## 模式识别实验五

实验人:叶平

实验内容:

## P. 475-476, Probs. 2-3

- 1、写程序实现 k-均值算法(算法 1)和模糊 k-均值聚类算法(算法 2)并用表中的三维数据进行测试。其中距离采用 Euclid 距离  $d(x,y)=\|x-y\|$ 。
  - (a) c=2, m1 (0) = (1,1,1) ', m2=(-1,1,-1) '.
- (b) c=2, m1 (0) = (0,0,0) ', m2=(1,1,-1) '。将得到的结果与(a) 中的结果进行比较,并解释差别,包括迭代次数的差别。
  - (c) c=3, m1 (0) = (0,0,0) ', m2 (0) = (1,1,1) ', m3 (0) = (-1,0,2) '
- (d) c=3,m1 (0) = (-0.1,0,0.1) ',m2 (0) = (0,-0.1,0.1) ',m3 (0) = (-0.1,-0.1,0.1) '。 将得到的结果与 (c) 中的结果进行比较,并解释差别,包括迭代次数的差别。
  - 2、重做 1,其中距离改为 d(x,y)^2=1-exp(- $\beta*$   $\|x-y\|^2$ )。其中 β 分别取 0.001, 0.01, 0.1, 1, 10, 100.

## 实验结果:

在 Command Window 下运行实验 5 文件夹下 Res.m 文件。其中 Res 函数有三个参数,Res(ForK,s,beta),前两个必填,后一个选填。第一个参数为 1 表示 K-means,否则 FCM,第二个参数选择数据中心,有 1 2 3 4 四个选项,对应四问,第三参数不填则为欧式距离,填则为给定度量函数的参数 beta 值。

用 Kmeans 得到的 (a)(b)的结果如下:

采用K-means)	⊈法,使用図	大野悪度軍	>> Res(1,2) 采用K-means算法,使用欧式距离度				
分类中心为:	1,000	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	分类中心为:				
center =			center =				
5.5845	-0.0418	1.6600	-6.3082	0.4152	-0.732		
-6.3991	0.5061	-0.8236	5. 5 <mark>8</mark> 45	-0.0418	1.478		
第一类的数据	有:		第一类的数据	有:			
ans =			ans =				
4.3600	-2.1900	2.0900	-7.8200	-4.5800	-3.970		
6.7200	0.8800	2.8000	-6. 6800	3.1600	2.710		
4.4700	-2.6200	5.7600	-8.6400	3.0600	3.500		
6.7300	-2.0100	4.1800	-6.8700	0.0570	-5.450		
6.1800	2.8100	5.8200	-7.7100	2.3400	-6.330		
6.7200	-0.9300	-4.0400	-6.9100	-0.4900	-5.680		
8.0900	0.2000	2.2500	-6.2500	-0.2600	0.560		
6.8100	0.1700	-4.1500	-6.9400	-1.2200	1.130		
4.0800	1.3000	5. 3300	-5.1900	4.2400	4.040		
6. 2700	0.9300	-2.7800	-6.3800	-1.7400	1.430		
第二类的数据	有:		第二类的数据	有:			
ans =			ans =				
-7.8200	-4.5800	-3.9700	4.3600	-2.1900	2.090		
-6.6800	3.1600	2.7100	6.7200	0.8800	2.800		
-8.6400	3.0600	3.5000	4.4700	-2.6200	5.760		
-6.8700	0.0570	-5.4500	6.7300	-2.0100	4.180		
-7.7100	2.3400	-6.3300	6.1800	2.8100	5.820		
-6. 9100	-0.4900	-5.6800	6.7200	-0.9300	-4.040		
-6.2500	-0.2600	0.5600	8.0900	0.2000	2.250		
-6.9400	-1.2200	1.1300	6.8100	0.1700	-4.150		
-5.1900	4.2400	4.0400	4.0800	1.3000	5.330		
-6.3800	-1.7400	1.4300	6.2700	0.9300	-2.780		
未分开的数据	有:		未分开的数据有:				
ans =			ans =				
[]			[]				

由上图可以看出,分类的数据结果几乎一样,除了类别反过来,除去初始给定的聚类中心,结果一样。可见 Kmeans 的分类和初始类中心的选择有关。

用 Kmeans 得到的 (c)(d)的结果如下:

>> Res(1,3)			>> Res(1,4) 采用K-means:	首法,使用网	なっています。
采用K-means)	算法,使用图	式距离度重	分类中心为:	#/A / [X](]E)	V14110141175 ±
分类中心为:			center =		
center =			-		
			-0.1000	0	0.1000
-5.8620	-0.5346	-4.2860	5. 4936	1070	
5.5845	-0.0418	1.6600		0.4061	
-5.8686	1.0343	2. 1957	17,000 900,000	31.5145	
			第一类的数据	有:	
第一类的数据 <sup>。</sup>	有:		ans =		
ans =					
			Empty ma	trix: 0-by	-3
-7.8200	-4.5800	-3.9700			
-6.8700	0.0570	-5.4500	第二类的数据	有:	
-7.7100	2.3400	-6.3300	ans =		
-6.9100	-0.4900	-5.6800			
			4.3600	-2.1900	2.0900
第二类的数据	有:		6.7200	0.8800	2.8000
ans =			4.4700	-2.6200	5.7600
			6. 7300	-2.0100	4.1800
	-2.1900	2.0900	6.1800	2.8100	5.8200
6.7200	0.8800	2.8000	6.7200	-0.9300	-4.0400
4.4700	-2.6200	5.7600	8. 0900	0.2000	2.2500
6.7300	-2.0100	4.1800	6.8100	0.1700	-4.1500
6. 1800	2.8100	5.8200	4.0800	1.3000	5.3300
6.7200	-0.9300	-4.0400	6.2700	0.9300	-2.7800
8.0900	0.2000	2. 2500			
6.8100	0.1700	-4.1500	第三类的数据	有:	
4.0800	1.3000	5. 3300	ans =		
6.2700	0.9300	-2.7800			
Y 100	0.00		-7.8200	-4.5800	-3.9700
第三类的数据	有:		-6. 6800	3.1600	2.7100
ans =			-8.6400	3.0600	3.5000
			-6.8700	0.0570	-5.4500
-6.6800	3.1600	2.7100	-7.7100	2.3400	-6.3300
-8.6400	3.0600	3.5000	-6.9100	-0.4900	-5.6800
-6.2500	-0.2600	0.5600	-6.2500	-0.2600	0.5600
-6.9400	-1.2200	1.1300	-6.9400	-1.2200	1.1300
-5. 1900	4. 2400	4.0400	-5.1900	4.2400	4.0400
-6.3800	-1.7400	1.4300	-6.3800	-1.7400	1.4300

由上图可以看出,由于初始的中心有很大的差别,结果导致分类的结果大大的不同。还 是说明 Kmeans 太过依赖于初始的聚类中心和类别数量。 用 FCM 算法得到的 (a)(b)的结果如下:

>> Res(2,1) 采用FCM算法,	使用欧式距	密度量:	>> Res(2,2) 采用FCM算法,使用欧式距离度量:				
迭代次数为:1		.147.5.1					
分类中心为:			迭代次数为:18 分类中心为:				
center =			の条件心内・ center =				
			center -				
4.5239	-0.0530	1.6204	-5.4755	0.2384	-0.507		
-5. <b>47</b> 57	0.2384	-0.5074		-0.0531			
第一类的数据	有:		第一类的数据	有:			
ans =			ans =				
4.3600	-2.1900	2.0900	/// <b>/</b>				
6.7200	0.8800	2.8000	-7. 8200	-4.5800			
4.4700	-2.6200	5.7600	-6.6800	3. 1600	2.710		
6. 7300	-2.0100	4. 1800	-8.6400	3.0600	3.500		
6. 1800	2.8100	5.8200	-6.8700	0.0570	-5.450		
6. 7200	-0.9300	-4.0400	-7.7100	2.3400	-6.330		
8. 0900	0.2000	2. 2500	-6.9100	-0.4900	-5.680		
6.8100	0.1700	-4. 1500	-6. 2500	-0.2600	0.560		
4. 0800	1.3000	5. 3300	-6.9400	-1.2200	1.130		
			-5. 1900	4.2400	4.040		
6. 2700	0.9300	-2.7800	-6.3800	-1.7400	1.430		
第二类的数据	有:		第二类的数据	有:			
ans =			ans =				
-7.8200	-4.5800	-3.9700	4.3600	-2.1900	2.090		
-6.6800	3.1600	2.7100	6.7200	0.8800	2.800		
-8.6400	3.0600	3.5000	4.4700	-2.6200	5.760		
-6.8700	0.0570	-5.4500	6.7300	-2.0100	4.180		
-7.7100	2.3400	-6.3300	6.1800	2.8100	5.820		
-6.9100	-0.4900	-5.6800	6.7200	-0.9300	-4.040		
-6.2500	-0.2600	0.5600	8.0900	0.2000	2.250		
-6.9400	-1.2200	1.1300	6.8100	0.1700	-4.150		
-5.1900	4.2400	4.0400	4.0800	1.3000	5. 330		
-6.3800	-1.7400	1.4300	6.2700	0.9300	-2.780		

由上图可以看出,分类的结果和 Kmeans 一样。除了聚类中心不同之外。

用 FCM 算法得到的 (c)(d)的结果如下:

>> Res(2, 3)			>> Res(2,4)				
采用FCM算法	,使用欧式距	离度里:	采用FCM算法	采用FCM算法,使用欧式距离度量:			
迭代次数为:5	58		迭代次数为:41				
分类中心为:			分类中心为:				
center =			center =				
-4.3675	0.2863	-0.4392	-4.3836	0.2848	-0.431		
5.1762	-0.1179	2.0186	5. 1762	-0.1179	2.018		
-4.3838	0.2849	-0.4315	-4.3676	0.2863	-0.439		
第一类的数据	有:		第一类的数据	有:			
ans =			ans =				
Empty ma	trix: 0-by	-3	-7.8200	-4.5800	-3.970		
			-6.6800	3.1600	2.710		
第二类的数据	有:		-8.6400	3.0600	3.500		
ans =			-6.8700	0.0570	-5.450		
			-7.7100	2.3400	-6.330		
4.3600	-2.1900	2.0900	-6.9100	-0.4900	-5.680		
6.7200	0.8800	2.8000	-6.2500	-0.2600	0.560		
4.4700	-2.6200	5.7600	-6.9400	-1.2200	1.130		
6.7300	-2.0100	4.1800	-5. 1900	4.2400	4.040		
6.1800	2.8100	5.8200	-6.3800	-1.7400	1.430		
6.7200	-0.9300	-4.0400					
8.0900	0.2000	2.2500	第二类的数据	精:			
6.8100	0.1700	-4.1500	ans =				
4.0800	1.3000	5.3300					
6.2700	0.9300	-2.7800	4.3600	-2.1900	2.090		
			6.7200	0.8800	2.800		
第三类的数据	有:		4.4700	-2.6200	5.760		
ans =			6.7300	-2.0100	4.180		
			6.1800	2.8100	5.820		
-7.8200	-4.5800	-3.9700	6.7200	-0.9300	-4.040		
-6.6800	3.1600	2.7100	8.0900	0.2000	2.250		
-8.6400	3.0600	3.5000	6.8100	0.1700	-4.150		
-6.8700	0.0570	-5.4500	4.0800	1.3000	5.330		
-7.7100	2.3400	-6.3300	6.2700	0.9300	-2.780		
-6.9100	-0.4900	-5.6800					
-6.2500	-0.2600	0.5600	第三类的数据	精:			
-6.9400	-1.2200	1.1300	ans =				
-5.1900	4.2400	4.0400					
-6.3800	-1.7400	1.4300	fx Empty ma	trix: 0-by	-3		

由上图可以看出,FCM 算法的分类结果和 Kmeans 有很大的不同,在 FCM 中,第一类和第三类没有数据,分类结果和 k=2 时一致,表明 FCM 不依赖初始聚类中心和聚类数目。

当采用 d(x,y)^2=  $1-\exp(-\beta*\left\|x-y\right\|^2)$ 作为度量时。

Beta 取 0.001 时,kmeans 算法:

>> Res(1,1,		<b>六</b>					
	算法 <b>,</b> 使用给	定度量函数,beta=0.001000:	>> Res(1,1)				
分类中心为:			采用K-means算法,使用欧式距离度				
center =			分类中心为:		(3-40-141)3		
5.5845	-0.0418	1.6600	center =				
-6.3991	0.5061	-0.8236	5 5845	-0.0418	1 660		
	123		-6.3991	0.5061			
第一类的数据	有:				ARREST SE		
ans =			第一类的数据	有:			
F 2 F 2 F 2 F 2 F 2	121 (121/2)	12 B250100	ans =				
4.3600	-2.1900	2.0900					
6,7200	0.8800	2.8000	4.3600	-2.1900	2.090		
4.4700	-2.6200	5. 7600	6.7200	0.8800	2.800		
6.7300	-2.0100	4.1800	4.4700	-2.6200	5.760		
6.1800	2.8100	5. 8200	6.7300	-2.0100	4.180		
6.7200	-0.9300	-4.0400	6.1800	2.8100	5.820		
8.0900	0.2000	2. 2500	6.7200	-0.9300	-4.040		
6.8100	0.1700	-4.1500	8.0900	0.2000	2. 250		
4.0800	1.3000	5. 3300	6.8100	0.1700	-4.150		
6,2700	0.9300	-2.7800	4.0800	1.3000	5. 330		
			6.2700	0.9300	-2.780		
第二类的数据	有:						
ans =			第二类的数据	有:			
			ans =				
-7.8200	-4.5800	-3.9700					
-6.6800	3.1600	2.7100	-7.82 <b>0</b> 0	-4.5800	-3.970		
-8.6400	3.0600	3.5000	-6.6800	3. 1600	2.710		
-6.8700	0.0570	-5.4500	-8.6400	3.0600	3.500		
-7.7100	2.3400	-6.3300	-6.8700	0.0570	-5.450		
-6.9100	-0.4900	-5.6800	-7.7100	2.3400	-6.330		
-6.2500	-0.2600	0.5600	-6.9100	-0.4900			
-6.9400	-1.2200	1.1300	-6. 2500	-0.2600	0.560		
-5.1900	4.2400	4.0400	-6.9400				
-6.3800	-1.7400	1.4300	-5.1900	4. 2400	4.040		
			-6.3800	-1.7400	1.430		
未分开的数据	有:		未分开的数据	右•			
ans =				П.			
			ans =				
[]			[]				

可以看出和欧式度量结果一样。事实上,beta 取 0.01 0.1 时结果都一样。而 beta 取 1 时则有些分不出来,当 beta 取 100 时则完全分不出来了。如下图:

				采用K-means算法,使用给定度量函数,beta=100.0000		
>> Res(1, 1,				分类中心为:		
100 70	算法,使用给	定度量函数, b	eta=1.000000:	center =		
分类中心为:						
center =				1	1 1	
				-1	1 -1	
	-0.0787					
-5.1520	0.4040	1.2320		第一类的数据	有:	
				ans =		
第一类的数据	有:					
ans =				Empty ma	trix: 0-by	-3
4.3600	-2.1900	2.0900		第二类的数据	有:	
6.7200	0.8800	2.8000		ans =	(673)	
4.4700	-2.6200	5.7600		**************************************		
6.7300	-2.0100	4.1800		Empty ma	trix: 0-by	-3
6.1800	2.8100	5.8200				
8.0900	0.2000	2.2500		未分开的数据	有:	
4.0800	1.3000	5.3300		ans =		
第二类的数据	有:			-7. <mark>8200</mark>	-4.5800	-3.9700
ans =					3.1600	2.7100
					-2.1900	
-6.2500	-0.2600	0.5600			0.8800	
-6.9400	-1.2200	1.1300			3.0600	
-5.1900	4.2400	4.0400			0.0570	
-6.3800	-1.7400	1.4300		4.4700	-2.6200	5.7600
				6.7300	-2.0100	4.1800
未分开的数据	有:			-7.7100	2.3400	
ans =				-6.9100	-0.4900	-5. 6800
				6.1800	2.8100	5.8200
-7.8200	-4.5800	-3.9700		6.7200	-0.9300	-4.0400
-6.6800	3.1600	2.7100		-6.2500	-0.2600	0.5600
-8.6400	3.0600	3.5000		-6.9400	-1.2200	1.1300
-6.8700	0.0570	-5.4500		8.0900	0.2000	2. 2500
-7.7100	2.3400	-6.3300		6.8100	0.1700	-4.1500
-6.9100	-0.4900	-5.6800		-5.1900	4.2400	4.0400
6.7200	-0.9300	-4.0400		-6.3800	-1.7400	1.4300
6.8100	0.1700	-4.1500		4.0800	1.3000	5. 3300
6.2700	0.9300	-2.7800		6.2700	0.9300	-2.7800

在 FCM 算法中,我们查看(c)中的分类情况如下:

				>> Res(2, 3,	0.1)			
>> Res (2, 3,				采用FCM算法,使用给定度量函数,beta=0.10000				
		里函数,beta=0.0	01(	迭代次数为:5				
迭代次数为:1	8			分类中心为:				
分类中心为:				center =				
center =								
				-0.4480	0.1554	0.4600		
-6.8795	-0.3199	-5.2090		-0.4479	0.1553	0.4600		
6.0529	-0.1745	2.2109		-0.4481	0.1554	0.4600		
-6.4454	1.2404	2.1240						
oun mornoussassion	02000			第一类的数据	有:			
第一类的数据	有:			ans =				
ans =								
				Empty ma	trix: 0-by	-3		
-7 <b>.</b> 8200	-4.5800	-3.9700						
-6.8700	0.0570	-5.4500		第二类的数据	有:			
-7.7100	2.3400	-6.3300		ans =				
-6.9100	-0.4900	-5.6800						
				4.3600	-2.1900	2.0900		
第二类的数据	有:			6.7200	0.8800	2.8000		
ans =				4.4700	-2.6200	5.7600		
				6.7300	-2.0100	4.1800		
4.3600	-2.1900	2.0900		6.1800	2.8100	5.8200		
6.7200	0.8800	2.8000		6.7200	-0.9300	-4.0400		
4.4700	-2.6200	5.7600		8.0900	0.2000	2. 2500		
6.7300	-2.0100	4.1800		6.8100	0.1700	-4.1500		
6.1800	2.8100	5.8200		4.0800	1.3000	5. 3300		
6.7200	-0.9300	-4.0400		6.2700	0.9300	-2.7800		
8.0900	0.2000	2.2500						
6.8100	0.1700	-4.1500		第三类的数据	有:			
4.0800	1.3000	5.3300		ans =				
6.2700	0.9300	-2.7800						
				-7.8200	-4.5800	-3.9700		
第三类的数据	有:			-6.6800	3.1600	2.7100		
ans =				-8.6400	3.0600	3.5000		
				-6.8700	0.0570	-5.4500		
-6.6800	3.1600	2.7100		-7.7100	2.3400	-6.3300		
-8.6400	3.0600	3.5000		-6.9100	-0.4900	-5.6800		
-6.2500	-0.2600	0.5600		-6, 2500	-0.2600	0.5600		
-6.9400	-1.2200	1.1300		-6.9400	-1.2200	1.1300		
-5.1900	4.2400	4.0400		-5. 1900	4. 2400	4.0400		
-6.3800	-1.7400	1.4300	fx	-6.3800	-1.7400	1.4300		

令人奇怪的是 beta 取 0.1 时结果与欧式度量一样。而 beta 取 0.001 时则反而有些分错,而当 beta 取 1 和 100 时如下图,可以看出,两者都已经完全分不出来了。

	>> Res(2, 3,	1)			>> Res(2, 3, 100)					
			里函数,beta=1	1.0	采用FCM算法,使用给定度里函数,beta=100.					
	迭代次数为:3				迭代次数为:3					
	分类中心为:				分类中心为:					
	center =				center =					
	-0.4480	0.1554	0.4600		-0.4480	0.1554	0.4600			
	-0.4480	0.1554	0.4600		-0.4480	0.1554	0.4600			
	-0.4480	0.1554	0.4600		-0.4480	0.1554	0.4600			
	第一类的数据	有:			第一类的数据	有:				
	ans =	1.4			ans =					
	THE STATE OF THE S									
	Empty ma	trix: 0-by-	-3		Empty mat	rix: 0-by	-3			
	77 - W 45 11 10				第二类的数据	有:				
	第二类的数据	<b>有:</b>			ans =					
	ans =									
	Empty mat	trix: 0-by-	-3		Empty matrix: 0-by-3					
					Jahr - Star & Cald Com.	1.000				
	第三类的数据有:				第三类的数据有:					
	ans =				ans =					
					Ennter not	rix: 0-by	_2			
	Empty ma	trix: 0-by-	-3		Емріу ма	.rix: 0-by	-3			
	未分开的数据	有:			未分开的数据有:					
	ans =	****			ans =					
	-7.8200	-4.5800	-3.9700		-7.8200	-4.5800	-3.9700			
	-6.6 <mark>80</mark> 0	3.1600	2.7100		-6.6800	3.1600	2.7100			
	4.3600	-2.1900	2.0900		4.3600	-2.1900	2.0900			
	6.7200	0.8800	2.8000		6.7200	0.8800	2.8000			
	-8.6400	3.0600	3.5000		-8.6400	3.0600	3.5000			
	-6.8700	0.0570	-5.4500		-6.8700	0.0570	-5. 4500			
	4.4700	-2.6200	5.7600		4.4700	-2.6200	5. 7600			
	6.7300	-2.0100	4.1800		6.7300	-2.0100	4. 1800			
	-7.7100	2.3400	-6.3300		-7.7100	2.3400	-6.3300			
	-6.9100	-0.4900	-5.6800		-6.9100	-0.4900	-5.6800			
	6.1800	2.8100	5.8200		6. 1800	2.8100	5.8200			
	6.7200	-0.9300	-4.0400		6.7200	-0.9300	-4.0400			
£	-6.2500	-0.2600	0.5600	fx	-6.2500	-0.2600	0.5600			
fx	-6.9400	-1.2200	1.1300	J.\$	-6.9400	-1.2200	1.1300			

综上所述,我们可以看出 Kmeans 算法依赖于初始的聚类中心和聚类数量,若给的不恰当, Kmeans 得出的结果差异很大,即 Kmeans 算法对初始条件敏感。而 FCM 算法不依赖于初始的聚类中心甚至是聚类数量,算法有着很好的鲁棒性。收敛的速度也非常快,在实验中所给定的数据条件下,迭代次数没有超过 100 次。

但是,Kmeans 算法和 FCM 算法都依赖与度量函数。当度量函数不同时,甚至是其中的参数有改变时,得出的结果也不相同。在实验所给定的度量函数中,通过实验发现,当 beta 取 0.1 时,实验的结果和欧式距离相同。