

# Homework 02 PCA

马倩

2016 年 5 月 25 日

## 1 实验内容

使用 PCA 方法对手写数字数据进行降维，选取最大的两个特征向量，将数据投影到这两个特征向量对应的方向上。

## 2 实验方法

PCA 过程：

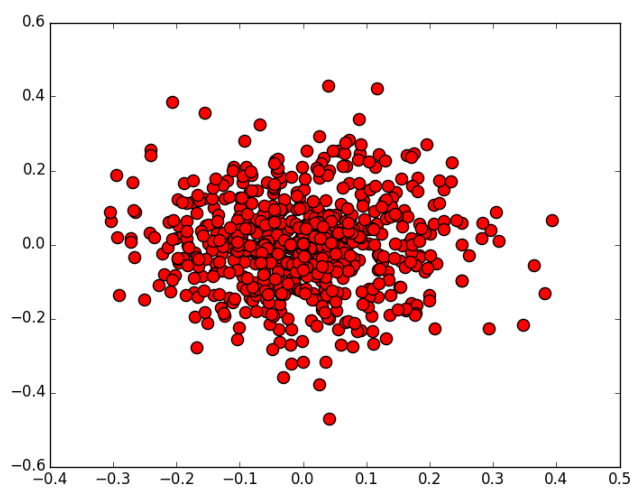
- (1) 特征中心化：每一维特征数据都减去该维的均值。
- (2) 计算特征中心化后数据的协方差矩阵。
- (3) 计算协方差矩阵的特征值和特征向量。
- (4) 选取最大的两个特征值对应的特征向量。
- (5) 根据所选特征向量，将数据映射到对应的方向。

## 3 实验数据及结果

数据：optdigits-orig.wdep 数据集中所有手写数字 3 的数据， $1024 \times 105$ 。  
结果如图。

## 4 讨论与分析

PCA 在对高维数据进行降维的同时去除了数据中的噪声，发现了数据中的模式。PCA 把原先的  $N$  个特征用数目更少的  $M$  个特征取代，新特征是旧特征



的线性组合，这些线性组合最大化样本方差，尽量使新的  $M$  个特征互不相关。用从旧特征到新特征的映射捕获数据中的固有变异性。

## References

- [1] <http://www.cnblogs.com/zhangchaoyang/articles/2222048.html>