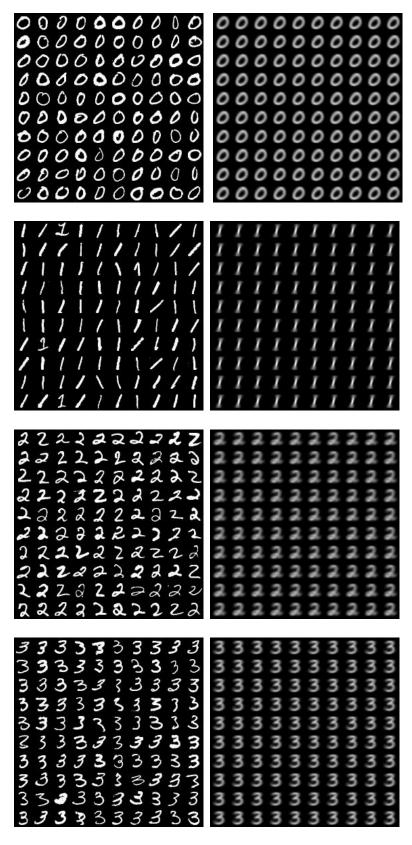
Programming 1: PCA on MNIST

首先从 MNIST_train 数据集中每类数字选取前 1000 张图像进行 PCA,具体代码见 PCA_on_MNIST.py 文件,得到可视化结果如图 1 所示。

由结果图可知,最大的主成分即可以很好地表征出数字的结构特征,虽然有点模糊但仍 能够成功分类。另外还忽略了粗细形状笔触等细节特征。



```
44
 4
   4
    4
            4
             4
                444
    4
      4444
                4444
    4
      444
             4
                44444
            44
   4444
          4
 4
                444
                    4
                      4
 44
    4444444
                444
 4
   4
      444
    4
            4
                444
    4444
 44
          4
                444
                      4
                    4
         4444444
 4.4
    4 4 K
   555
       5
             5
         5
         5
   S
      5
                    5
          5
   5
    55
       5 5
            5 5
                5
         5
          55
              5
 5
   5555
                5
    5555
          55
                5
5
 5
  555555
            55
                5
                 5
 55
    5555
          5 5 5
                5
                 5
   55
      55
            5
         5
           5
             5
                5
 555555
          55
             5
                555
                    5
                      5
                       5
6666666
           666
    6666666
666
666666666666
               666
                      6
   66666666
                666
            66
66666666
                666
6666666666
               666
               6666
 666666666
                      6
                       6
 666666666
               666
6666666666
                66666
                       6
                         66
                            6
                              6
6666666666666
                    6
                      6
                       666
         7
 77
                   7
                    7
                      7
                              7
                       7
                         7
 7
            7
7
   777
       7
          7
              7
                    7
                      7
         7
             マフ
    7
           7
 2
                7
                 7
                      7
       1
      7
            7
         7
           7
            7
       1
 7
   7
    7
      7
                7
                 7
                    7
                      7
 777
          7
              7
      77
                7
         7
            7
                 7
                   7
                7
    7
         7
           7
      7
 7
   7
    7
       7
         7
   7
         7
    777
              7
            7
 7
                7
                 7
                      7
          7
                   7
    77
   7
```

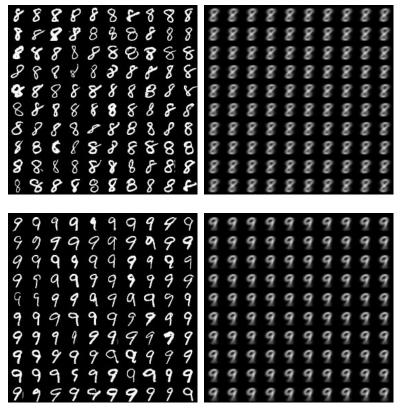


图 1: PCA 结果图

Programming 2: MDS and city distance

实现 MDS 算法,并用 ppt 中 8 个城市之间的交通所需时间来表征城市之间的不相似程度。结果如图 2 所示,基本上能够表示城市间的关系,例如北京到湖北、黑龙江、广州都比较近,四川九龙离其他城市都比较远。

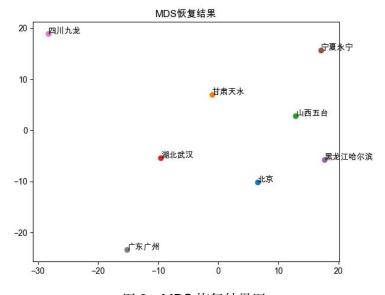


图 2: MDS 恢复结果图