实验六　聚类算法

一．简单介绍 K-means、层次聚类，DBSCAN 和密度峰值聚类(DPC)算法的原理。

二．鸢尾花数据实验

1.对鸢尾花数据集进行探索性分析与预处理。

2.选取兰德系数和轮廓系数作为评价指标，对四种算法在该数据集上的性能进行分析。

三．算法参数影响探究

1.介绍三种算法中的几个主要参数（K-Means 中的 k 参数、DBSCAN 中的 eps 与 min_samples 参数、DPC 中的 t_0 参数--t_0 的含义为圆中样本个数占数据集总样本数的比例）。

2.以鸢尾花数据为例，选取合适的评价指标，探究 K-Means 算法中 k 参数对算法的性能影响，并尝试找出确定 k 参数的方法。

3.以模拟数据为例(如：高斯分布数据集，Spiral 数据集，Circle 数据集)，选取合适的评价指标，探究另外三个参数（eps 与 min_samples、t_0）对各自算法的性能影响。