Python
机器学习实战

聚类算法

聚类算法

聚类和分类的区别

聚类是无监督的学习

分类是有监督的学习

聚类的场景

使用聚类不需要提前被告知要划分的组是什么样的，

在我们不知道找什么时就自动完成分组

聚类算法

无监督学习算法

聚类算法一览

层次聚类法

Kmeans

密度聚类法

谱聚类

GMM

LDA

…… ……

聚类算法

基于距离

- 层次聚类法
- K-means聚类法
- K-center聚类法

基于密度

- 密度距离法（DBCSAN）

数学中距离满足三个要求

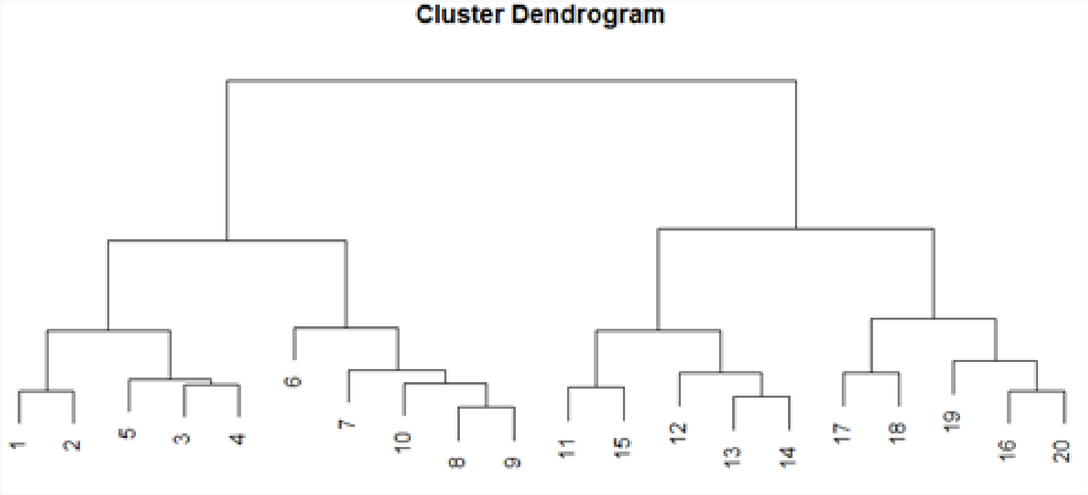
- 必须为正数
- 必须对称
- 满足三角不等式

层次聚类法

1、找到两个最近的两个点（类）；

2、把他们合成一个点（类）；（这个新的点不是数据集中本来存在的点，把原来那两个点删掉，用这个点代替原来两个点）

3、重复1、2。最终得到一棵树。

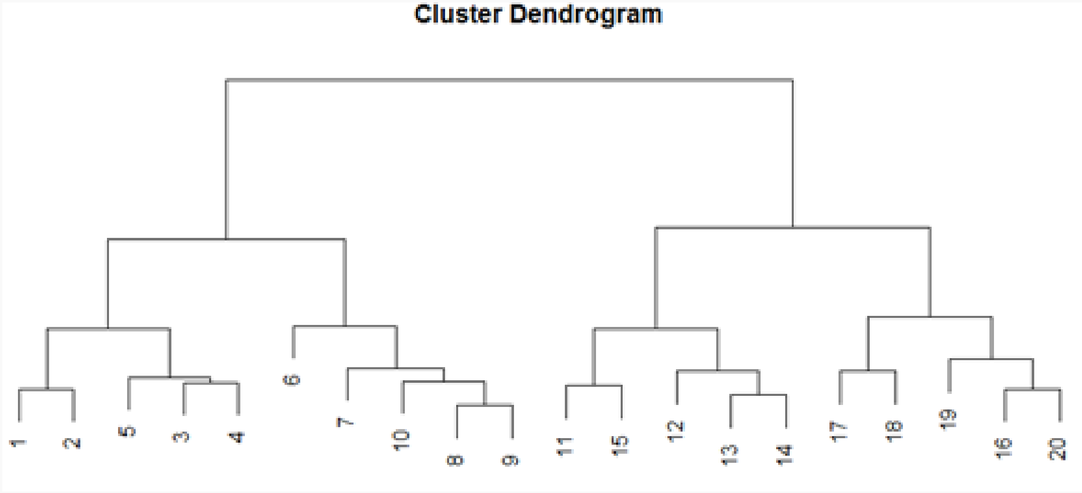


聚类算法

类间距离的计算

- 最短距离法
- 最长距离法
- 中间距离法
- 类平均法

不同的距离方法，得到不同的结果



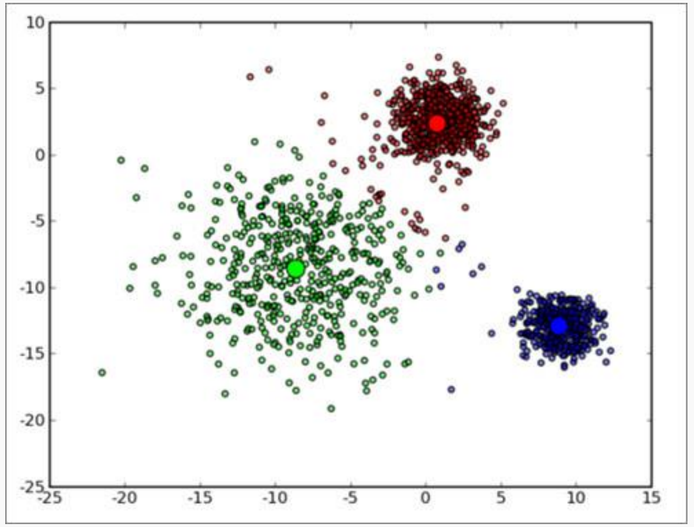
K-means聚类法

1、随机选择k个点作为初始质心

2、把每个点按照距离分配给最近的质心，形成k个簇

3、重新计算每个簇的质心

4、重复2、3步，直到质心不再变化



数据变换

中心化变换：均值为0，方差不变

$$x_{ij}^{*} = x_{ij} - \overline{x}_{j},$$

标准化变换：均值为0，方差为1

$$x_{ij}^{*} = \frac{x_{ij} - \overline{x}_{j}}{s_{j}}$$

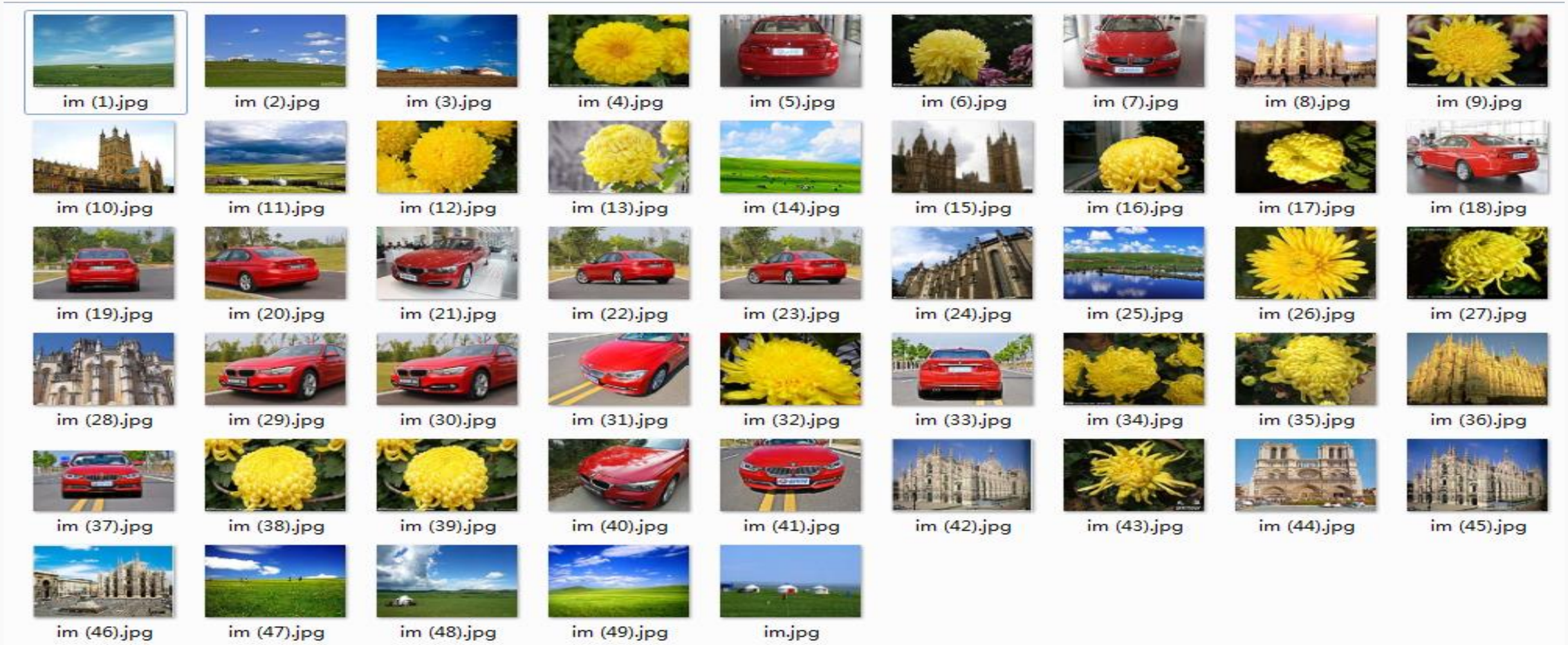
极差正规化变换：极差为1，数值在[0,1]之间

$$x_{ij}^{*} = \frac{x_{ij} - \min\limits_{1 \le k \le n} x_{kj}}{R_{j}}$$

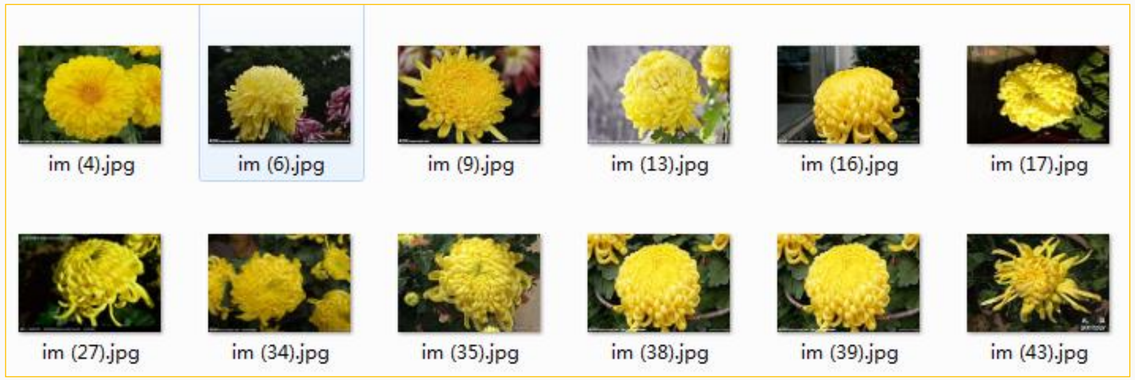
实战

图像按照色彩聚类

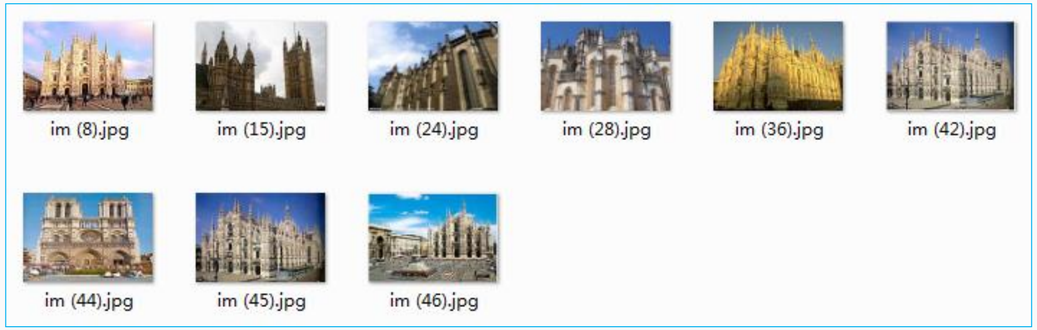
聚类算法

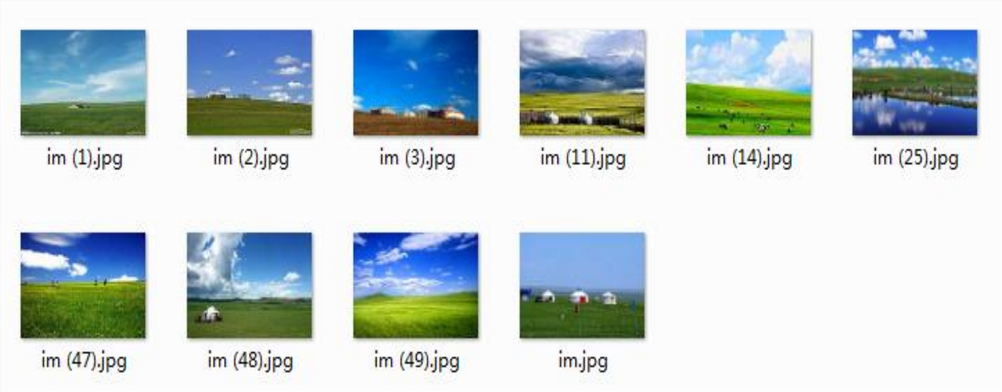

im (1).jpg
im (2).jpg
im (3).jpg
im (4).jpg
im (5).jpg
im (6).jpg
im (7).jpg
im (8).jpg
im (9).jpg
im (10).jpg
im (11).jpg
im (12).jpg
im (13).jpg
im (14).jpg
im (15).jpg
im (16).jpg
im (17).jpg
im (18).jpg
im (19).jpg
im (20).jpg
im (21).jpg
im (22).jpg
im (23).jpg
im (24).jpg
im (25).jpg
im (26).jpg
im (27).jpg
im (28).jpg
im (29).jpg
im (30).jpg
im (31).jpg
im (32).jpg
im (33).jpg
im (34).jpg
im (35).jpg
im (36).jpg
im (37).jpg
im (38).jpg
im (39).jpg
im (40).jpg
im (41).jpg
im (42).jpg
im (43).jpg
im (44).jpg
im (45).jpg
im (46).jpg
im (47).jpg
im (48).jpg
im (49).jpg
im.jpg

聚类算法


im (4).jpg
im (6).jpg
im (9).jpg
im (13).jpg
im (16).jpg
im (17).jpg
im (27).jpg
im (34).jpg
im (35).jpg
im (38).jpg
im (39).jpg
im (43).jpg


im (5).jpg
im (7).jpg
im (10).jpg
im (18).jpg
im (19).jpg
im (20).jpg
im (21).jpg
im (22).jpg
im (23).jpg
im (29).jpg
im (30).jpg
im (31).jpg
im (33).jpg
im (37).jpg
im (40).jpg
im (41).jpg


im (8).jpg
im (15).jpg
im (24).jpg
im (28).jpg
im (36).jpg
im (42).jpg
im (44).jpg
im (45).jpg
im (46).jpg


im (1).jpg
im (2).jpg
im (3).jpg
im (11).jpg
im (14).jpg
im (25).jpg
im (47).jpg
im (48).jpg
im (49).jpg
im.jpg


im (12).jpg
im (26).jpg
im (32).jpg