha<X(i,1) && X(i,1)<6+alpha && 3+alpha<X(i,2) && X(i,2)<6+alpha

Y(i) = 5;

end

if 3+alpha<X(i,1) && X(i,1)<6+alpha && 6+2\*alpha<X(i,2) && X(i,2)<9+2\*alpha

Y(i) = 6;

end

if 6+2\*alpha<X(i,1) && X(i,1)<9+2\*alpha && 0<X(i,2) && X(i,2)<3

Y(i) = 7;

end

if 6+2\*alpha<X(i,1) && X(i,1)<9+2\*alpha && 3+alpha<X(i,2) && X(i,2)<6+alpha

Y(i) = 8;

end;

if 6+2\*alpha<X(i,1) && X(i,1)<9+2\*alpha && 6+2\*alpha<X(i,2) && X(i,2)<9+2\*alpha

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

n = length(Y); % 去除掉白色间隔剩下的点的个数

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

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;

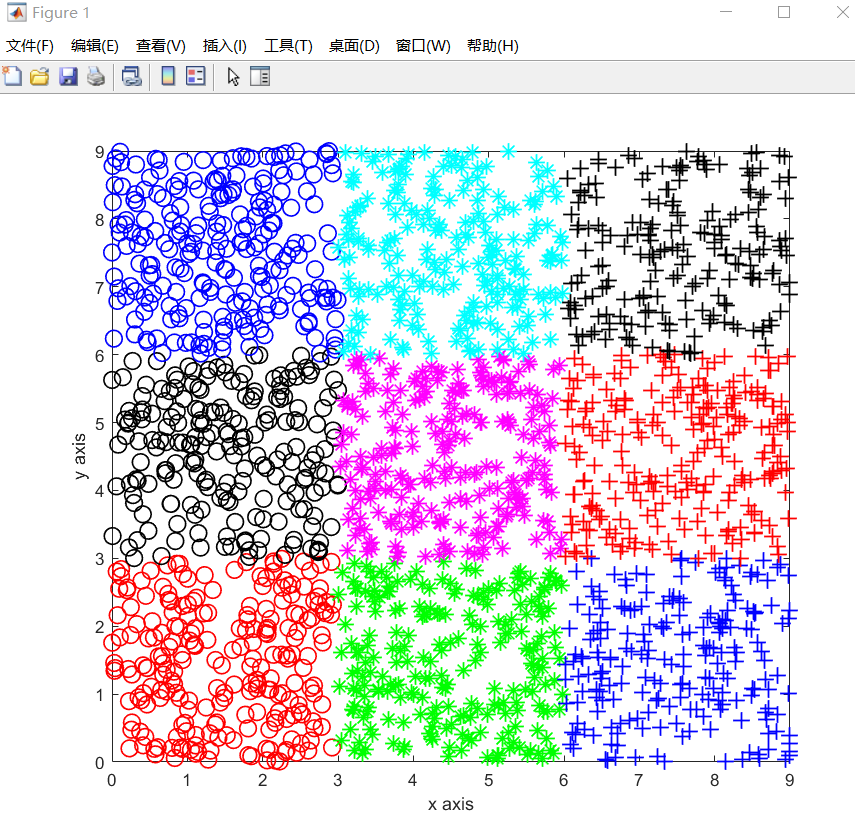
hold on;

xlabel('x axis');

ylabel('y axis');

clear Y; % 类别信息仅用与生成数据

运行结果：



1. 进行聚类分类

代码：

%%%%%%%%%%%%%%%%%% K-means算法：学生实现 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%% 画出聚类结果，注意类别信息Y是不能使用的 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

K = 9; % 中心点个数

Ym = zeros(n,1); % 每个数据点的预测输出类别

meanpoint = rand(K,2)\*10; % K个初始中心

h=2;

Xdis=zeros(n,9);%用来存储n个样本点对9个数据中心的距离

llq=0; %用来计数类别中心点迭代了多少次

for u=1:1:250 %设置迭代次数上限

for i=1:1:n

for j=1:1:9

Xdis(i,j)=(X(i,1)-meanpoint(j,1))^h+(X(i,2)-meanpoint(j,2))^h;%距离度量函数

Xdis(i,j)=Xdis(i,j)^(1/h);%%通过改变h的值来取用不同的距离度量函数

end

[P,Ym(i,1)]=min(Xdis(i,:));

for j=1:1:9 %对9个类别中心坐标进行更新

sum1=0.001;

sum2=0.001;

sumx1=0;

sumx2=0;

for p=1:1:n

if Ym(p,1)==j

sumx1=sumx1+X(p,1);

sum1=sum1+1;

sumx2=sumx2+X(p,2);

sum2=sum2+1;

end

end

if (meanpoint(j,1)~=sumx1/sum1)||(meanpoint(j,2)~=sumx2/sum2)

meanpoint(j,1)=sumx1/sum1;

meanpoint(j,2)=sumx2/sum2;

llq=llq+1; %完成一次类别中心坐标迭代，将计数器加一

else

break %%如果发现不再改变，则跳出循环

end

end

end

end

1. 画出分类结果

代码：%%%%%%%%%%%%%%%%%% 画出聚类结果及中心点 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

figure(2)

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

set(gca,'Fontsize',18)

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

hold on;

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

hold on;

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

hold on;

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

hold on;

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

hold on;

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

hold on;

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

hold on;

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

hold on;

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

hold on;

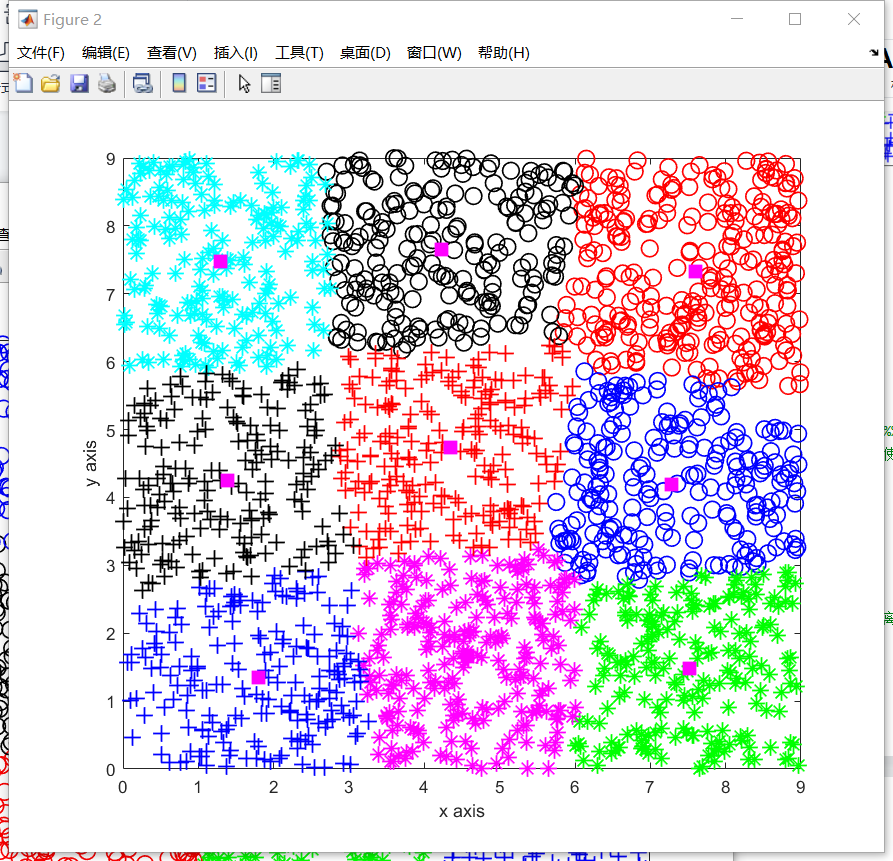
plot(meanpoint(:,1),meanpoint(:,2),'ms','MarkerFaceColor','m','LineWidth',1,'MarkerSize',10); % 画出中心点

hold on;

xlabel('x axis');

ylabel('y axis');

运行结果：



四．附加题：

选用不同的距离度量函数：

由代码

Xdis(i,j)=(X(i,1)-meanpoint(j,1))^h+(X(i,2)-meanpoint(j,2))^h;%距离度量函数

Xdis(i,j)=Xdis(i,j)^(1/h);%%通过改变h的值来取用不同的距离度量函数

知，改变h即可。

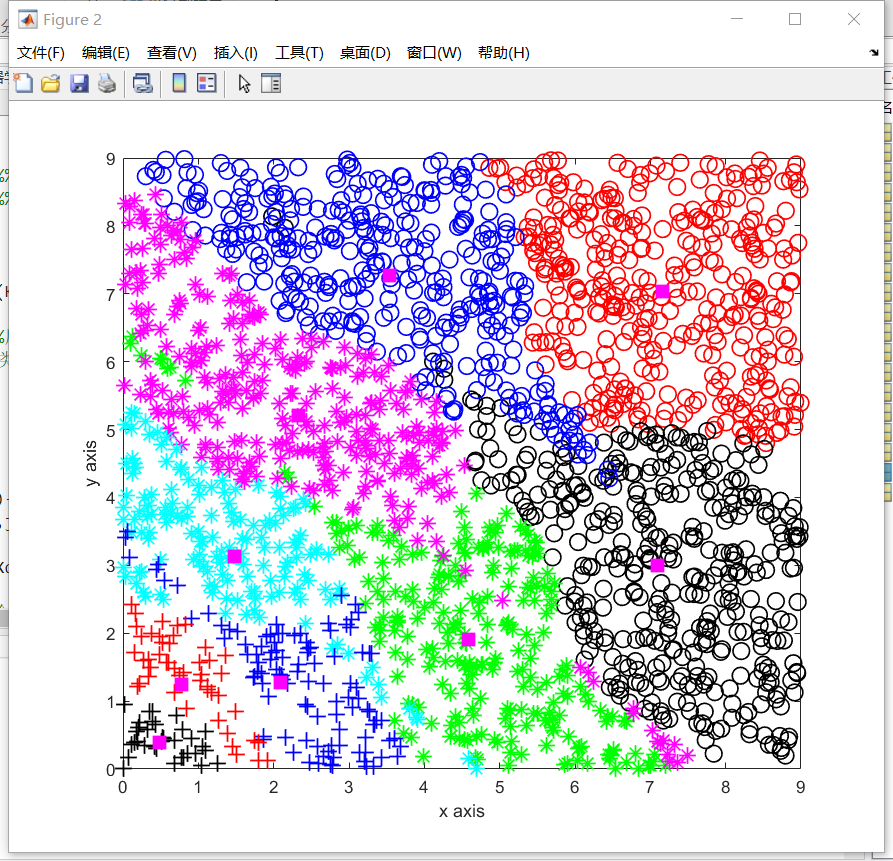
当p=1时，就是曼哈顿距离

当p=2时，就是[欧氏距离](https://so.csdn.net/so/search?q=%E6%AC%A7%E6%B0%8F%E8%B7%9D%E7%A6%BB&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/weixin_43935696/article/details/_blank)

当p→∞时，就是切比雪夫距离

取h=3

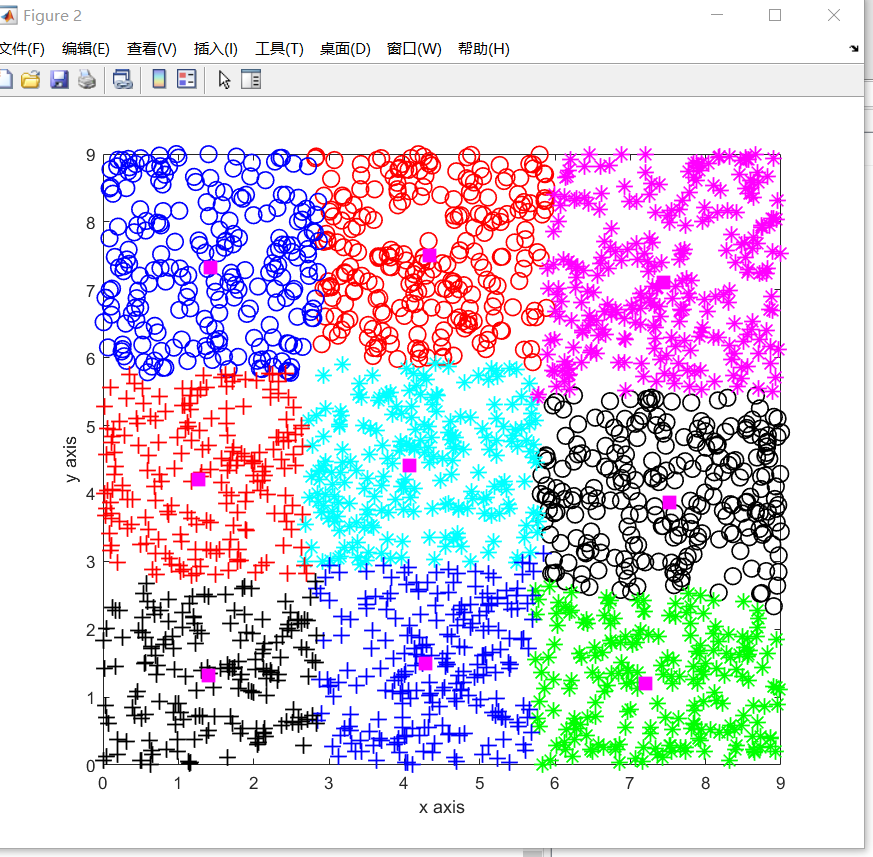
运行得：



发现举例效果有所下降

再取h=55

运行得：



发现聚类效果有显著提升。