数据挖掘实验报告

一、实验要求

学习并应用模糊C均值聚类算法处理数据。

二、实验题目

利用模糊C均值聚类算法对下列22维数据进行聚类分析,c=3, 加权指数m取2.

1 12 22 16 6 18 6 13 1 0 2 0 2 0 9 5 55 4 1 15 4 38

0 0 0 0 0 42 0 51 0 0 0 0 0 0 3 1 0 1 1 3 0 0

25 13 74 26 20 6 6 1 6 2 5 15 11 12 8 7 10 17 8 19 10 30

65 19 22 26 30 9 14 14 12 7 24 30 26 29 12 8 13 25 8 11 24 38

12 0 5 8 5 0 1 0 0 0 0 0 5 1 0 1 1 3 2 9 0 5

0 0 0 0 0 40 1 11 0 0 0 1 0 0 2 0 3 21 1 0 1 0

33 13 125 88 28 132 39 58 9 3 15 15 7 9 74 52 85 77 15 114 20 121

339 269 404 253 152 144 73 58 30 35 127 117 185 151 88 40 36 166 43 108 70 279

4 5 5 5 1 6 1 0 1 1 9 3 5 2 7 1 7 7 1 1 1 11

18 1 0 29 38 112 339 44 24 16 53 55 77 52 68 136 52 86 134 16 144 111

9 4 4 5 2 1 0 6 3 1 11 11 15 7 2 1 2 11 1 2 5 3

118 59 60 29 73 107 20 80 13 7 52 35 66 47 200 41 13 94 13 16 38 157

1 54 0 1 0 7 8 2 17 20 63 94 138 94 18 2 0 31 7 1 5 70

1 0 0 0 0 0 3 0 2 0 7 1 1 2 1 0 0 2 2 0 0 6

2725 87 124 62 312 145 48 101 12 10 39 45 36 45 56 25 30 89 25 41 16 66

88 91 140 265 187 176 119 128 64 62 189 213 210 153 170 71 86 250 97 96 79 307

876 863 2393 739 226 819 578 389 155 127 551 374 560 523 976 506 434 1428 257 704 547 1352

2 8 0 12 20 4 3 7 1 1 5 2 2 3 9 4 0 4 2 0 2 9

0 0 0 0 2 61 5 9 3 2 20 25 38 27 7 0 0 12 3 1 4 42

524 93 514 1085 507 37 661 115 237 165 802 542 806 712 105 40 8 93 231 158 120 128

179 2 11 766 0 261 14 49 1 0 6 1 2 7 124 92 192 151 19 96 693 191

三、实验代码

function [U,P,Dist,Cluster\_Res,Obj\_Fcn,iter]=fuzzycm(Data,C,plotflag,M,epsm)

% 模糊 C 均值聚类 FCM: 从随机初始化划分矩阵开始迭代

% [U,P,Dist,Cluster\_Res,Obj\_Fcn,iter] = fuzzycm(Data,C,plotflag,M,epsm)

% 输入:

% Data: N×S 型矩阵,聚类的原始数据,即一组有限的观测样本集,

% Data 的每一行为一个观测样本的特征矢量,S 为特征矢量

% 的维数,N 为样本点的个数

% C: 聚类数,1<C<N

% plotflag: 聚类结果 2D/3D 绘图标记,0 表示不绘图,为缺省值

% M: 加权指数,缺省值为 2

% epsm: FCM 算法的迭代停止阈值,缺省值为 1.0e-6

% 输出:

% U: C×N 型矩阵,FCM 的划分矩阵

% P: C×S 型矩阵,FCM 的聚类中心,每一行对应一个聚类原型

% Dist: C×N 型矩阵,FCM 各聚类中心到各样本点的距离,聚类中

% 心 i 到样本点 j 的距离为 Dist(i,j)

% Cluster\_Res: 聚类结果,共 C 行,每一行对应一类

% Obj\_Fcn: 目标函数值

% iter: FCM 算法迭代次数

% See also: fuzzydist maxrowf fcmplot

Data=[1 12 22 16 6 18 6 13 1 0 2 0 2 0 9 5 55 4 1 15 4 38;

0 0 0 0 0 42 0 51 0 0 0 0 0 0 3 1 0 1 1 3 0 0;

25 13 74 26 20 6 6 1 6 2 5 15 11 12 8 7 10 17 8 19 10 30;

65 19 22 26 30 9 14 14 12 7 24 30 26 29 12 8 13 25 8 11 24 38;

12 0 5 8 5 0 1 0 0 0 0 0 5 1 0 1 1 3 2 9 0 5;

0 0 0 0 0 40 1 11 0 0 0 1 0 0 2 0 3 21 1 0 1 0;

33 13 125 88 28 132 39 58 9 3 15 15 7 9 74 52 85 77 15 114 20 121;

339 269 404 253 152 144 73 58 30 35 127 117 185 151 88 40 36 166 43 108 70 279;

4 5 5 5 1 6 1 0 1 1 9 3 5 2 7 1 7 7 1 1 1 11;

18 1 0 29 38 112 339 44 24 16 53 55 77 52 68 136 52 86 134 16 144 111;

9 4 4 5 2 1 0 6 3 1 11 11 15 7 2 1 2 11 1 2 5 3;

118 59 60 29 73 107 20 80 13 7 52 35 66 47 200 41 13 94 13 16 38 157;

1 54 0 1 0 7 8 2 17 20 63 94 138 94 18 2 0 31 7 1 5 70;

1 0 0 0 0 0 3 0 2 0 7 1 1 2 1 0 0 2 2 0 0 6;

2725 87 124 62 312 145 48 101 12 10 39 45 36 45 56 25 30 89 25 41 16 66;

88 91 140 265 187 176 119 128 64 62 189 213 210 153 170 71 86 250 97 96 79 307;

876 863 2393 739 226 819 578 389 155 127 551 374 560 523 976 506 434 1428 257 704 547 1352;

2 8 0 12 20 4 3 7 1 1 5 2 2 3 9 4 0 4 2 0 2 9;

0 0 0 0 2 61 5 9 3 2 20 25 38 27 7 0 0 12 3 1 4 42;

524 93 514 1085 507 37 661 115 237 165 802 542 806 712 105 40 8 93 231 158 120 128;

179 2 11 766 0 261 14 49 1 0 6 1 2 7 124 92 192 151 19 96 693 191

];

C=3;

if nargin<5

epsm=1.0e-6;

end

if nargin<4

M=2;

end

if nargin<3

plotflag=0;

end

[N,S]=size(Data);m=2/(M-1);iter=0;

Dist(C,N)=0; U(C,N)=0; P(C,S)=0;

% 随机初始化划分矩阵

U0 = rand(C,N);

U0=U0./(ones(C,1)\*sum(U0));

% FCM 的迭代算法

while true

% 迭代计数器

iter=iter+1;

% 计算或更新聚类中心 P

Um=U0.^M;

P=Um\*Data./(ones(S,1)\*sum(Um'))';

% 更新划分矩阵 U

for i=1:C

for j=1:N

Dist(i,j)=fuzzydist(P(i,:),Data(j,:));

end

end

U=1./(Dist.^m.\*(ones(C,1)\*sum(Dist.^(-m))));

% 目标函数值: 类内加权平方误差和

if nargout>4 | plotflag

Obj\_Fcn(iter)=sum(sum(Um.\*Dist.^2));

end

% FCM 算法迭代停止条件

if norm(U-U0,Inf)<epsm

break

end

U0=U;

end

% 聚类结果

if nargout > 3

res = maxrowf(U);

for c = 1:C

v = find(res==c);

Cluster\_Res(c,1:length(v))=v;

end

end

% 绘图

if plotflag

fcmplot(Data,U,P,Obj\_Fcn);

end

function fcmplot(Data,U,P,Obj\_Fcn)

% FCM 结果绘图函数

% See also: fuzzycm maxrowf ellipse

[C,S] = size(P); res = maxrowf(U);

str = 'po\*x+d^v><.h';

% 目标函数绘图

figure(1),plot(Obj\_Fcn)

title('目标函数值变化曲线','fontsize',8)

% 2D 绘图

if S==2

figure(2),plot(P(:,1),P(:,2),'rs'),hold on

for i=1:C

v=Data(find(res==i),:);

plot(v(:,1),v(:,2),str(rem(i,12)+1))

ellipse(max(v(:,1))-min(v(:,1)), ...

max(v(:,2))-min(v(:,2)), ...

[max(v(:,1))+min(v(:,1)), ...

max(v(:,2))+min(v(:,2))]/2,'r:')

end

grid on,title('2D 聚类结果图','fontsize',8),hold off

end

% 3D 绘图

if S>2

figure(2),plot3(P(:,1),P(:,2),P(:,3),'rs'),hold on

for i=1:C

v=Data(find(res==i),:);

plot3(v(:,1),v(:,2),v(:,3),str(rem(i,12)+1))

ellipse(max(v(:,1))-min(v(:,1)), ...

max(v(:,2))-min(v(:,2)), ...

[max(v(:,1))+min(v(:,1)), ...

max(v(:,2))+min(v(:,2))]/2, ...

'r:',(max(v(:,3))+min(v(:,3)))/2)

end

grid on,title('3D 聚类结果图','fontsize',8),hold off

end

function D=fuzzydist(A,B)

% 模糊聚类分析: 样本间的距离

% D = fuzzydist(A,B)

D=norm(A-B);

function mr=maxrowf(U,c)

% Çó¾ØÕó U Ã¿ÁÐµÚ c ´óÔªËØËùÔÚÐÐ,c µÄÈ±Ê¡ÖµÎª 1

% µ÷ÓÃ¸ñÊ½: mr = maxrowf(U,c)

% See also: addr

if nargin<2

c=1;

end

N=size(U,2);mr(1,N)=0;

for j=1:N

aj=addr(U(:,j),'descend');

mr(j)=aj(c);

end

function ellipse(a,b,center,style,c\_3d)

% 绘制一个椭圆

% 调用: ellipse(a,b,center,style,c\_3d)

% 输入:

% a: 椭圆的轴长(平行于 x 轴)

% b: 椭圆的轴长(平行于 y 轴)

% center: 椭圆的中心 [x0,y0],缺省值为 [0,0]

% style: 绘制的线型和颜色,缺省值为实线蓝色

% c\_3d: 椭圆的中心在 3D 空间中的 z 轴坐标,可缺省

if nargin<4

style='b';

end

if nargin<3 | isempty(center)

center=[0,0];

end

t=1:360;

x=a/2\*cosd(t)+center(1);

y=b/2\*sind(t)+center(2);

if nargin>4

plot3(x,y,ones(1,360)\*c\_3d,style)

else

plot(x,y,style)

end

function f = addr(a,strsort)

% 返回向量升序或降序排列后各分量在原始向量中的索引

% 函数调用:f = addr(a,strsort)

% strsort: 'ascend' or 'descend'

% default is 'ascend'

% -------- example --------

% addr([ 4 5 1 2 ]) returns ans:

% [ 3 4 1 2 ]

if nargin==1

strsort='ascend';

end

sa=sort(a); ca=a;

la=length(a);f(la)=0;

for i=1:la

f(i)=find(ca==sa(i),1);

ca(f(i))=NaN;

end

if strcmp(strsort,'descend')

f=fliplr(f);

end

四、实验结果

