

PCA 算法:

自己的一些设定:

1) 直接调用了 numpy 中许多矩阵的操作。

2) 数据标准化是按照缩放到[0,1]之内做的。

缩放值=(当前值-最小值)/(最大值-最小值)。

思路:

1) 原始数据按每行一条数据组成 n 行 m 列矩阵

2) 将属性的每一列进行 0 均值化, 即减去这一列的均值

3) 利用 numpy 中的 cov 函数求出协方差矩阵

4) 对于协方差矩阵求出特征值及其对应的特征向量

5) 利用传入的特征值累计贡献率计算取前多少个特征值

6) 取出特征值对应的特征向量形成新的矩阵

运行结果图:

特征值贡献率 0.9

```
开始 插入 [[ 0.65097987 -0.41221174 -0.07124726 0.13250736 0.13877752 0.21443167
-0.03448881 0.10244441]]
A [[ 0.71343063 -0.20437147 -0.24815545 -0.32063043 -0.10935007 0.02291729
-0.00540241 -0.12008188]]
X2 A - 2 [[ 0.256576 -0.64530584 -0.24819576 -0.25927833 -0.2785368 0.56785091
-0.19785604 0.24451152]]
numpy 中 [[ 0.37364252 -0.44353228 -0.11828998 -0.22211758 -0.21996265 0.53970093
-0.16861824 0.06824512]]
是按照缩放 [[ 0.67501204 -0.14345928 -0.12705076 0.06252094 0.03968583 0.20268131
-0.02108487 -0.25079771]]
值-最小值) [[ 0.51748468 -0.22918728 0.04051356 -0.14443885 0.08609243 -0.19676843
0.031768 0.04669577]]
每行一条数 [[ 0.62606956 -0.04628052 -0.14037643 -0.22200846 0.06207065 -0.12293465
0.07368096 0.00224031]]
列进行 0 [[ 0.57900853 -0.13907482 0.14146187 0.13695703 0.02311716 -0.13778017
0.09172621 -0.05212501]]
中的 cov 函 [[ 0.62437568 -0.34886831 0.1599144 0.03292312 -0.03914209 0.06694715
-0.20834084 0.15296275]]
矩阵求出特 [[ 0.76634979 -0.22301853 -0.07287329 0.04945748 0.21935984 -0.04424873
-0.00806222 0.20828706]]
特征值累计 [[ 0.56301119 -0.4765656 -0.07389449 0.00103894 -0.18273702 0.0235175
0.03410113 0.0987567 ]]
对应的特征 [[ 0.62438891 -0.28876937 0.14386427 0.10494247 0.00157391 0.09887445
-0.09553005 -0.06507237]]
[[ 0.47251421 -0.36078103 -0.02121867 -0.03403024 -0.25801679 -0.02505048
-0.10237244 0.16960613]]
[[ 0.48239336 -0.4159888 -0.25043078 0.32438214 -0.23437364 -0.19827251
-0.10627871 -0.00577922]]
[[ 0.67918228 0.06575189 0.16871982 0.13808317 -0.01897717 -0.07136612
-0.00175037 -0.12263152]]
[[ 0.77453037 -0.18191366 0.21294202 -0.02350404 -0.02597415 0.18823704
-0.2098698 -0.09291572]]
[[ 0.53765639 -0.4624047 0.137019 -0.03480618 -0.0104513 0.13442427
-0.20731956 0.12782535]]
; k 个聚类点 [[ 0.7395095 -0.47190115 -0.20936033 0.14972687 0.29176872 -0.08383375
0.09019078 -0.00981782]]
将数据分为 [[ 0.58178114 -0.34836575 -0.08359029 0.07522574 -0.05031563 -0.02529549
0.04801419 0.04728405]]
的是欧式距 [[ 0.62631289 -0.54685701 0.03049476 0.17013732 -0.0737382 0.03170277
0.18633392 -0.03713893]]
值的方法计 [[ 0.5729911 -0.42551609 0.09453672 -0.13739317 -0.15055918 0.11225401
0.14892236 -0.02466179]]
步, 直到中 [[ 0.701764 -0.51350498 -0.29390996 -0.08137147 0.02513817 0.0906167
-0.1134516 0.11174192]]
迭代次数 PS F:\作业2020\人工智能作业\PB17111609_lab2\unsupervise\src>
```

特征值贡献率 0.5:

