

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) $c=2$, $m_1(0) = (1,1,1)'$, $m_2 = (-1,1,-1)'$ 。

(b) $c=2$, $m_1(0) = (0,0,0)'$, $m_2 = (1,1,-1)'$ 。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) $c=3$, $m_1(0) = (0,0,0)'$, $m_2(0) = (1,1,1)'$, $m_3(0) = (-1,0,2)'$

(d) $c=3$, $m_1(0) = (-0.1,0,0.1)'$, $m_2(0) = (0,-0.1,0.1)'$, $m_3(0) = (-0.1,-0.1,0.1)'$ 。

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:

```
>> Res(1,1)
```

采用K-means算法, 使用欧式距离度量
分类中心为:

center =

5.5845	-0.0418	1.6600
-6.3991	0.5061	-0.8236

第一类的数据有:

ans =

4.3600	-2.1900	2.0900
6.7200	0.8800	2.8000
4.4700	-2.6200	5.7600
6.7300	-2.0100	4.1800
6.1800	2.8100	5.8200
6.7200	-0.9300	-4.0400
8.0900	0.2000	2.2500
6.8100	0.1700	-4.1500
4.0800	1.3000	5.3300
6.2700	0.9300	-2.7800

第二类的数据有:

ans =

-7.8200	-4.5800	-3.9700
-6.6800	3.1600	2.7100
-8.6400	3.0600	3.5000
-6.8700	0.0570	-5.4500
-7.7100	2.3400	-6.3300
-6.9100	-0.4900	-5.6800
-6.2500	-0.2600	0.5600
-6.9400	-1.2200	1.1300
-5.1900	4.2400	4.0400
-6.3800	-1.7400	1.4300

未分开的数据有:

ans =

[]

```
>> Res(1,2)
```

采用K-means算法, 使用欧式距离度量
分类中心为:

center =

-6.3082	0.4152	-0.7327
5.5845	-0.0418	1.4782

第一类的数据有:

ans =

-7.8200	-4.5800	-3.9700
-6.6800	3.1600	2.7100
-8.6400	3.0600	3.5000
-6.8700	0.0570	-5.4500
-7.7100	2.3400	-6.3300
-6.9100	-0.4900	-5.6800
-6.2500	-0.2600	0.5600
-6.9400	-1.2200	1.1300
-5.1900	4.2400	4.0400
-6.3800	-1.7400	1.4300

第二类的数据有:

ans =

4.3600	-2.1900	2.0900
6.7200	0.8800	2.8000
4.4700	-2.6200	5.7600
6.7300	-2.0100	4.1800
6.1800	2.8100	5.8200
6.7200	-0.9300	-4.0400
8.0900	0.2000	2.2500
6.8100	0.1700	-4.1500
4.0800	1.3000	5.3300
6.2700	0.9300	-2.7800

未分开的数据有:

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)
采用K-means算法，使用欧式距离度量
分类中心为：
center =

    -5.8620    -0.5346    -4.2860
     5.5845    -0.0418     1.6600
    -5.8686     1.0343     2.1957

第一类的数据有：
ans =

    -7.8200    -4.5800    -3.9700
    -6.8700     0.0570    -5.4500
    -7.7100     2.3400    -6.3300
    -6.9100    -0.4900    -5.6800

第二类的数据有：
ans =

     4.3600    -2.1900     2.0900
     6.7200     0.8800     2.8000
     4.4700    -2.6200     5.7600
     6.7300    -2.0100     4.1800
     6.1800     2.8100     5.8200
     6.7200    -0.9300    -4.0400
     8.0900     0.2000     2.2500
     6.8100     0.1700    -4.1500
     4.0800     1.3000     5.3300
     6.2700     0.9300    -2.7800

第三类的数据有：
ans =

    -6.6800     3.1600     2.7100
    -8.6400     3.0600     3.5000
    -6.2500    -0.2600     0.5600
    -6.9400    -1.2200     1.1300
    -5.1900     4.2400     4.0400
    -6.3800    -1.7400     1.4300
```

```
>> Res(1,4)
采用K-means算法，使用欧式距离度量：
分类中心为：
center =

    -0.1000         0     0.1000
     5.4936    -0.1418     1.5782
    -6.3173     0.4061    -0.7236

第一类的数据有：
ans =

Empty matrix: 0-by-3

第二类的数据有：
ans =

     4.3600    -2.1900     2.0900
     6.7200     0.8800     2.8000
     4.4700    -2.6200     5.7600
     6.7300    -2.0100     4.1800
     6.1800     2.8100     5.8200
     6.7200    -0.9300    -4.0400
     8.0900     0.2000     2.2500
     6.8100     0.1700    -4.1500
     4.0800     1.3000     5.3300
     6.2700     0.9300    -2.7800

第三类的数据有：
ans =

    -7.8200    -4.5800    -3.9700
    -6.6800     3.1600     2.7100
    -8.6400     3.0600     3.5000
    -6.8700     0.0570    -5.4500
    -7.7100     2.3400    -6.3300
    -6.9100    -0.4900    -5.6800
    -6.2500    -0.2600     0.5600
    -6.9400    -1.2200     1.1300
    -5.1900     4.2400     4.0400
    -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, 使用欧式距离度量:

迭代次数为:15

分类中心为:

center =

4.5239	-0.0530	1.6204
-5.4757	0.2384	-0.5074

第一类的数据有:

ans =

4.3600	-2.1900	2.0900
6.7200	0.8800	2.8000
4.4700	-2.6200	5.7600
6.7300	-2.0100	4.1800
6.1800	2.8100	5.8200
6.7200	-0.9300	-4.0400
8.0900	0.2000	2.2500
6.8100	0.1700	-4.1500
4.0800	1.3000	5.3300
6.2700	0.9300	-2.7800

第二类的数据有:

ans =

-7.8200	-4.5800	-3.9700
-6.6800	3.1600	2.7100
-8.6400	3.0600	3.5000
-6.8700	0.0570	-5.4500
-7.7100	2.3400	-6.3300
-6.9100	-0.4900	-5.6800
-6.2500	-0.2600	0.5600
-6.9400	-1.2200	1.1300
-5.1900	4.2400	4.0400
-6.3800	-1.7400	1.4300

```
>> Res(2,2)
```

采用FCM算法, 使用欧式距离度量:

迭代次数为:18

分类中心为:

center =

-5.4755	0.2384	-0.5073
4.5237	-0.0531	1.6203

第一类的数据有:

ans =

-7.8200	-4.5800	-3.9700
-6.6800	3.1600	2.7100
-8.6400	3.0600	3.5000
-6.8700	0.0570	-5.4500
-7.7100	2.3400	-6.3300
-6.9100	-0.4900	-5.6800
-6.2500	-0.2600	0.5600
-6.9400	-1.2200	1.1300
-5.1900	4.2400	4.0400
-6.3800	-1.7400	1.4300

第二类的数据有:

ans =

4.3600	-2.1900	2.0900
6.7200	0.8800	2.8000
4.4700	-2.6200	5.7600
6.7300	-2.0100	4.1800
6.1800	2.8100	5.8200
6.7200	-0.9300	-4.0400
8.0900	0.2000	2.2500
6.8100	0.1700	-4.1500
4.0800	1.3000	5.3300
6.2700	0.9300	-2.7800

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:

<pre>>> Res(2,3) 采用FCM算法, 使用欧式距离度量: 迭代次数为:58 分类中心为: center = -4.3675 0.2863 -0.4392 5.1762 -0.1179 2.0186 -4.3838 0.2849 -0.4315 第一类的数据有: ans = Empty matrix: 0-by-3 第二类的数据有: ans = 4.3600 -2.1900 2.0900 6.7200 0.8800 2.8000 4.4700 -2.6200 5.7600 6.7300 -2.0100 4.1800 6.1800 2.8100 5.8200 6.7200 -0.9300 -4.0400 8.0900 0.2000 2.2500 6.8100 0.1700 -4.1500 4.0800 1.3000 5.3300 6.2700 0.9300 -2.7800 第三类的数据有: ans = -7.8200 -4.5800 -3.9700 -6.6800 3.1600 2.7100 -8.6400 3.0600 3.5000 -6.8700 0.0570 -5.4500 -7.7100 2.3400 -6.3300 -6.9100 -0.4900 -5.6800 -6.2500 -0.2600 0.5600 -6.9400 -1.2200 1.1300 -5.1900 4.2400 4.0400 -6.3800 -1.7400 1.4300</pre>	<pre>>> Res(2,4) 采用FCM算法, 使用欧式距离度量: 迭代次数为:41 分类中心为: center = -4.3836 0.2848 -0.4317 5.1762 -0.1179 2.0186 -4.3676 0.2863 -0.4390 第一类的数据有: ans = -7.8200 -4.5800 -3.9700 -6.6800 3.1600 2.7100 -8.6400 3.0600 3.5000 -6.8700 0.0570 -5.4500 -7.7100 2.3400 -6.3300 -6.9100 -0.4900 -5.6800 -6.2500 -0.2600 0.5600 -6.9400 -1.2200 1.1300 -5.1900 4.2400 4.0400 -6.3800 -1.7400 1.4300 第二类的数据有: ans = 4.3600 -2.1900 2.0900 6.7200 0.8800 2.8000 4.4700 -2.6200 5.7600 6.7300 -2.0100 4.1800 6.1800 2.8100 5.8200 6.7200 -0.9300 -4.0400 8.0900 0.2000 2.2500 6.8100 0.1700 -4.1500 4.0800 1.3000 5.3300 6.2700 0.9300 -2.7800 第三类的数据有: ans = Empty matrix: 0-by-3</pre>
--	--

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

Beta 取 0.001 时，kmeans 算法：

```
>> Res(1, 1, 0.001)
采用K-means算法，使用给定度量函数，beta=0.001000:
分类中心为:
center =

    5.5845   -0.0418    1.6600
   -6.3991    0.5061   -0.8236

第一类的数据有:
ans =

    4.3600   -2.1900    2.0900
    6.7200    0.8800    2.8000
    4.4700   -2.6200    5.7600
    6.7300   -2.0100    4.1800
    6.1800    2.8100    5.8200
    6.7200   -0.9300   -4.0400
    8.0900    0.2000    2.2500
    6.8100    0.1700   -4.1500
    4.0800    1.3000    5.3300
    6.2700    0.9300   -2.7800

第二类的数据有:
ans =

   -7.8200   -4.5800   -3.9700
   -6.6800    3.1600    2.7100
   -8.6400    3.0600    3.5000
   -6.8700    0.0570   -5.4500
   -7.7100    2.3400   -6.3300
   -6.9100   -0.4900   -5.6800
   -6.2500   -0.2600    0.5600
   -6.9400   -1.2200    1.1300
   -5.1900    4.2400    4.0400
   -6.3800   -1.7400    1.4300

未分开的数据有:
ans =

[]
```

```
>> Res(1, 1)
采用K-means算法，使用欧式距离度量
分类中心为:
center =

    5.5845   -0.0418    1.6600
   -6.3991    0.5061   -0.8236

第一类的数据有:
ans =

    4.3600   -2.1900    2.0900
    6.7200    0.8800    2.8000
    4.4700   -2.6200    5.7600
    6.7300   -2.0100    4.1800
    6.1800    2.8100    5.8200
    6.7200   -0.9300   -4.0400
    8.0900    0.2000    2.2500
    6.8100    0.1700   -4.1500
    4.0800    1.3000    5.3300
    6.2700    0.9300   -2.7800

第二类的数据有:
ans =

   -7.8200   -4.5800   -3.9700
   -6.6800    3.1600    2.7100
   -8.6400    3.0600    3.5000
   -6.8700    0.0570   -5.4500
   -7.7100    2.3400   -6.3300
   -6.9100   -0.4900   -5.6800
   -6.2500   -0.2600    0.5600
   -6.9400   -1.2200    1.1300
   -5.1900    4.2400    4.0400
   -6.3800   -1.7400    1.4300

未分开的数据有:
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：

```
>> Res(1,1,1)
采用K-means算法，使用给定度量函数，beta=1.000000:
分类中心为:
```

```
center =

    5.2037   -0.0787    3.6537
   -5.1520    0.4040    1.2320
```

第一类的数据有:

```
ans =

    4.3600   -2.1900    2.0900
    6.7200    0.8800    2.8000
    4.4700   -2.6200    5.7600
    6.7300   -2.0100    4.1800
    6.1800    2.8100    5.8200
    8.0900    0.2000    2.2500
    4.0800    1.3000    5.3300
```

第二类的数据有:

```
ans =

   -6.2500   -0.2600    0.5600
   -6.9400   -1.2200    1.1300
   -5.1900    4.2400    4.0400
   -6.3800   -1.7400    1.4300
```

未分开的数据有:

```
ans =

   -7.8200   -4.5800   -3.9700
   -6.6800    3.1600    2.7100
   -8.6400    3.0600    3.5000
   -6.8700    0.0570   -5.4500
   -7.7100    2.3400   -6.3300
   -6.9100   -0.4900   -5.6800
    6.7200   -0.9300   -4.0400
    6.8100    0.1700   -4.1500
    6.2700    0.9300   -2.7800
```

```
>> Res(1,1,100)
```

采用K-means算法，使用给定度量函数，beta=100.000000:

分类中心为:

```
center =

    1    1    1
   -1    1   -1
```

第一类的数据有:

```
ans =

Empty matrix: 0-by-3
```

第二类的数据有:

```
ans =

Empty matrix: 0-by-3
```

未分开的数据有:

```
ans =

   -7.8200   -4.5800   -3.9700
   -6.6800    3.1600    2.7100
    4.3600   -2.1900    2.0900
    6.7200    0.8800    2.8000
   -8.6400    3.0600    3.5000
   -6.8700    0.0570   -5.4500
    4.4700   -2.6200    5.7600
    6.7300   -2.0100    4.1800
   -7.7100    2.3400   -6.3300
   -6.9100   -0.4900   -5.6800
    6.1800    2.8100    5.8200
    6.7200   -0.9300   -4.0400
   -6.2500   -0.2600    0.5600
   -6.9400   -1.2200    1.1300
    8.0900    0.2000    2.2500
    6.8100    0.1700   -4.1500
   -5.1900    4.2400    4.0400
   -6.3800   -1.7400    1.4300
    4.0800    1.3000    5.3300
    6.2700    0.9300   -2.7800
```

fx

在 FCM 算法中，我们查看 (c) 中的分类情况如下:


```
>> Res(2, 3, 0.001)
采用FCM算法，使用给定度量函数，beta=0.001
迭代次数为:18
分类中心为:
center =
```

```
-6.8795 -0.3199 -5.2090
 6.0529 -0.1745  2.2109
-6.4454  1.2404  2.1240
```

第一类的数据有:

```
ans =
```

```
-7.8200 -4.5800 -3.9700
-6.8700  0.0570 -5.4500
-7.7100  2.3400 -6.3300
-6.9100 -0.4900 -5.6800
```

第二类的数据有:

```
ans =
```

```
4.3600 -2.1900  2.0900
6.7200  0.8800  2.8000
4.4700 -2.6200  5.7600
6.7300 -2.0100  4.1800
6.1800  2.8100  5.8200
6.7200 -0.9300 -4.0400
8.0900  0.2000  2.2500
6.8100  0.1700 -4.1500
4.0800  1.3000  5.3300
6.2700  0.9300 -2.7800
```

第三类的数据有:

```
ans =
```

```
-6.6800  3.1600  2.7100
-8.6400  3.0600  3.5000
-6.2500 -0.2600  0.5600
-6.9400 -1.2200  1.1300
-5.1900  4.2400  4.0400
-6.3800 -1.7400  1.4300
```

```
>> Res(2, 3, 0.1)
```

采用FCM算法，使用给定度量函数，beta=0.10000

迭代次数为:5

分类中心为:

```
center =
```

```
-0.4480  0.1554  0.4600
-0.4479  0.1553  0.4600
-0.4481  0.1554  0.4600
```

第一类的数据有:

```
ans =
```

```
Empty matrix: 0-by-3
```

第二类的数据有:

```
ans =
```

```
4.3600 -2.1900  2.0900
6.7200  0.8800  2.8000
4.4700 -2.6200  5.7600
6.7300 -2.0100  4.1800
6.1800  2.8100  5.8200
6.7200 -0.9300 -4.0400
8.0900  0.2000  2.2500
6.8100  0.1700 -4.1500
4.0800  1.3000  5.3300
6.2700  0.9300 -2.7800
```

第三类的数据有:

```
ans =
```

```
-7.8200 -4.5800 -3.9700
-6.6800  3.1600  2.7100
-8.6400  3.0600  3.5000
-6.8700  0.0570 -5.4500
-7.7100  2.3400 -6.3300
-6.9100 -0.4900 -5.6800
-6.2500 -0.2600  0.5600
-6.9400 -1.2200  1.1300
-5.1900  4.2400  4.0400
-6.3800 -1.7400  1.4300
```

fx

令人奇怪的是 β 取 0.1 时结果与欧式度量一样。而 β 取 0.001 时则反而有些分错，而当 β 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)
```

采用FCM算法，使用给定度量函数，beta=1.0

迭代次数为:3

分类中心为:

```
center =
```

-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:

```
ans =
```

```
Empty matrix: 0-by-3
```

第二类的数据有:

```
ans =
```

```
Empty matrix: 0-by-3
```

第三类的数据有:

```
ans =
```

```
Empty matrix: 0-by-3
```

未分开的数据有:

```
ans =
```

-7.8200	-4.5800	-3.9700
-6.6800	3.1600	2.7100
4.3600	-2.1900	2.0900
6.7200	0.8800	2.8000
-8.6400	3.0600	3.5000
-6.8700	0.0570	-5.4500
4.4700	-2.6200	5.7600
6.7300	-2.0100	4.1800
-7.7100	2.3400	-6.3300
-6.9100	-0.4900	-5.6800
6.1800	2.8100	5.8200
6.7200	-0.9300	-4.0400
-6.2500	-0.2600	0.5600
-6.9400	-1.2200	1.1300

fx

```
>> Res(2, 3, 100)
```

采用FCM算法，使用给定度量函数，beta=100.

迭代次数为:3

分类中心为:

```
center =
```

-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:

```
ans =
```

```
Empty matrix: 0-by-3
```

第二类的数据有:

```
ans =
```

```
Empty matrix: 0-by-3
```

第三类的数据有:

```
ans =
```

```
Empty matrix: 0-by-3
```

未分开的数据有:

```
ans =
```

-7.8200	-4.5800	-3.9700
-6.6800	3.1600	2.7100
4.3600	-2.1900	2.0900
6.7200	0.8800	2.8000
-8.6400	3.0600	3.5000
-6.8700	0.0570	-5.4500
4.4700	-2.6200	5.7600
6.7300	-2.0100	4.1800
-7.7100	2.3400	-6.3300
-6.9100	-0.4900	-5.6800
6.1800	2.8100	5.8200
6.7200	-0.9300	-4.0400
-6.2500	-0.2600	0.5600
-6.9400	-1.2200	1.1300

fx

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 β 取 0.1 时,实验的结果和欧式距离相同。