重画公开论文的 UMAP，形状不一样，要从哪一步开始比?

从 Scanpy 的公开 PBMC3K 参考对象取出相同的 562 个条码，分别检验整体旋转和从上游更改特征重新拟合所造成的图形差别。



运行教学示例

Python 3.10 或以上和 requirements.txt。默认用已拟合坐标；从头拟合还需 umap-learn，版本记录在 fit 中。

```
python run.py --input example --out results
```

在解压后的资料包根目录运行。example 是教学输入，results 保存本次输出，expected 保存可供核对的参考结果。

输入字段

reference.npz、cells.tsv、genes.txt 真实表达层、原 PCA 与 UMAP、条码顺序及参考标签。

fit_umap.py、fit/changed_umap.tsv 重做特征、PCA 和 UMAP 的脚本及已拟合坐标。

核对输出，再换自己的数据

stage_audit.tsv、edges.tsv 差异首次出现的环节、图上及 PCA 邻居比较、每条有向邻边。

three_embeddings.png 三种情况的同条码同标签对照。

用示例验算

1 只旋转时，PCA 和二维邻居的交并比中位数都为 1。

2 在 919 个特征上重拟合时，PCA 邻域中位交并比约 0.154，二维约 0.111。

3 标签一致率由约 0.887 降至 0.742。

换用真实数据

先建立共同条码顺序，确认输入层、特征、PCA 和图空间；记下邻居是从子集重建还是从原图诱导。

本例提供确切参考数据；变更示例同时涉及子集重拟合，不能把数值差异全归因于特征个数。

方法资料

Scanpy 公开 PBMC3K 教程

UMAP 作者对随机性与复现的说明