

机器学习第八章实验内容

详细内容见第七周实验内的 jupyter notebook。

一、实验内容

1.1 使用 sklearn 的 DBSCAN 和 GaussianMixture 完成聚类

1. 使用 sklearn 的 DBSCAN 和 GaussianMixture 在两个数据集上完成聚类任务
2. 对聚类结果可视化
3. 对比外部指标 FMI 和 NMI
4. 选做：调整密度聚类的 eps 参数，绘制聚类结果

1.2 使用 sklearn 的 KMeans 完成聚类

1. 使用 sklearn 的 Kmeans 完成两个数据集的聚类任务
2. 计算外部指标 FMI 和 NMI
3. 对聚类结果可视化

1.3 使用 sklearn 的 AgglomerativeClustering 完成层次聚类

1. 使用 sklearn 的 AgglomerativeClustering 完成两个数据集的层次聚类
2. 计算外部指标 FMI 和 NMI
3. 对 6 种形式的聚类效果可视化

1.4 实现 K-means

1. 实现一个 K-means 聚类算法
2. 计算外部指标 FMI 和 NMI
3. 对聚类结果可视化
4. 完成第二个数据集上 myKmeans 与层次聚类(single)算法的对比

1.5 实现层次聚类，并与其他聚类算法进行对比（选做）

1. 实现三种形式的层次聚类（最小距离、最大距离、平均距离）
2. 至少选择两个数据集
3. 在这两个数据集上运行自己实现的层次聚类算法和 sklearn 的 KMeans, GMM, DBSCAN, 共 4 个聚类算法

4. 计算外部指标 FMI 和 NMI
5. 绘制 4 种聚类的聚类效果