

第一次作业（2024.9.22）

链接数据包括 4 组数据，每组数据包括 xxxcounts.csv, xxxlogcounts.csv 和 xxxclusterxxx.csv。其中 xxxcounts.csv 为原始数据矩阵，每个元素均大于等于 0，行为基因，列为细胞。xxxlogcounts.csv 为对该数据的预处理，通常加 1 后取对数或者先对数据矩阵每列乘以权重使每列加和相同的情况下再加 1 后取对数。xxxcluster.csv 为细胞类型数据。

- (1) mclust, Leiden 是两种常用的聚类方法，请查阅文献，深入理解这两种方法。
- (2) 请基于我们学习的内容和 mclust、Leiden 方法，查阅相关文献，自行设计方法对数据中的细胞进行聚类分析，例如可以用 PCA+kmeans 对数据进行聚类。对聚类效果，请用轮廓系数、ARI 和 NMI 等指标来度量聚类结果，

作业提交：报告一份，至少包括：问题、方法、结果分析三部分；程序包一份，程序包与报告分开。

数据链接：

<https://pan.baidu.com/s/1KomRqDNDDoSzsIDSvAXtqg?pwd=8fkq>

提取码：8fkq