**第二次上机实验任务：单细胞聚类**

1. 数据介绍

单细胞RNA-seq测序数据介绍

1. **Zeisel\_df**

预处理过后的单细胞数据集，已经进行了细胞类型筛选。行代表细胞，列代表基因。

1. **Sclabel**

Zeisel\_df中细胞对应的真实细胞类型标签。

1. 任务分析
2. 特征选择和降维

从所有基因中选择最具信息量的基因，例如通过差异表达分析或方差筛选。使用降维方法（如主成分分析（PCA）或t-SNE）将高维数据映射到较低维空间，以便于可视化和聚类。

1. 聚类算法应用

选择合适的聚类算法，如层次聚类、DBSCAN等，将细胞分配到不同的聚类簇中。

1. 聚类结果评估和可视化

通过内部指标（如轮廓系数）或外部指标（如调整兰德指数ARI）评估聚类结果的质量。使用二维或三维降维方法（如t-SNE、UMAP）对聚类结果进行可视化展示，以便于观察细胞群集的分布和相互作用关系。

1. 实验报告撰写

1、针对该数据分析任务，撰写实验报告，包含任务描述、数据描述、详细分析步骤实结合实验结果及可视化，但不限于此。

2、以自己的数据探索结果为主要内容，给出丰富的分析结果，并探讨以上结果对于后续数据挖掘的影响和启示。

1. 验收

最后一次上机验收实验代码和结果。

最终提交材料为实验报告+实验代码、实验结果。

执笔人：郭杏莉

2023-11-27