

机器学习

第3次实验： 对航空公司客户价值进行聚类分析

实验目的

01

会用Python创建Kmeans聚类分析模型

02

使用KMeans模型对航空公司客户价值进行聚类分析

03

会对聚类结果进行分析评价

实验内容

使用sklearn.cluester的KMeans类对航空公司客户数据进行聚类分析，把乘客分到不同的类别中

- ✓ **数据集：air_data.csv**
- ✓ **数据集大小：62052条不重复数据**
- ✓ **原数据有40个属性，为了大家训练模型方便，本实验使用预处理后的标准化数据，该数据有5个属性**

实验内容

✓ 数据说明

- ZL: 入会至当前时长, 反应可能的活跃时间
- ZR: 最近消费时间间隔, 反应最近一段时间活跃程度
- ZF: 消费频次, 反应客户忠诚度
- ZM: 消费里程总额, 反应客户对乘机的依赖程度
- ZC: 舱位等级对应折扣系数, 一般舱位等级越高, 折扣系数越大

实验内容

1

载入训练数据、显示读入数据的前5行

2

训练KMeans聚类模型，把数据聚成5类

```
from sklearn.cluster import KMeans
k = 5
model = ...
KMeans(algorithm='auto', copy_x=True, init='k-means++',
        max_iter=300, n_clusters=5, n_init=10, n_jobs=None,
        precompute_distances='auto', random_state=None,
        tol=0.0001, verbose=0)
```

实验内容

3

检查每个聚类类别样本数、每个聚类类别中心点，统计聚类个数及中心点

4

画出5个聚类中心点在每个维度上的散点图，并按统一类别把聚类中心用线连接起来

5

分析聚类结果

执行结果图例

```
1. 载入训练数据, 输出数据前5条.....

>>>测试数据前5条:

      ZL      ZR      ZF      ZM      ZC
0  1.445701 -0.944998 14.034510 26.762112 1.294939
1  1.304181 -0.911945  9.073559 13.127371 2.866311
2  1.339561 -0.889910  8.719206 12.653972 2.879075
3  0.667342 -0.416159  0.781685 12.541109 1.993551
4  0.384302 -0.922963  9.924008 13.899268 1.343695

2. 训练KMeans聚类模型.....
3. 检查聚类结果.....

>>>查看每个聚类类别的样本数量:
1      23900
0      15525
4      11890
2       5389
3       5347
dtype: int64

>>>查看客户样本类别的聚类标准中心值:

      0      1      2      3      4
0  1.168936 -0.380394 -0.083019 -0.090589 -0.165740
1 -0.701537 -0.419798 -0.154753 -0.151863 -0.299190
2 -0.061701  0.001555 -0.268473 -0.286520  1.906873
3  0.485956 -0.799832  2.482688  2.423800  0.322613
4 -0.307776  1.698690 -0.574984 -0.536152 -0.200488
```

>>>查看聚类个数及中心点统计:

	聚类个数	ZL	ZR	ZF	ZM	ZC
0	15525	1.168936	-0.380394	-0.083019	-0.090589	-0.165740
1	23900	-0.701537	-0.419798	-0.154753	-0.151863	-0.299190
2	5389	-0.061701	0.001555	-0.268473	-0.286520	1.906873
3	5347	0.485956	-0.799832	2.482688	2.423800	0.322613
4	11890	-0.307776	1.698690	-0.574984	-0.536152	-0.200488

