

# 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

```
from keras.datasets import mnist
(X_train, Y_train), (X_test, Y_test) = mnist.load_data()
print(X_train.shape)
```

(60000, 28, 28)

可以观察到 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

```
X_train=X_train.reshape(60000,-1)
Y_train=Y_train.reshape(60000,-1)
print(X_train.shape)
print(Y_train.shape)
```

(60000, 784)  
(60000, 1)

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

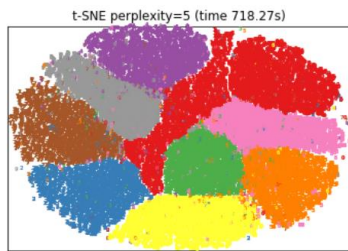
查阅 sklearn 的官方文档后，我认为可以修改参数呈现出不同效果的参数有两个，分别是 perplexity 复杂度和 init 初始化方法。

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。

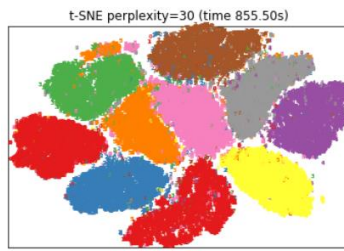
init 初始化方式分为 random 初始化和 pca 初始化两种。

**random:**

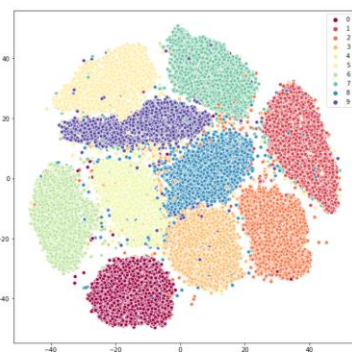
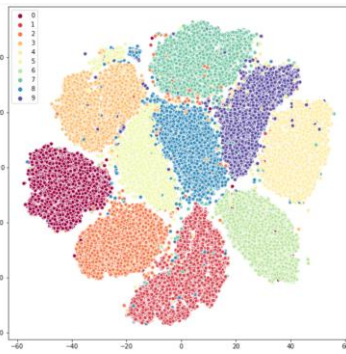
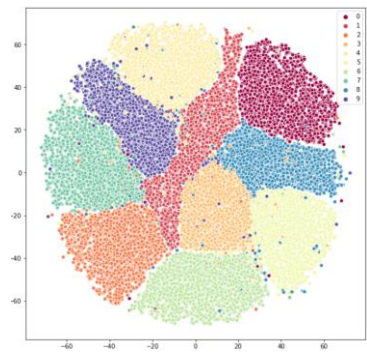
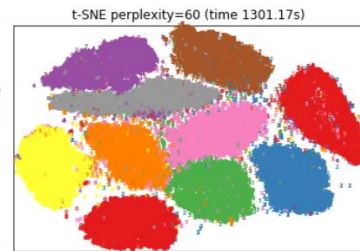
**perplexity = 5**



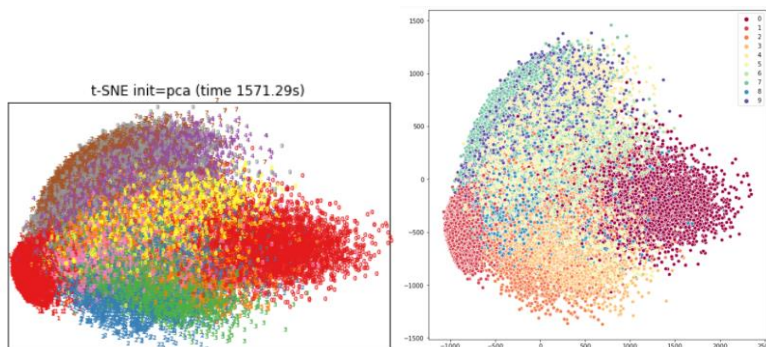
**perplexity= 30**



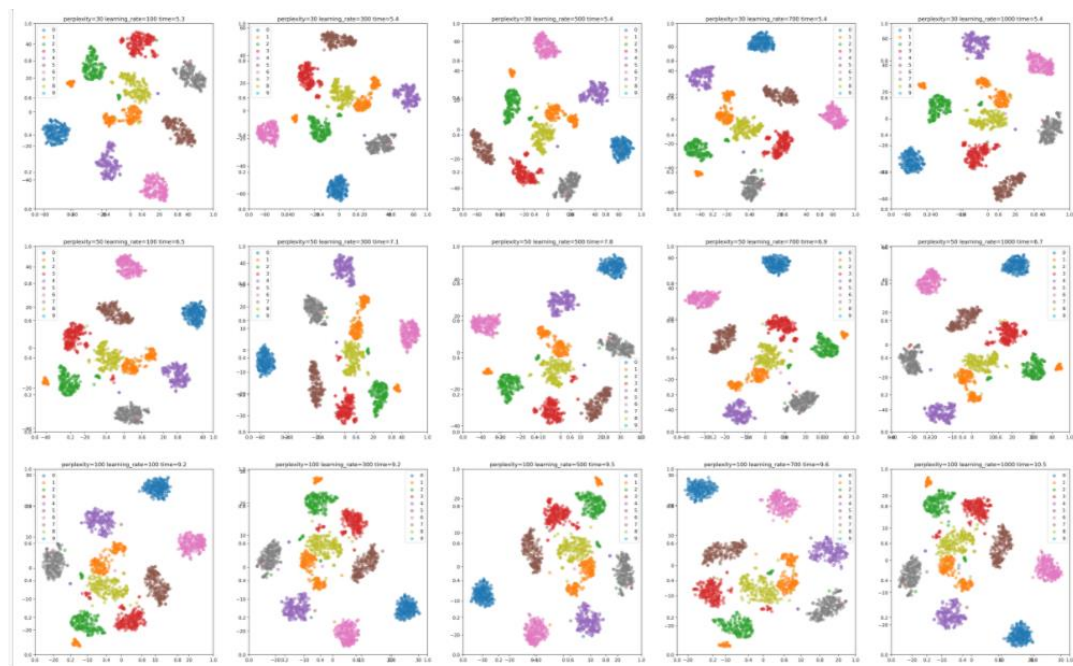
**perplexity=60**



**pca:**

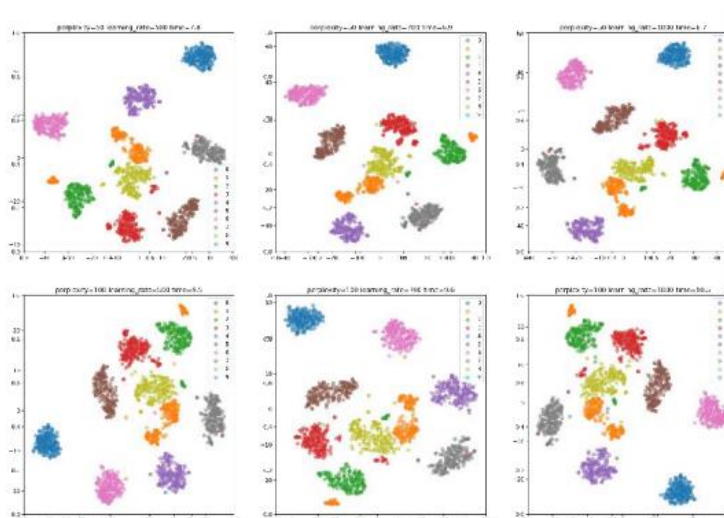


为比较尝试更多、更大范围的参数，从 60000 个训练数据中抽样，提高速度。



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。