

作业

计算机科学与技术系 52 班杨定澄 学号：2015011274

E-mail:892431401@qq.com

1 生成数据

rand.py 用于产生数据，运行得到的 A1,B1,C1,A2,B2,C2 的数据存在对应名字的.txt 里。

现已生成好。

2 k-means

kmeans.py 实现了用 k-means 算法对 D2 进行聚类的结果，程序大约需花 4 秒左右。

聚类的结果如下：

```
第一簇的中心点  
[ 1.65198626  1.62790533  1.63353907]  
第二簇的中心点  
[ 6.2109197   7.39503424  6.42141664]  
第三簇的中心点  
[ 7.57100239  8.17216752 11.08752164]
```

真实的情况是每组数据的中心点大约在 $[1, 1, 1]$, $[6, 6, 6]$, $[7, 8, 9]$ 附近，可以看出结果比较稳合。

3 MLE 估计与分类正确率

MLE.py 用来估计参数与分类，程序要运行约 6 秒。

参数估计与分类结果如下：

```

A, B, C类的先验概率=
[0.3071756791043033, 0.19444619765330937, 0.4983781232423868]
A类的mean=
[ 0.9687576  1.03817006  1.03122787]
B类的mean=
[ 2.98369849  2.98686134  3.05688423]
C类的mean=
[ 7.05228389  7.91440706  8.94998992]
A类的协方差矩阵=
[[ 1.04249526  0.05122529  0.01894823]
 [ 0.05122529  0.97669313 -0.007496 ]
 [ 0.01894823 -0.007496   1.02095665]]
B类的协方差矩阵=
[[ 2.03787334 -0.0175603  0.13028412]
 [-0.0175603  3.16461339  0.28948231]
 [ 0.13028412  0.28948231  4.13014483]]
C类的协方差矩阵=
[[ 5.87567894  0.12705768  0.1865166 ]
 [ 0.12705768  6.13800228  0.10583541]
 [ 0.1865166  0.10583541  8.88380628]]
A类分类正确概率为0.970000
B类分类正确概率为0.930000
C类分类正确概率为0.960000
总的分类正确概率为0.953333

```

可以看到最后能有 95% 的正确率，是相当不错的分类方法。

初始的 μ 和协方差矩阵使用的是真实值， $P(\omega_i)$ 使用的是 $\frac{1}{3}$ ，而更改成任意随机值后都会收敛至较优的解。（也可以通过调整迭代次数要结果更优，我这里是强制迭代 30 次）