

程序运行说明及聚类文件

1. 利用 `datatrans.m` 导入数据，该函数的说明，在第三次作业中已经说过。功能是将数据从.txt 文件变到数组中。最后全部数据存在 `dataset` 中。（得到的 `label`, `womenset`, `menset` 等数据在该实验中没有用到。）

2. 运行 `pca(dataset')` 这是一个较为统一的函数，注释已经在源文件中写的比较清楚。

2.1 如果需要画图，将 25——31 行的注释去掉，并选择自己需要留下的维数和分的类别即可。画图只能花 3 个特征一下维数的图。

2.2 如果要考察不同的距离，在第 22 行 `kmeans` 的函数后添加,如（“Distance”, “cityblock”）即可（具体参见 `Kmeans` 的帮助文档）

3. 如果需要知道划分的聚类结果与最后真实结果的差距。将 25 行、32 行的注释去掉，并在中间设置断点，运行到断点处，运行

```
for i=1:class
a=sum(cidx(486:size(cidx,1))==i)
b=sum(cidx(1:485)==i)
end
```

即可得到实验报告中每个表的值。其中 `remaindem` 和 `class` 为需要保留的维数和分类的类数。

对于 `raw data`, 过程与之类似。

4. 对于五个特征的情况，分三类和分两类的聚类结果分别保存在了 `cluster3.mat` 和 `cluster2.mat` 中。行代表样本（954 个样本），列代表保留的特征数目，1,2,3 分别代表属于 1,2,3 类。