# 第11章习题参考答案

11.1 试阐述聚类和分类的区别。

回答：聚类和分类本质上都是将样本划分到不同的类别中。在使用数据上：聚类针对的是未标记的训练样本，而分类针对的是带有标签的样本数据，其本质是一个监督问题；在目标上：聚类是通过各种模型算法揭示未标记数据的内在性关联规律，而分类则是寻找一个模型预测类别标签。

11.2 给定道路路段的行驶速度，设定，试使用K-means算法将其聚类。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 1 | 2 | 3 | 4 | 5 |
| 速度 | 2 | 11 | 6 | 5 | 12 |

回答：（1）随机选择两个样本点作为初始类的中心，假设选择，。

（2）以和为两类别和的中心，分别计算到和的距离。

对，，，由于，因此将分类到。

对，，，由于，因此将分类到。

对，，，由于，因此将分类到。

（3）得到新的类，，计算新的类别的中心和：，。

（4）重复步骤（2）和步骤（3）：

对，，，由于，因此将分类到。

对，，，由于，因此将分类到。

对，，，由于，因此将分类到。

对，，，由于，因此将分类到。

对，，，由于，因此将分类到。

由于得到的新类，并未发生改变，聚类过程到此停止。因此，最终速度的聚类结果为：，。

11.3 简述聚合聚类算法原理。

回答：首先计算样本两两之间的欧式距离，然后合并类间距离最小的两个子类构成一个新类并计算新类与当前各类的距离，持续进行此步骤直至最终全部合并为一类，结束聚类。

11.4 查找车辆行驶轨迹数据，试着使用DBSCAN算法判别道路通行状态。

回答：略（具体可参照11.5中的案例应用实现）。