**Ex7 Report**

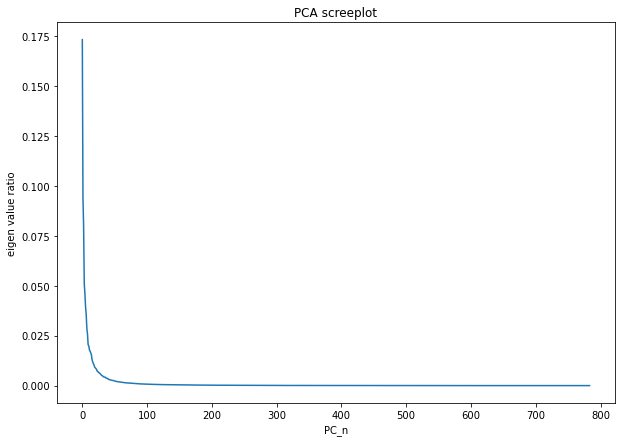
自硕21 崔晏菲 2021210976

**Experiment 13：**

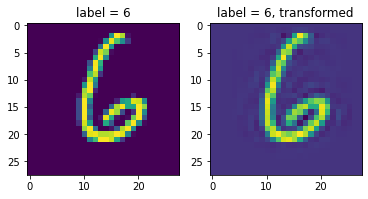
**1. 选择数字6**

**1.1 PCA降维**

碎石图如图所示，找到拐点并不是很合理，因为“拐点”处只占用了80%的特征值。为了占用95%以上的特征值，我选择n\_PCs=112



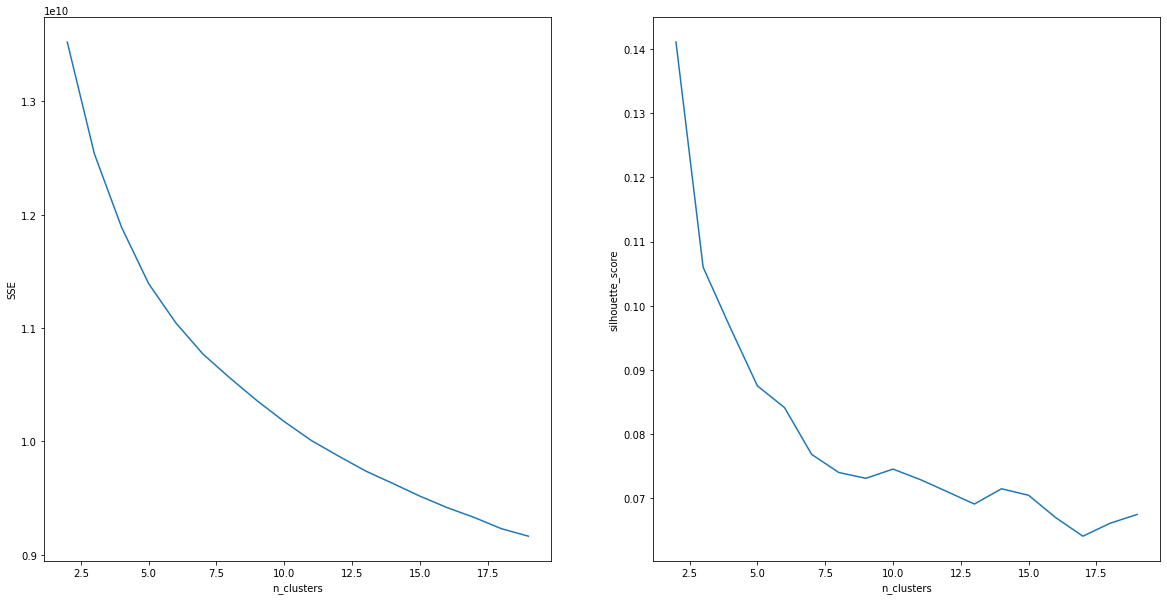
PCA降维前后的对比图如下所示：



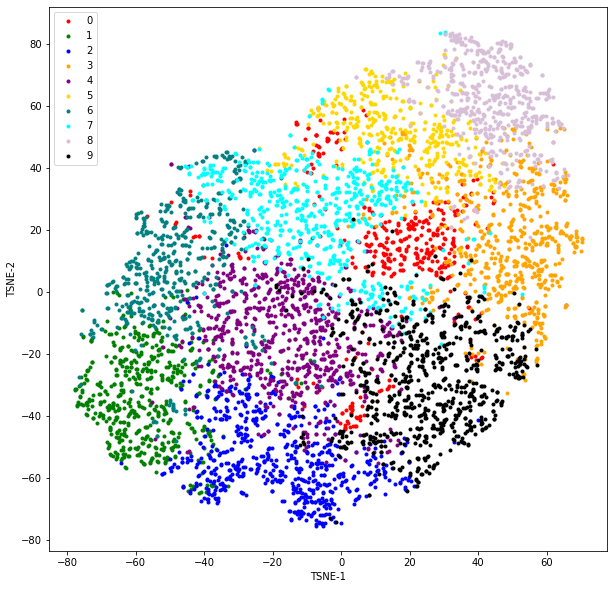
**1.2 聚类**

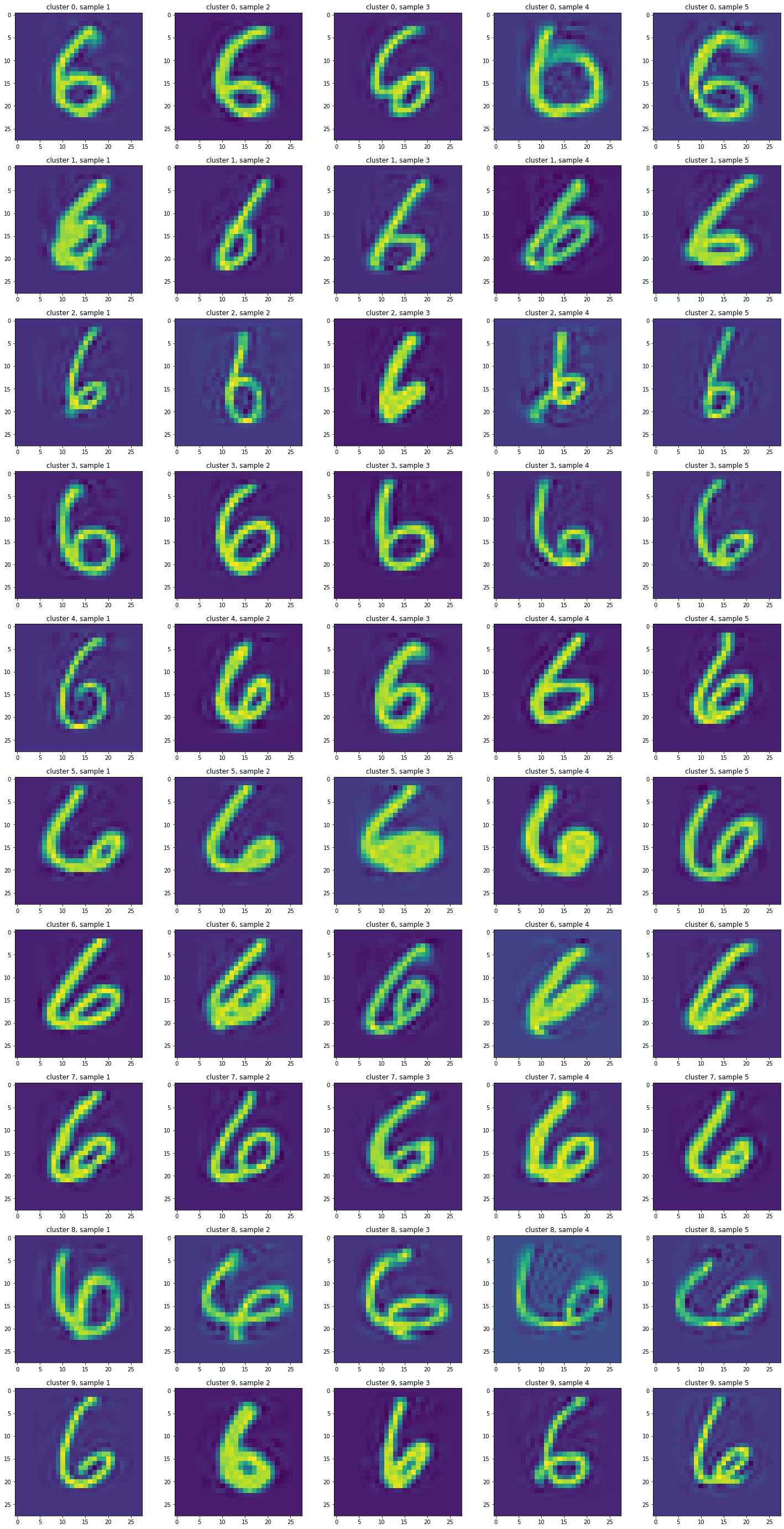
**1.2.1 KMeans**

根据SSE碎石图和轮廓系数图，决定取k=10。



聚类结果可视化如下图所示：

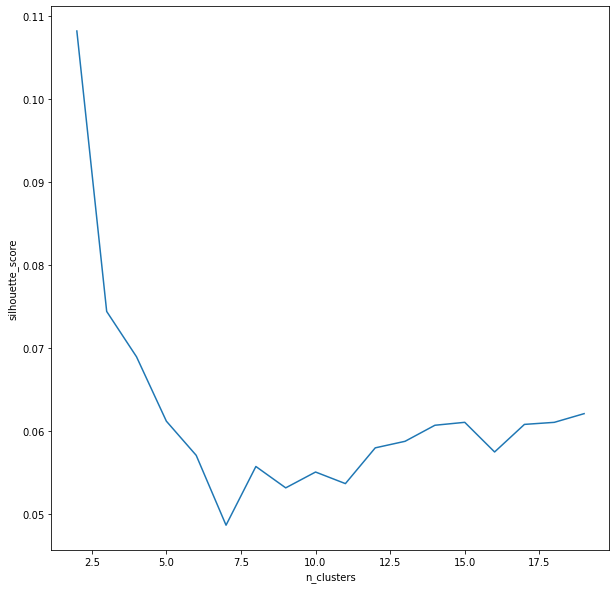




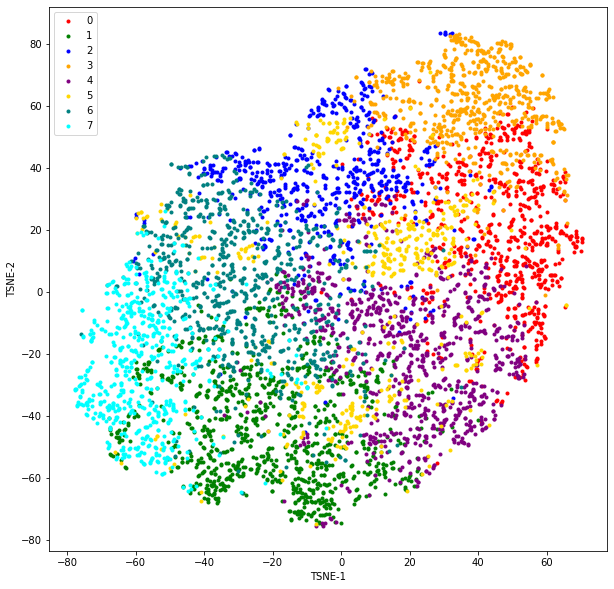
其中，每一行是一个cluster的5个随机采样。可以看到，一些cluster还是聚集出了比较有趣的特征，例如cluster 6 的数字“6”的方向都是倾斜的，cluster 8的数字“6”都像是躺下了一样，cluster 4 就像是正常人类写出的比较规规矩矩的“6”。因此，这个聚类是比较成功的。

**1.2.2 GMM**

根据轮廓系数图，决定取k=8。



聚类结果可视化如下图所示：

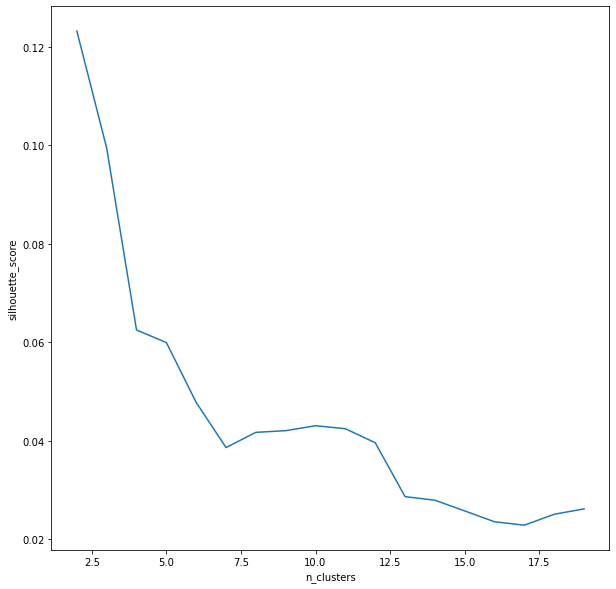




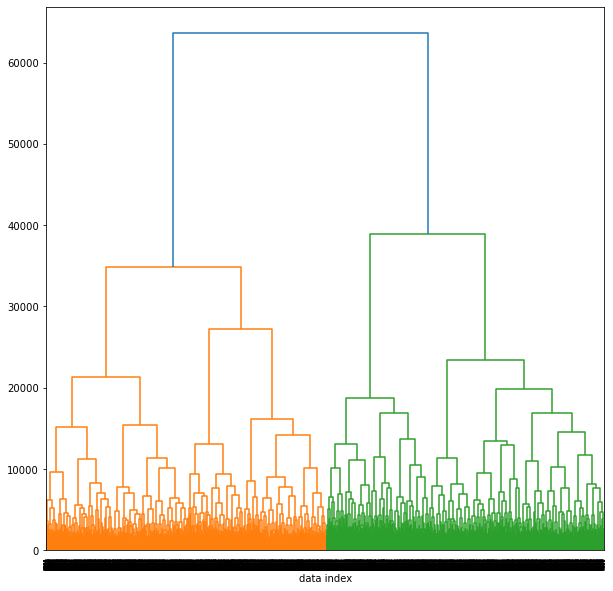
和KMeans类似，这个聚类也是提取出了一些有趣的特征的。

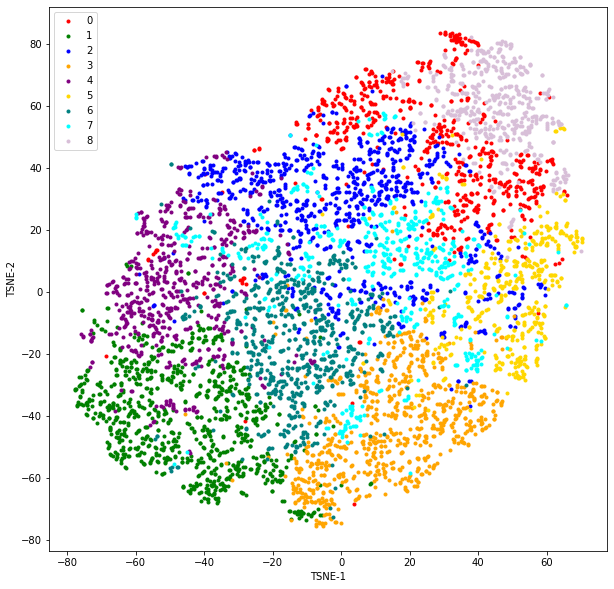
**1.2.3 Hierarchical clustering**

根据轮廓系数，取k = 9。



可视化结果如图所示：



****

****

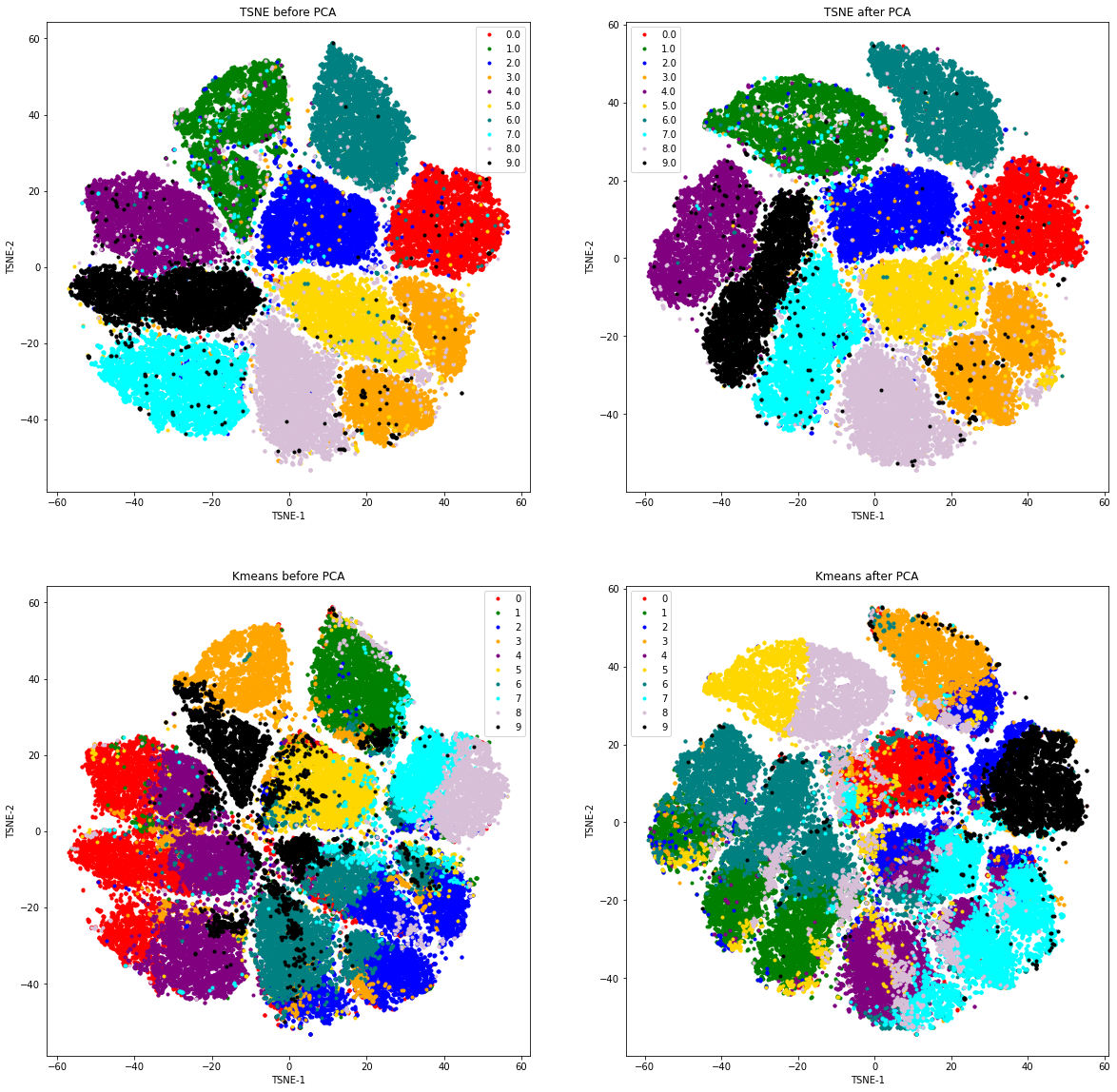
和KMeans类似，这个聚类也是提取出了一些有趣的特征的。

**2. 选择数字9**

在数字9上也进行了相同的实验，得到的结论和数字6一样。受限于篇幅，实验结果就不放在报告中了，负载代码文件里。

**3. 对所有数字进行降维、聚类**

结果如下：

****

发现PCA降维对可视化和聚类都并没有什么作用。