# 做笔记

* 概述

**主成分分析算(PCA)是最常用的线性降维方法，它的目标是通过某种线性投影，将高维的数据映射到低维的空间中，并期望在所投影的维度上数据的信息量最大（方差最大），以此使用较少的数据维度，同时保留住较多的原数据点的特性。**

**PCA降维的目的，就是为了在尽量保证“信息量不丢失”的情况下，对原始特征进行降维，也就是尽可能将原始特征往具有最大投影信息量的维度上进行投影。将原特征投影到这些维度上，使降维后信息量损失最小。**

**应用场景**

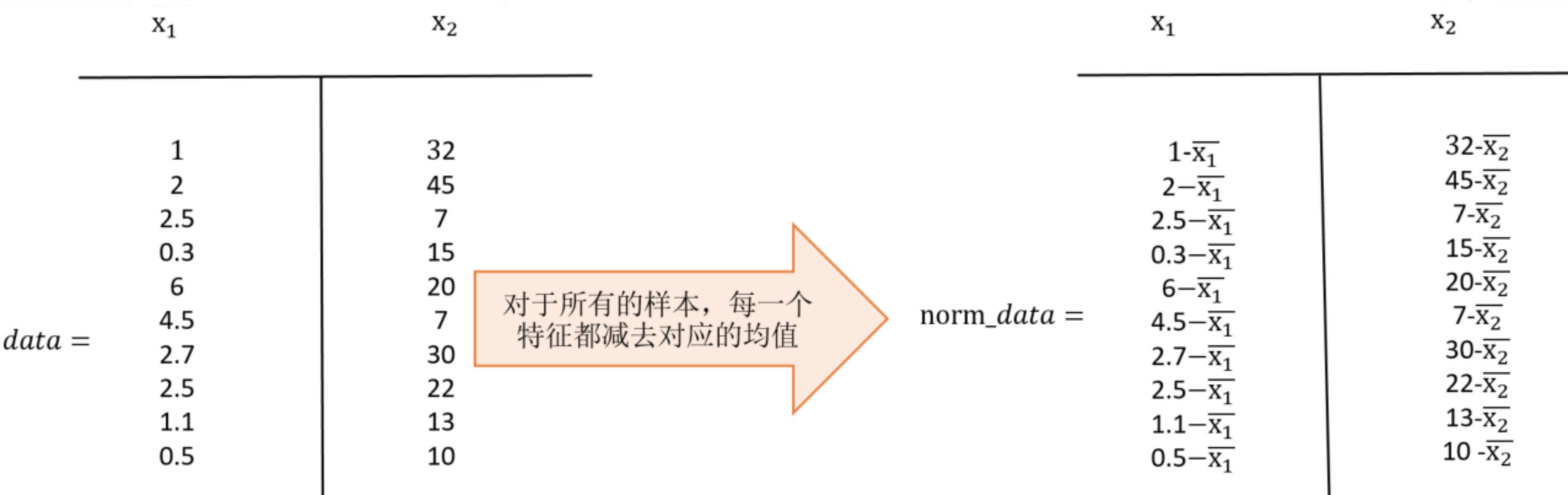
**1.降维**

**2.降噪**

**求解步骤**

**1.样本均值化(中心化)**

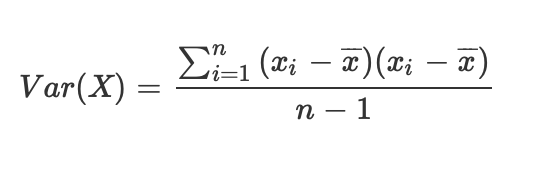
**求每一个特征的平均值，然后对于所有的样本，每一个特征都减去自身的均值**

****

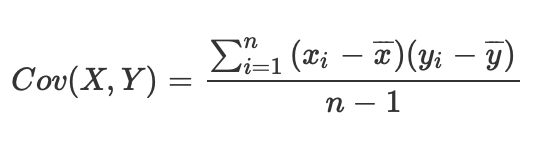
**2.计算协方差矩阵**

**协方差就是衡量两个变量相关性的变量。当协方差为正时，两个变量呈正相关关系(同增同减)当协方差为负时，两个变量呈负相关关系(一增一减)。**

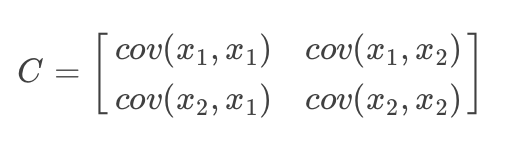
**方差**

****

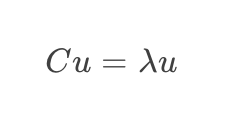
**协方差**

****

**协方差矩阵**



**3.计算协方差矩阵的特征值和特征向量**

****

**特征值λ会有N个，每一个λi对应一个特征向量ui，将特征值λ按照从大到小 的顺序排序，选择最大的前k个，并将其相对应的k个特征向量拿出来，我们会得到一组{(λ1，u1)，(λ2，u2)，...，(λk，uk)}**

**4.将特征值排序**

**选取最大的前k个特征值和相对应的特征向量，并进行投影的过程，**

**就是降维的过程。**

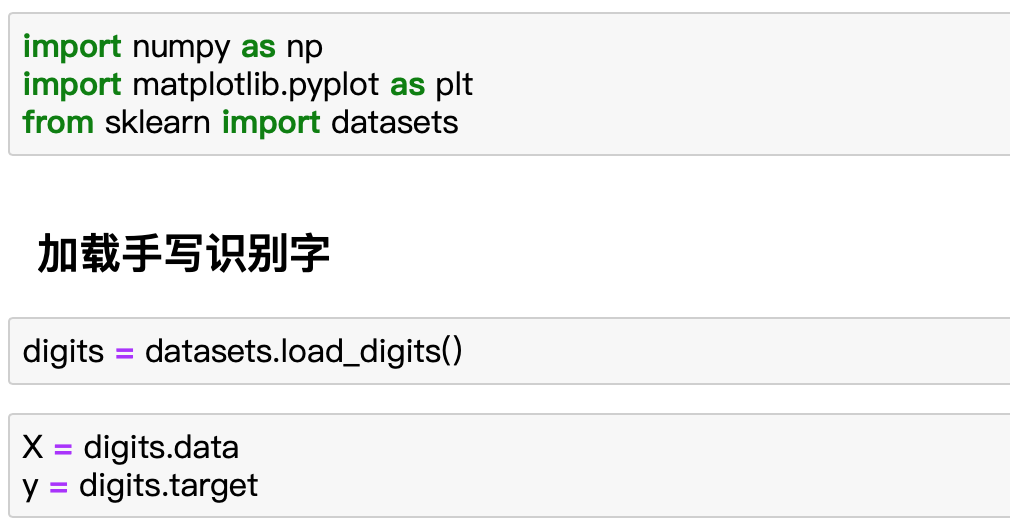
**5.保留前N个最大的特征值对应的特征向量**

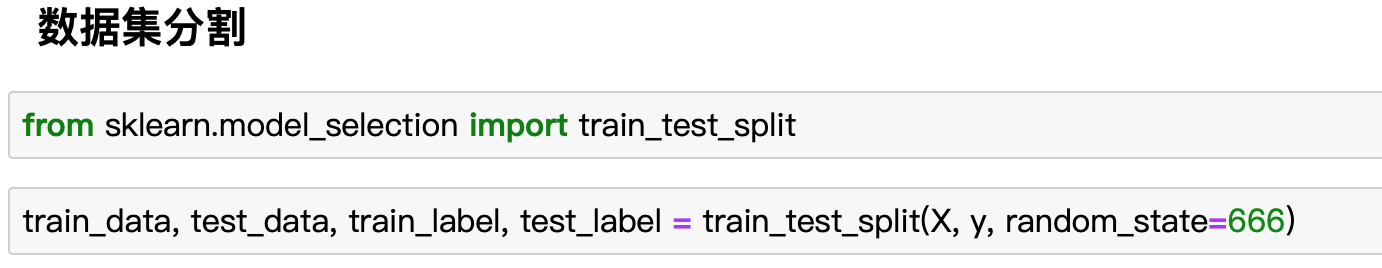
**6.将原始特征转换到上面得到的N个特征向量构建的新空间中**

**为什么样本在（协方差矩阵C的最大K个特征值所对应的特征向量）上的投影就是k维理想特征？**

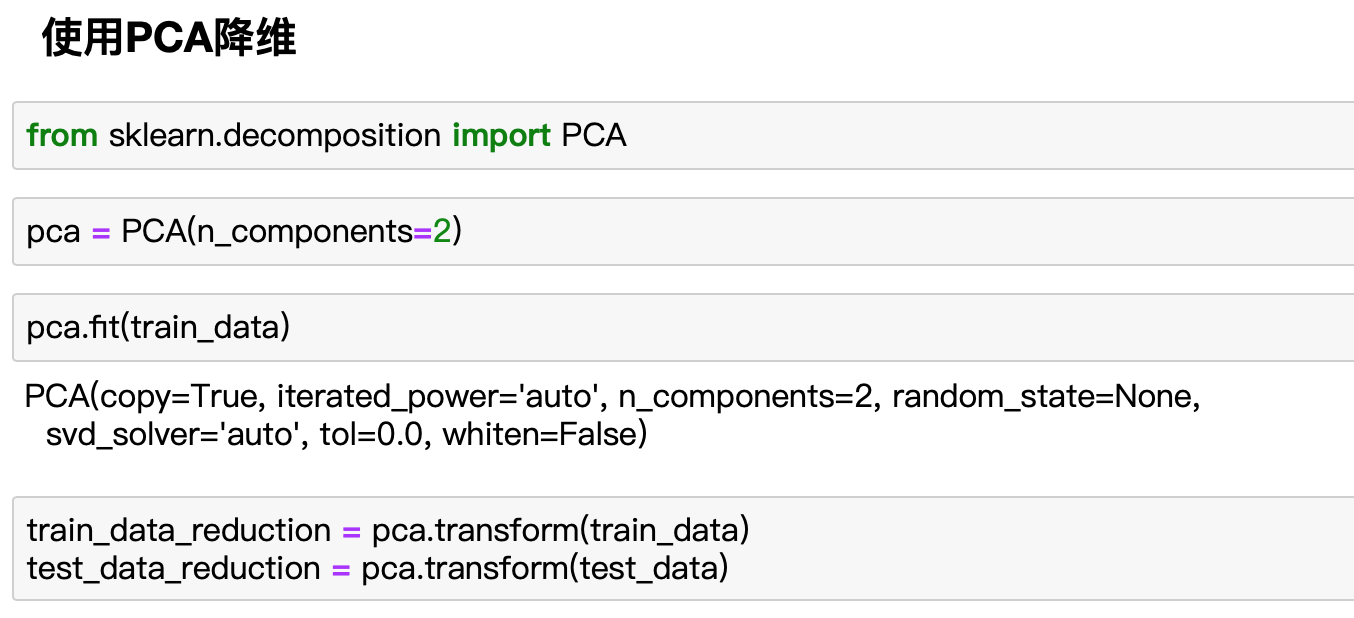
**最大方差理论:方差越大，信息量就越大。协方差矩阵的每一个特征向量就是一个投影面，每一个特征向量所对应的特征值就是原始特征投影到这个投影面之后的方差。由于投影过去之后，我们要尽可能保证信息不丢失，所以要选择具有较大方差的投影面对原始特征进行投影，也就是选择具有较大特征值的特征向量。然后将原始特征投影在这些特征向量上，投影后的值就是新的特征值。每一个投影面生成一个新的特征，k个投影面就生成k个新特征。**

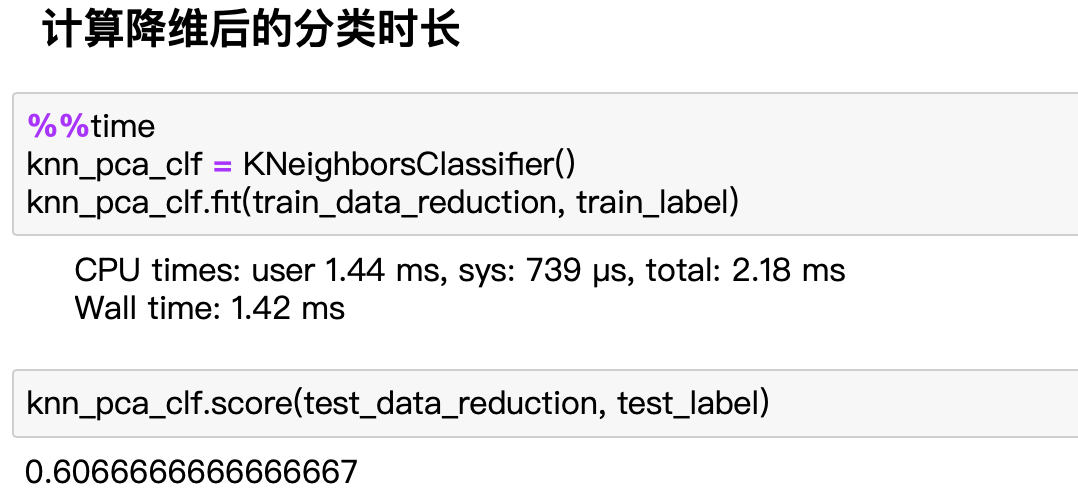
* PCA降维实现

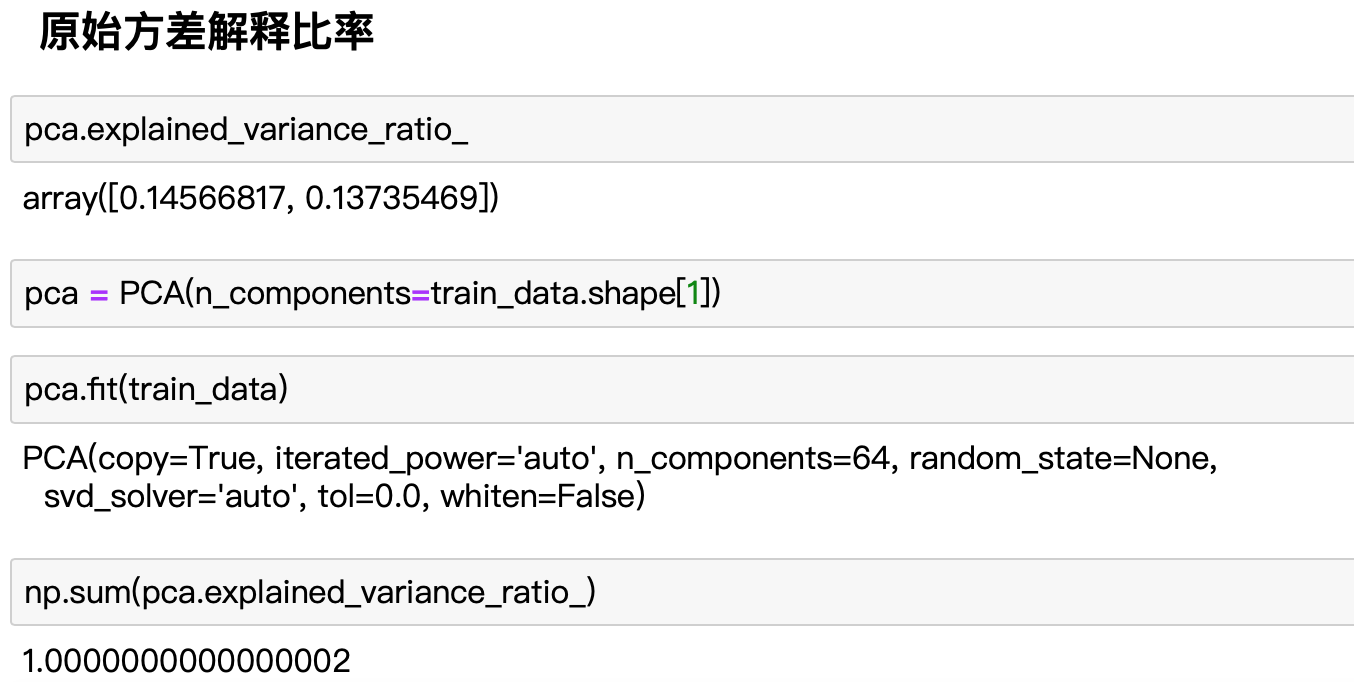
****

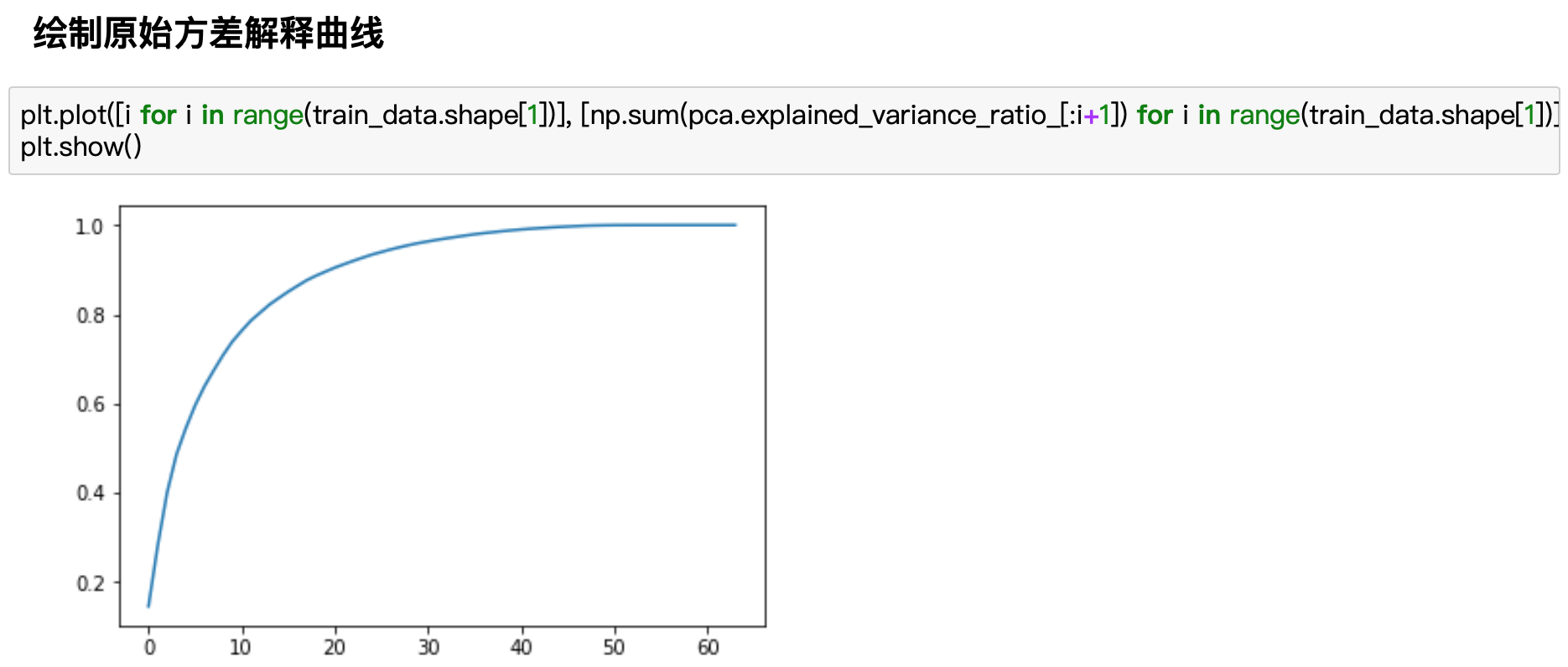
****

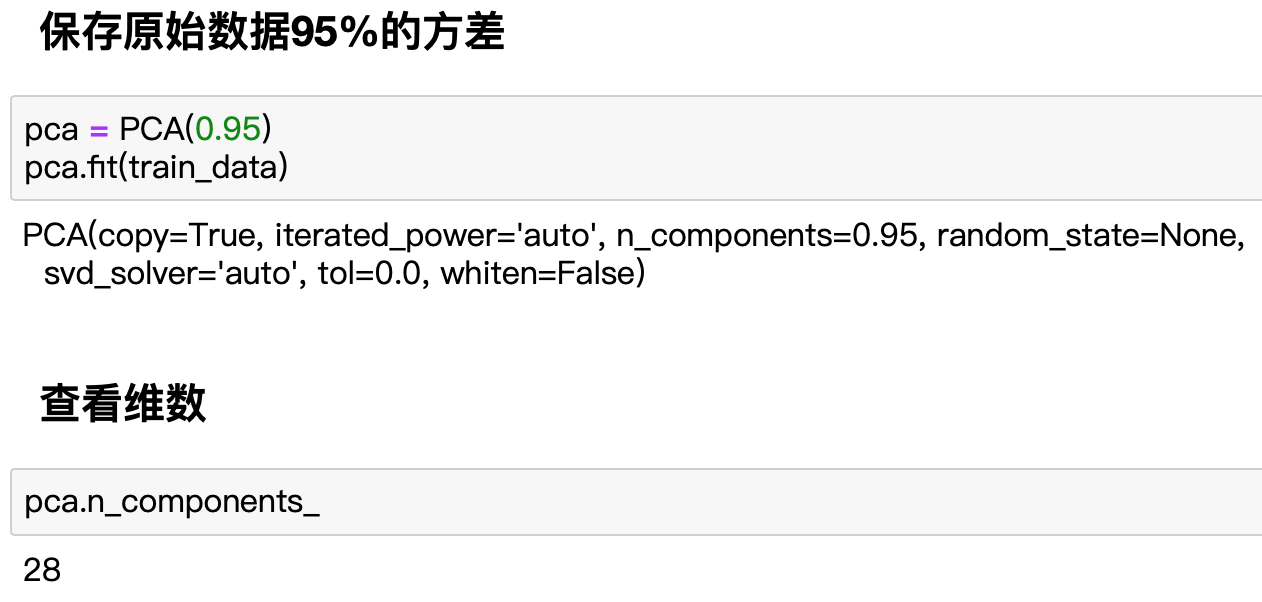
****

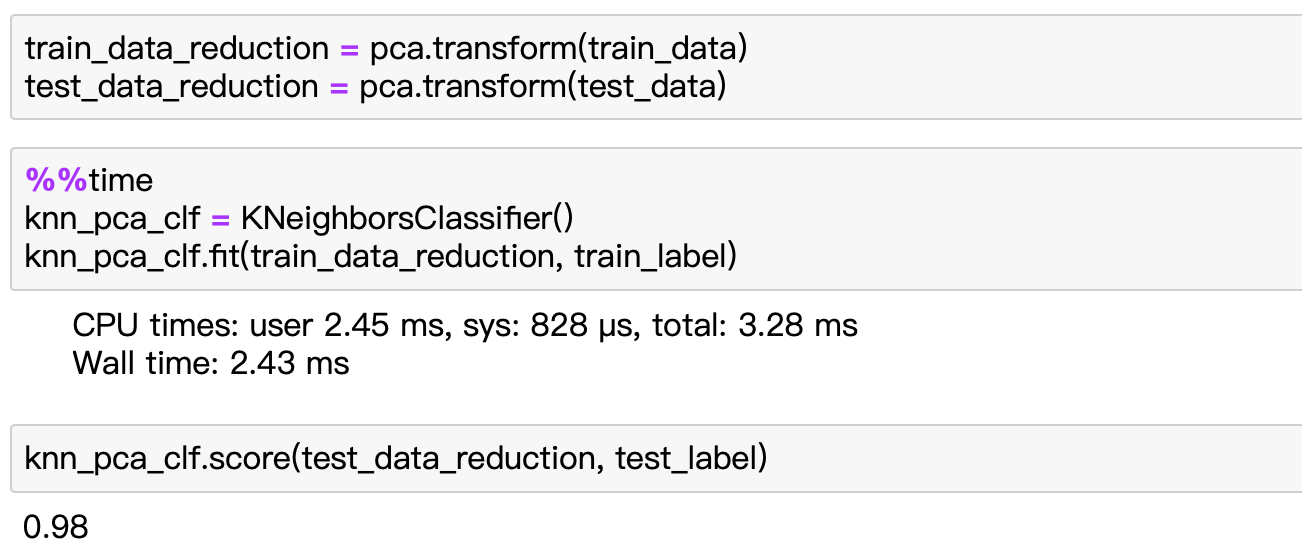
****

****

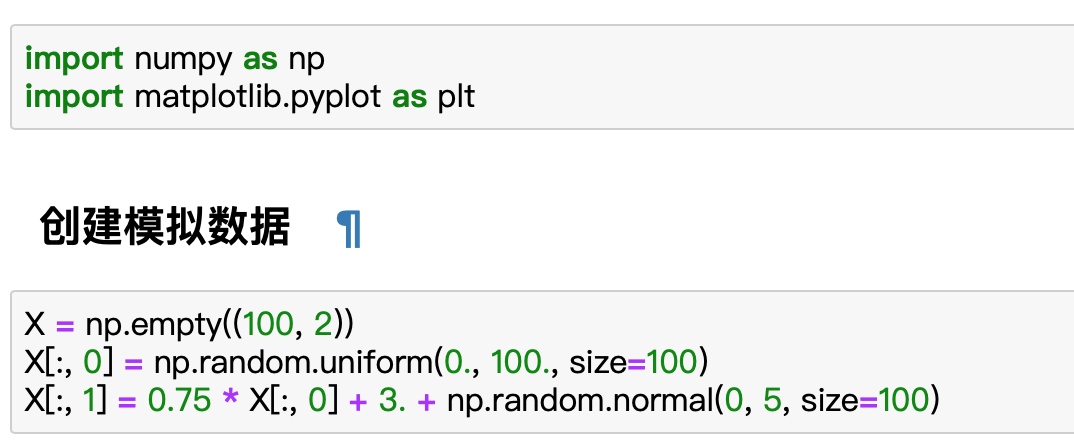
****

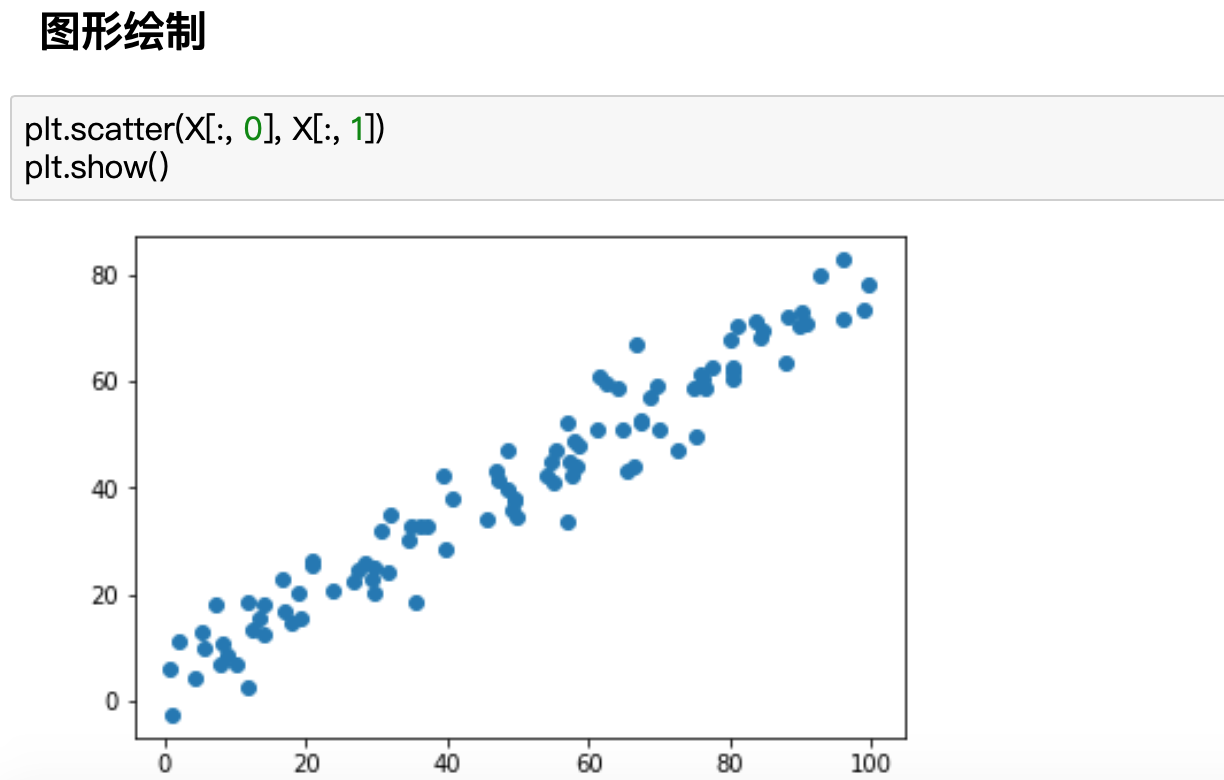
****

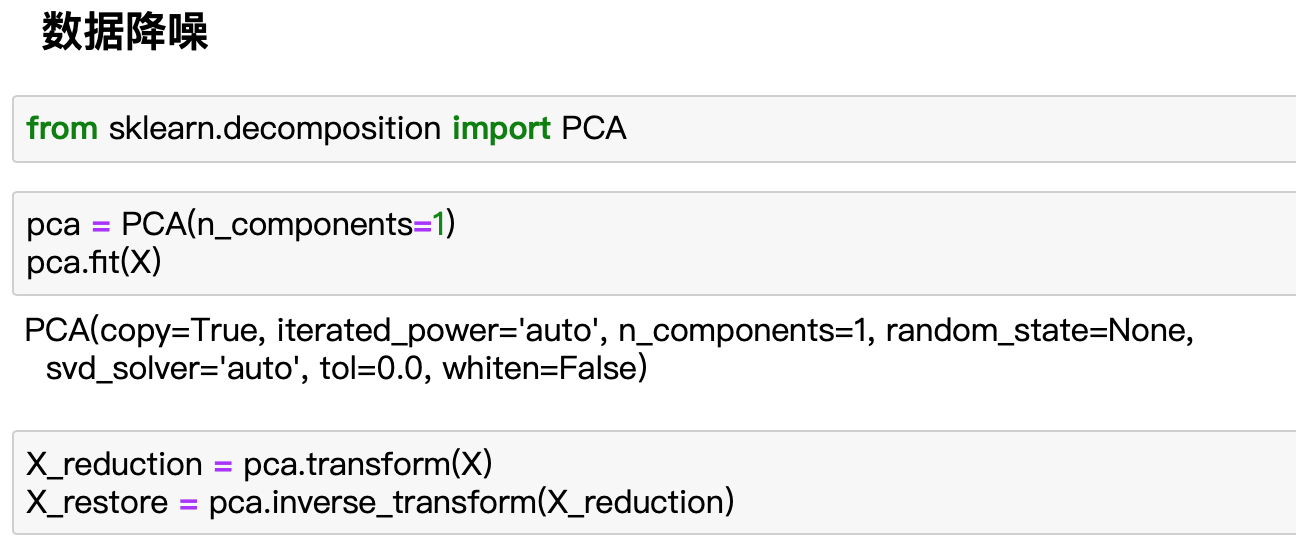
****

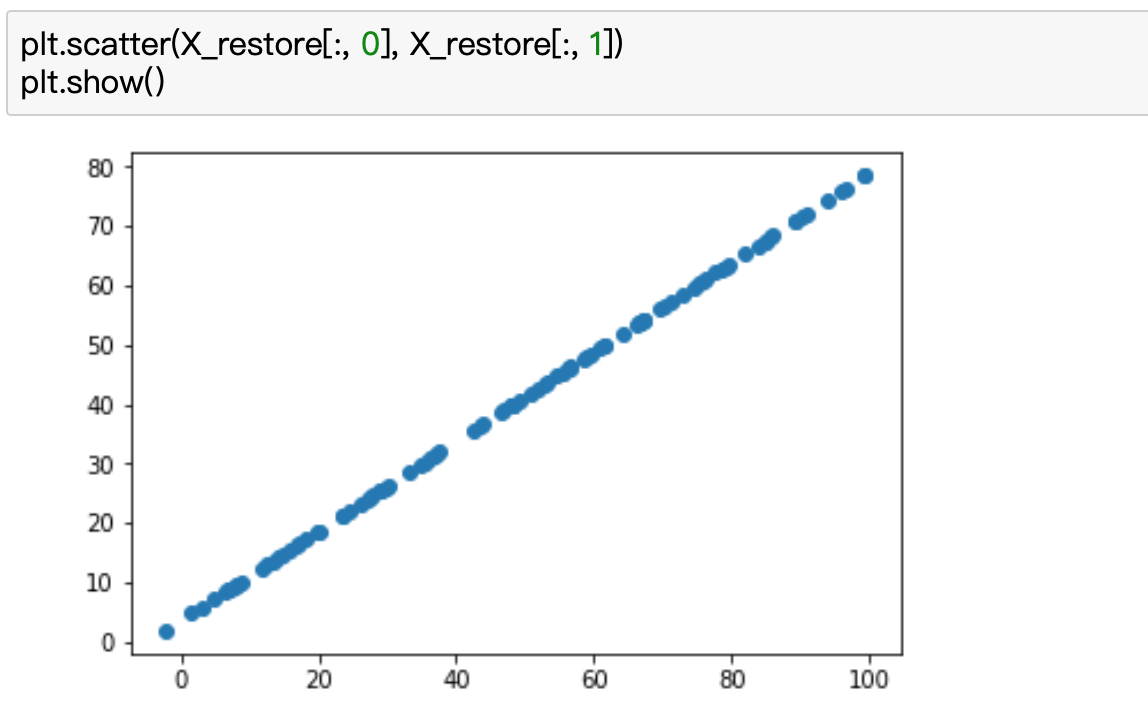
****

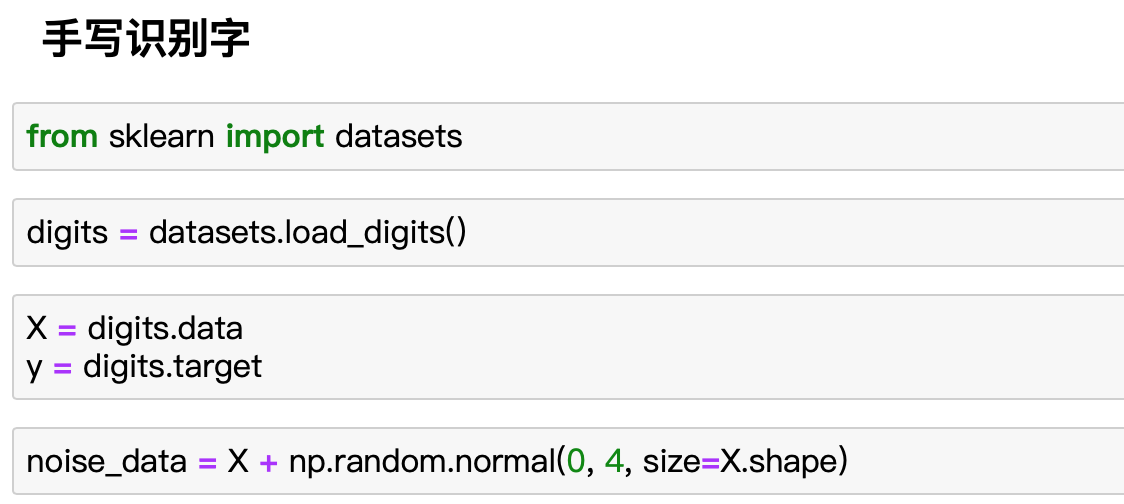
* PCA降噪实现

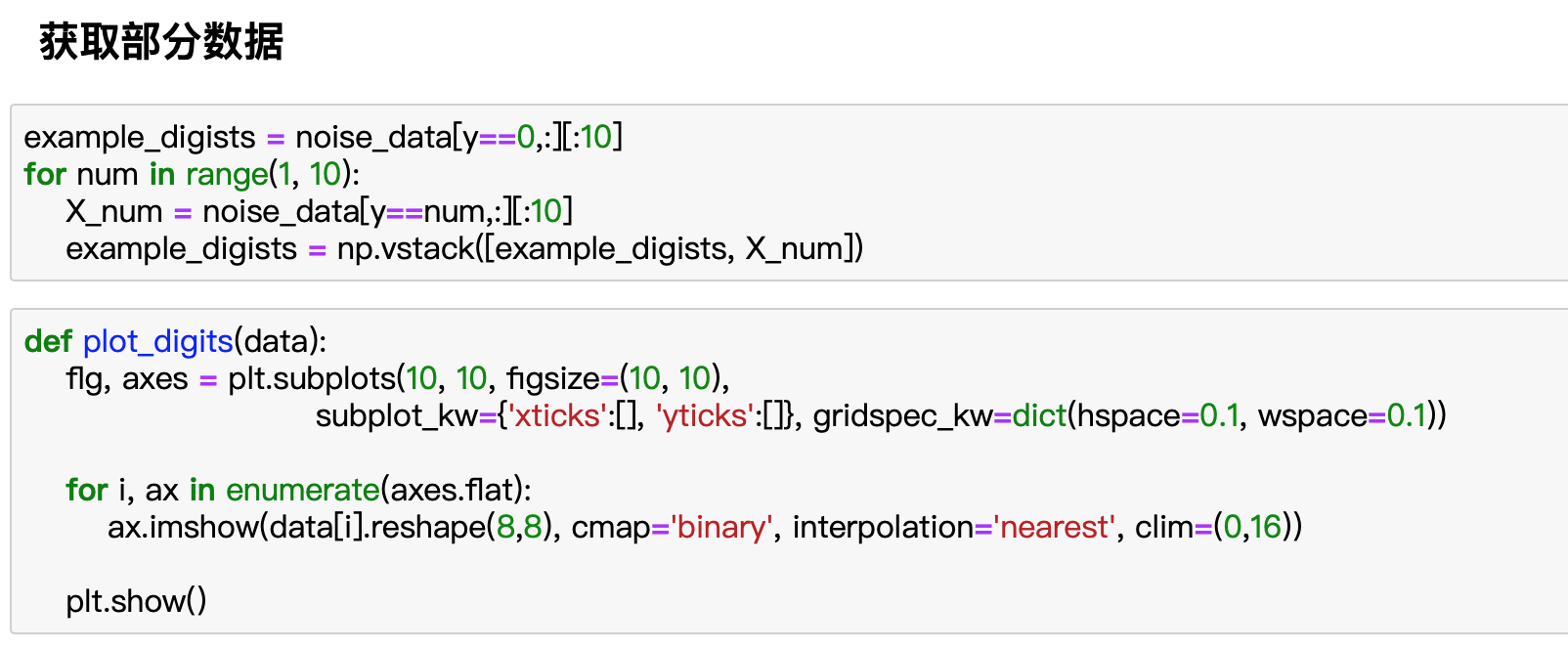
****

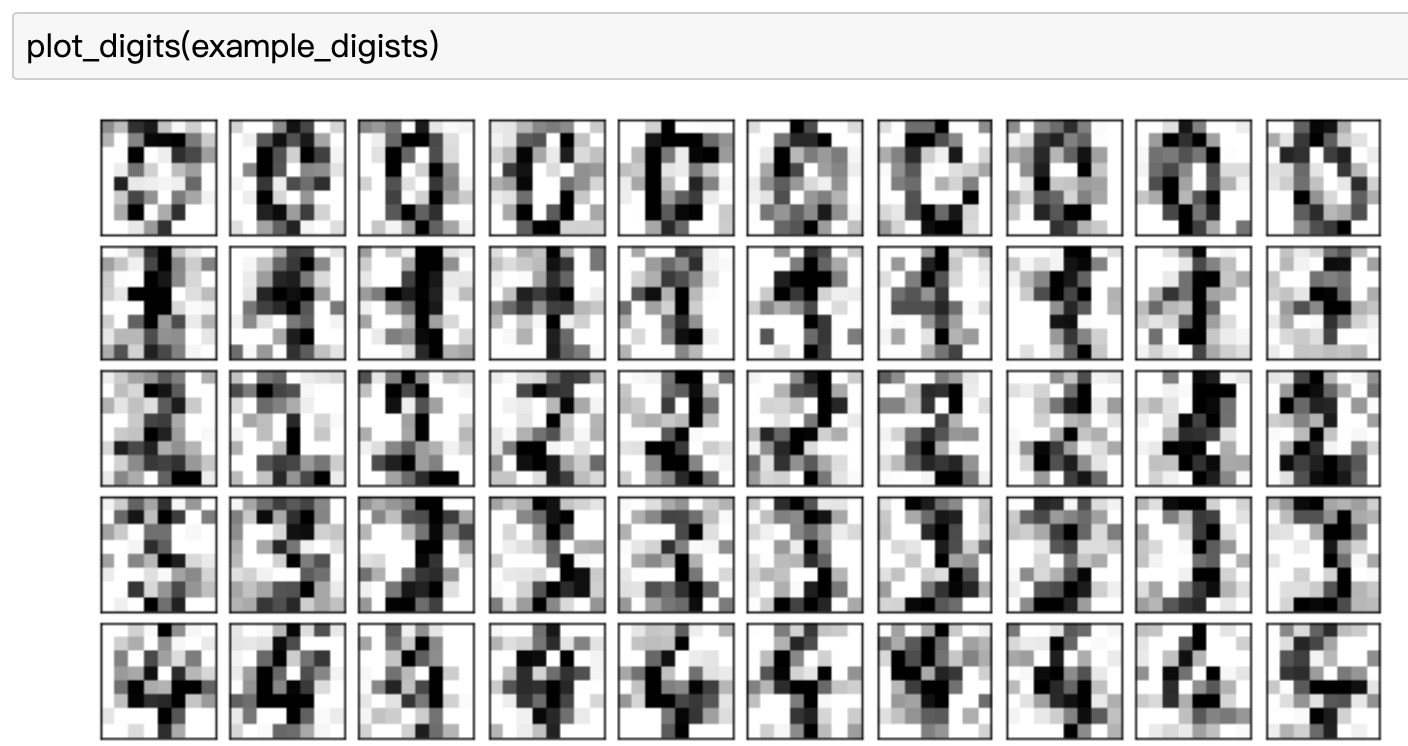
****

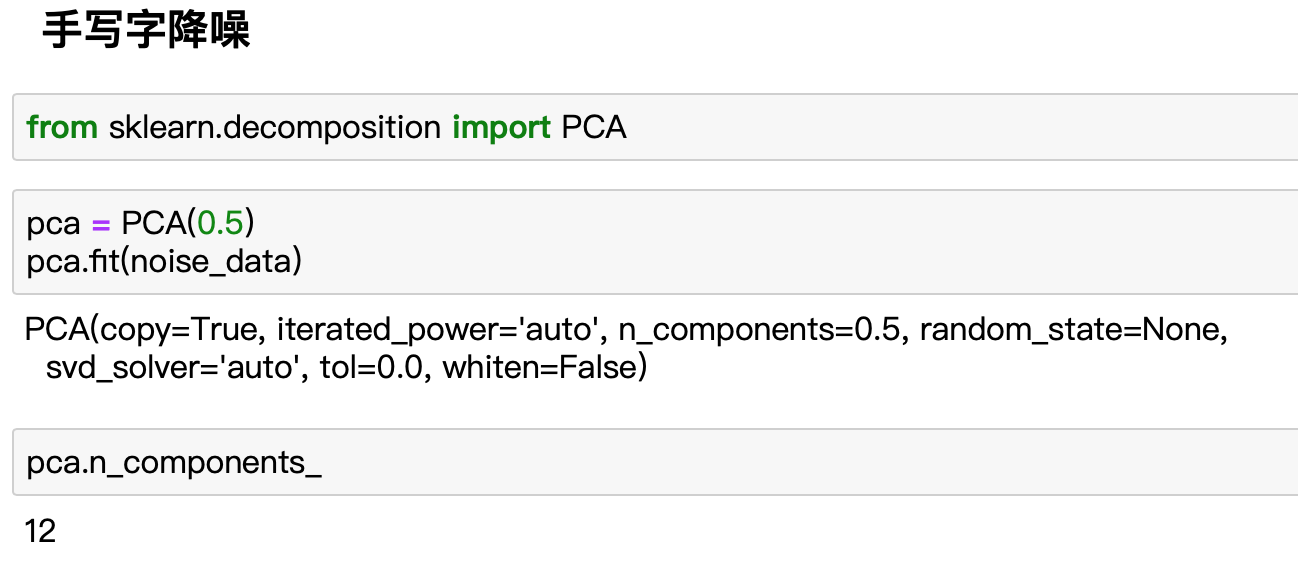
****

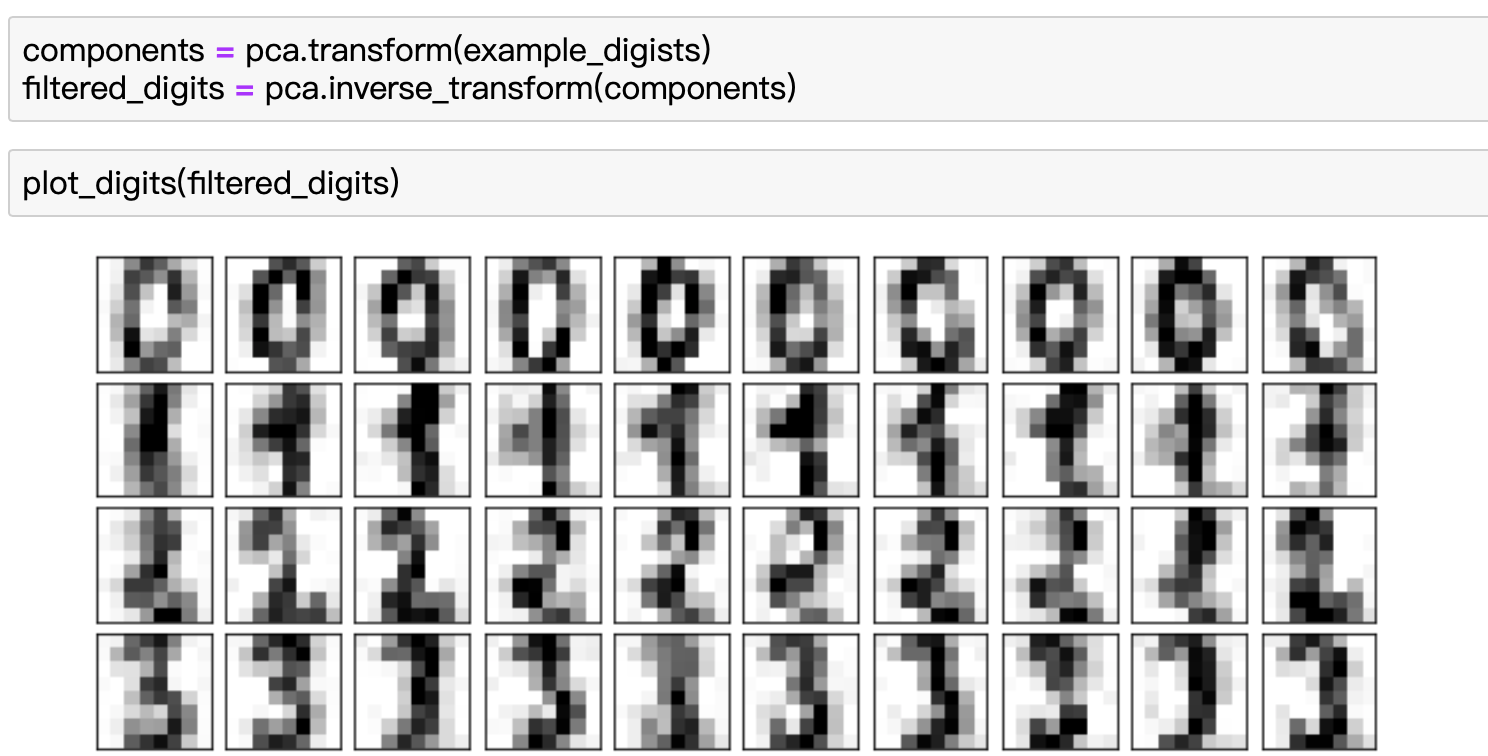
****

****

****

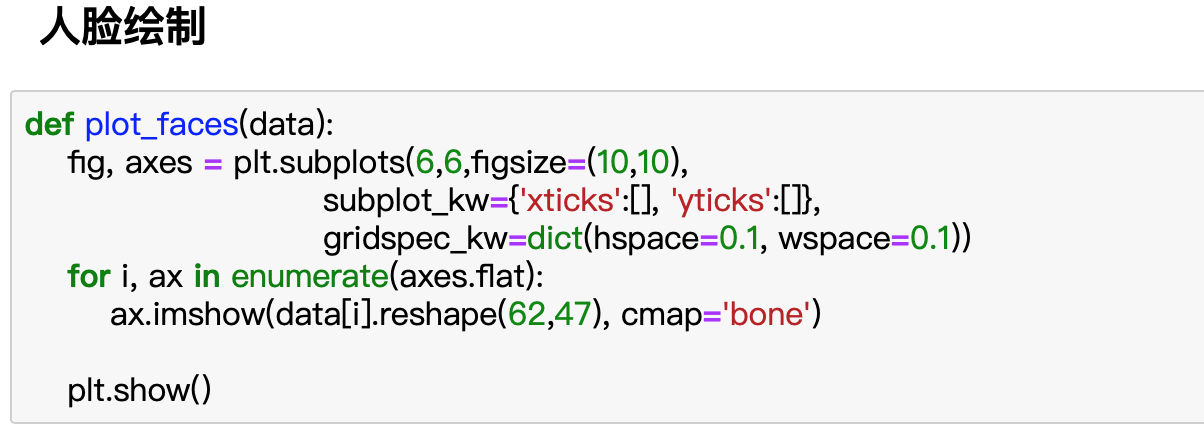
****

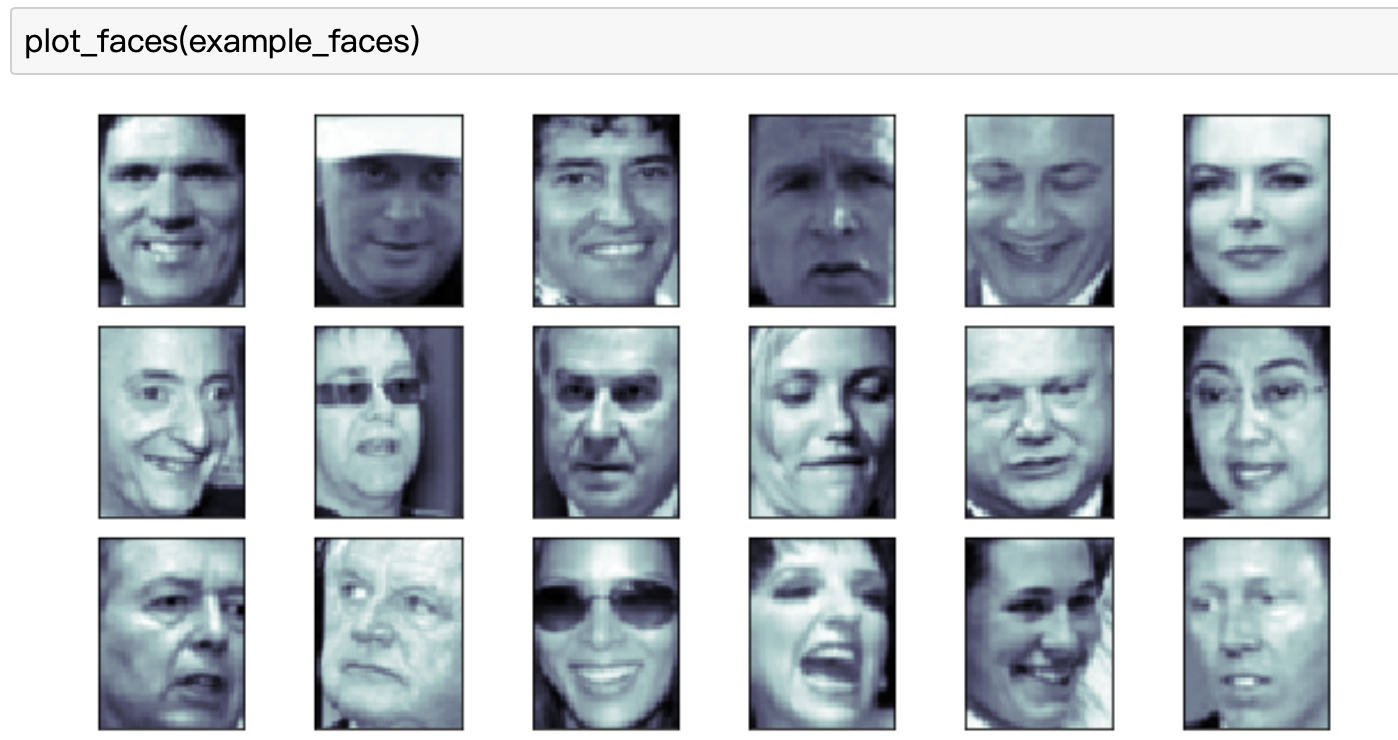
****

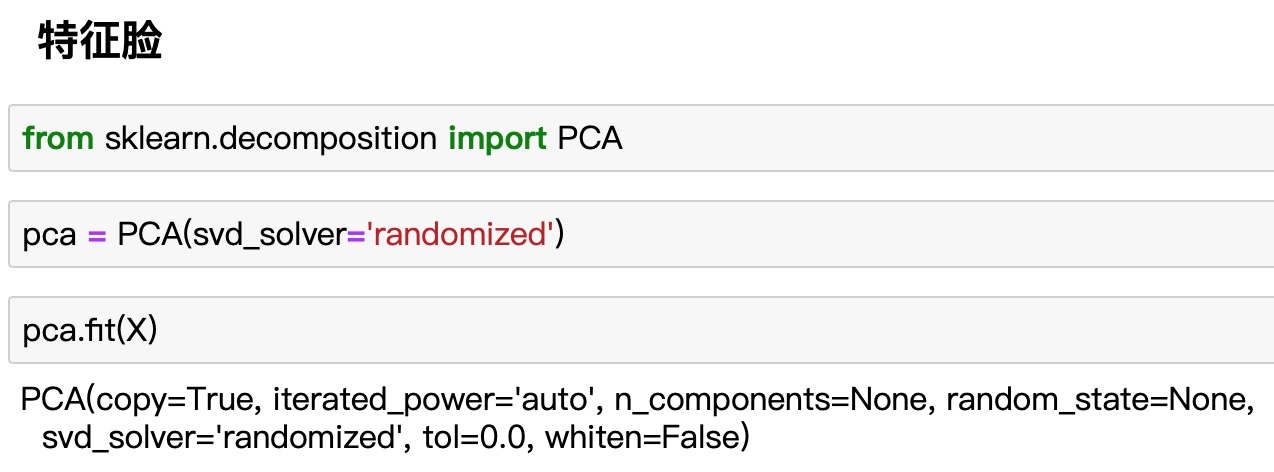
****

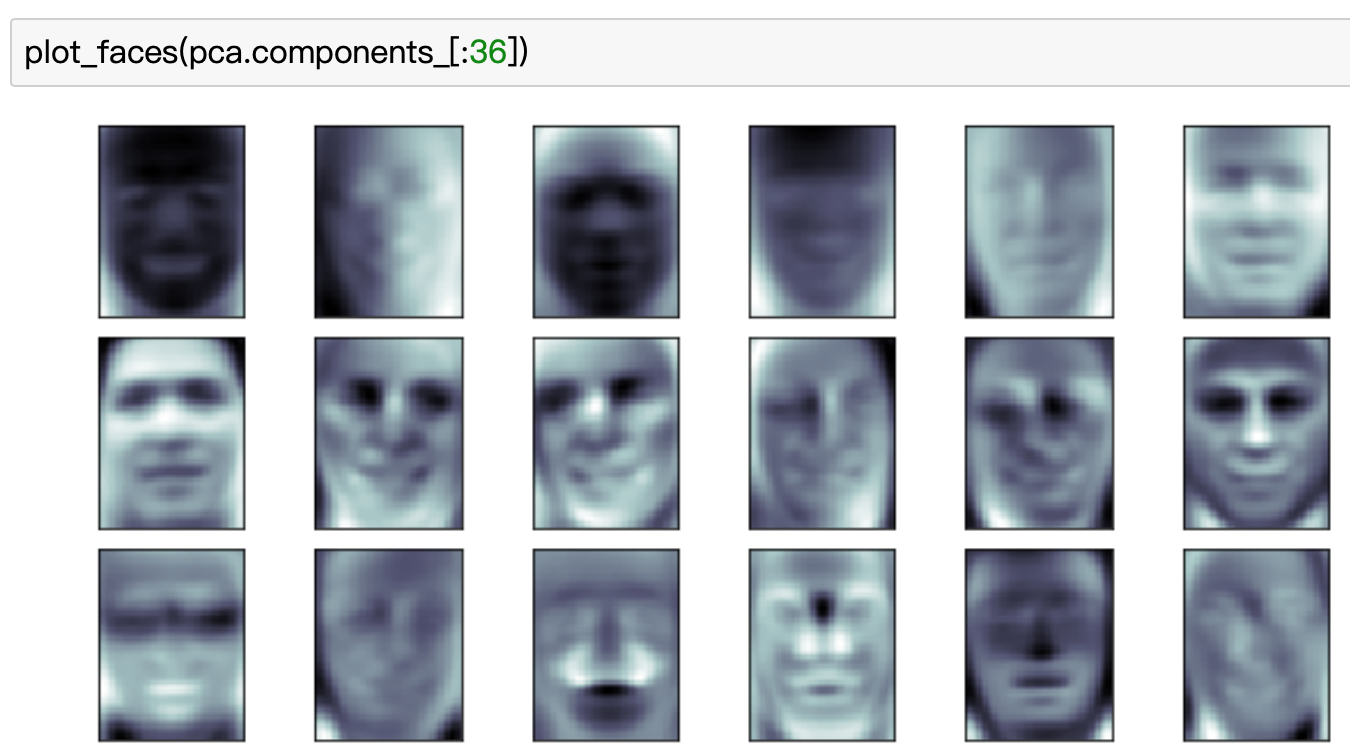
* PCA特征脸实现

****

****

****

****

****