可视化实验四报告

201900161140 张文浩

实验完成时间：10.15

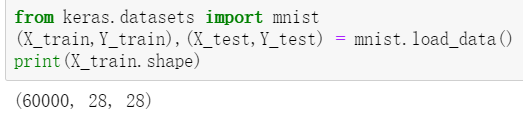
软件环境：vscode

### 实验要求

参考https://distill.pub/2016/misread-tsne/ 对mnist数据分析不同参数对t-sne结果的影响

### 实验步骤

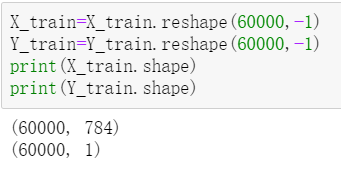
**1.加载mnist数据集**



可以观察出mnist的train数据集有60000条数据，每一个数据是一个大小为28\*28的图片。

**2.维度合并**

把二维的图片合并成一维



可以看到图片大小的两个维度变成了一个维度，784=28\*28

**3.对比不同参数的效果**

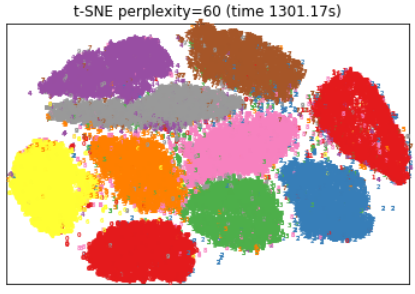
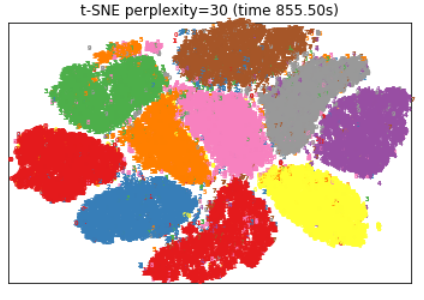
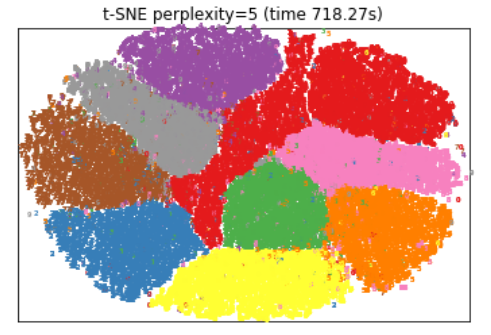
查阅sklearn的官方文档后，我认为可以修改参数呈现出不同效果的参数有两个，分别是perplexity复杂度和init初始化方法。

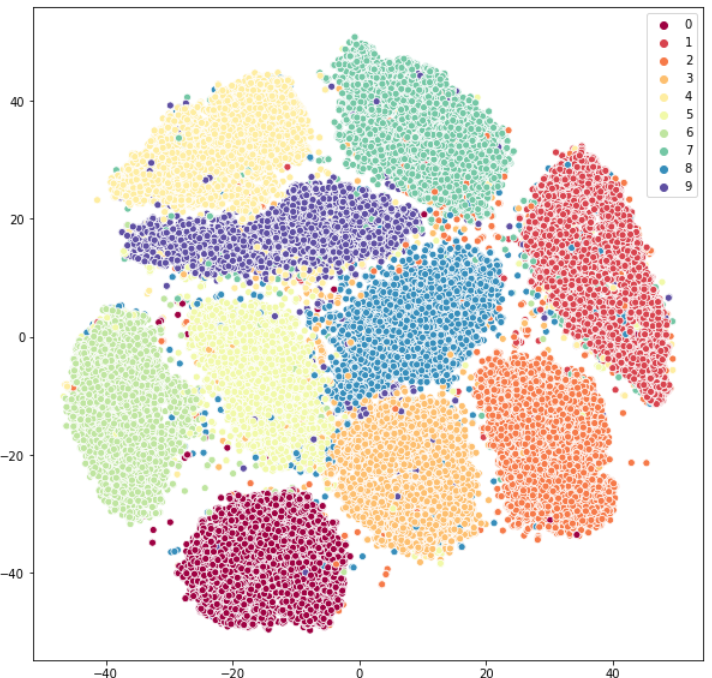
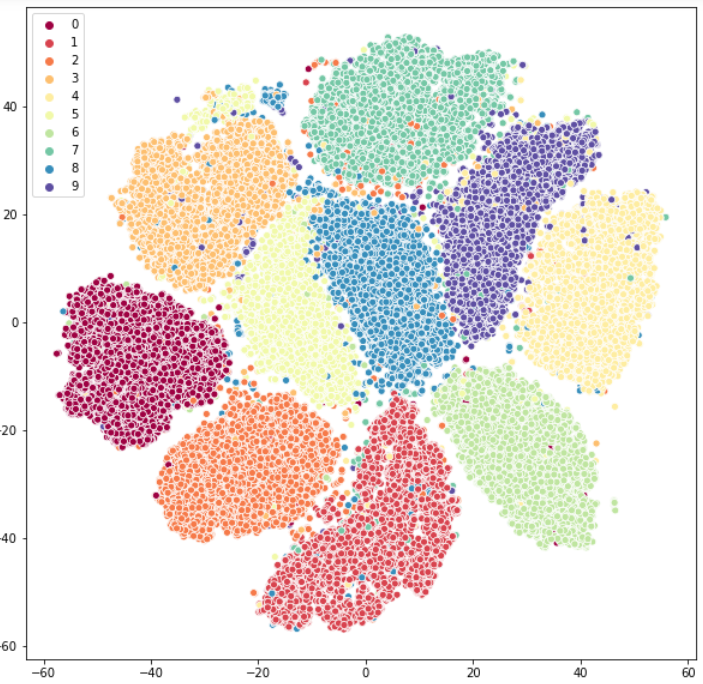
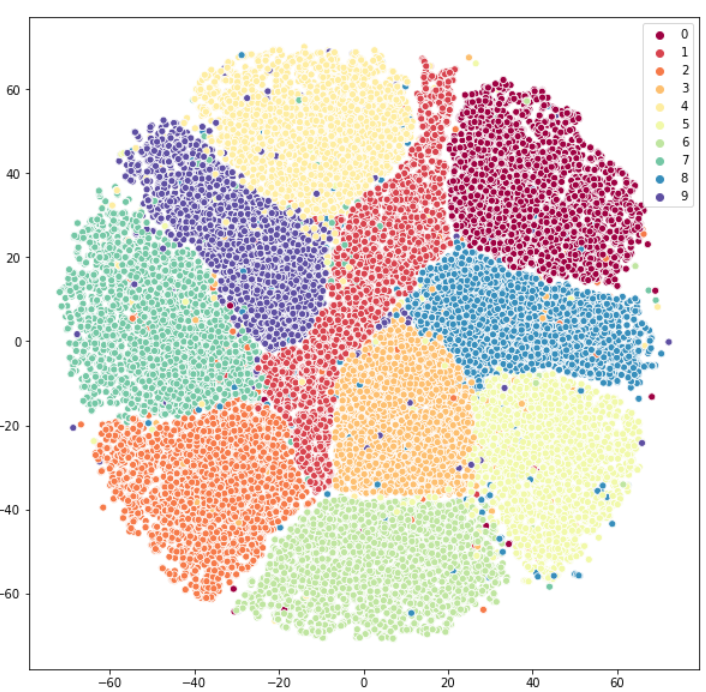
官方文档上对perplexity的描述是：perplexity与其他流形学习算法中使用的最近邻数有关。更大的数据集通常需要更大的perplexity。考虑选择一个介于5和50之间的值。不同的值可能导致显著不同的结果。

init初始化方式分为random初始化和pca初始化两种。

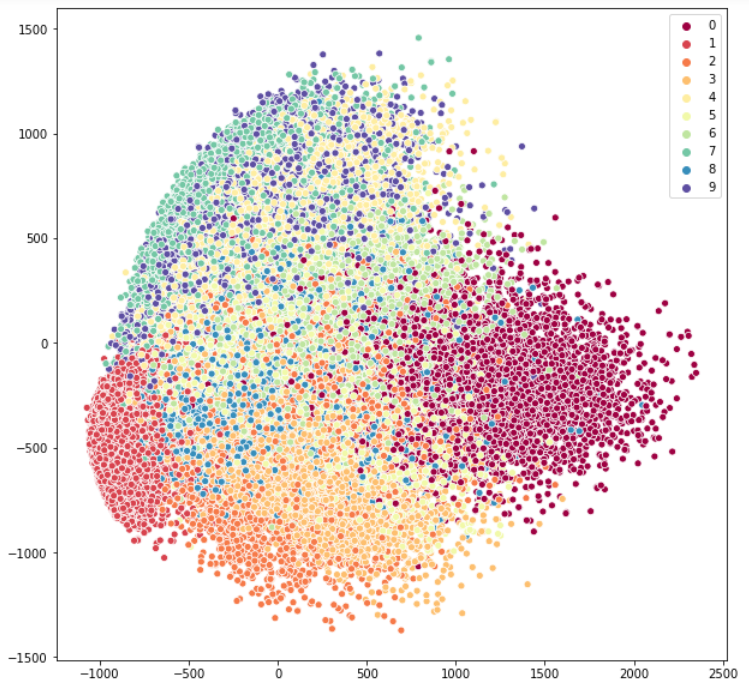
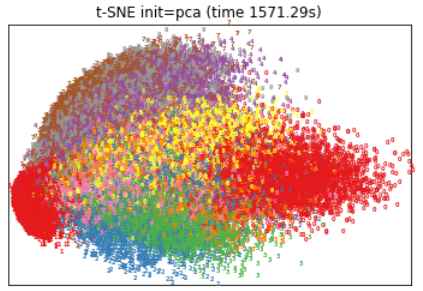
**random：**

**perplexity = 5**   **perplexity= 30 perplexity=60**

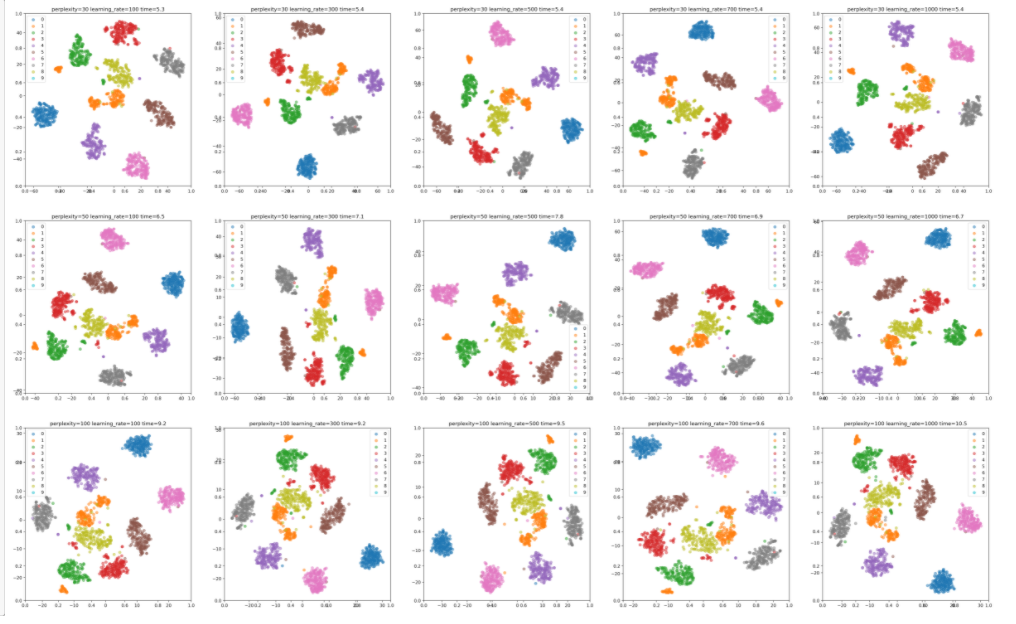




**pca：**

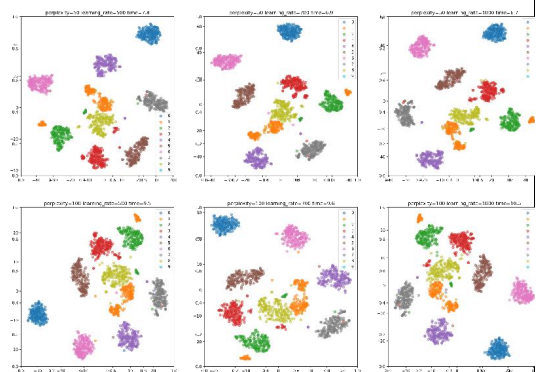
****

为比较尝试更多、更大范围的参数，从60000个训练数据中抽样，提高速度。

****

从上到下perplexity增大，从左到右学习率learning\_rate变大

将init参数换成pca，再次测试



### 实验总结

通过调整初始化方式和perplexity两个参数，观察了不同参数对可视化mnist数据集的的影响，发现random初始化效果比pca好很多。对于perplexity参数来说，perplexity类似于k近邻中的k一样的参数，根据实验结果，可视化的效果应该是参数perplexity越大，每个类区分的越明显，计算量更大，时间更长。修改learning\_rate效果变化不明显。