****

**模式识别与深度学习课程**

**实验2、PCA降维SVM分类算法实验报告**



**学 院 智能与计算学部**

**专 业 计算机与科学技术**

**学 号 3019244140**

**姓 名 郭思齐**

# 1．实验目标

实验二——PCA降维与SVM分类算法，实验总目标是帮助我们掌握PCA将维算法原理及代码实现，对已有算法灵活调用获取降维重建结果。根据实验数据的特性调节PCA降维数以及SVM线性函数或核函数。其中两个小实验的目标分别是：

1. 将PCA人脸降维重建算法理解，更改给出的重建代码，实现对本班级采集人脸数据进行PCA降维重建实验，并调试参数n\_components的不同取值，查看不同取值对实验结果的影响，将结果记录分析。
2. 补全pca\_svm.py代码，使用本班级人脸采集数据进行PCA降维，然后使用降维后的数据进行SVM分类实验。调试不同参数，查看不同参数对实验结果的影响，将不同参数对应的训练集和测试集准确率通过表格记录，并对结果进行分析。

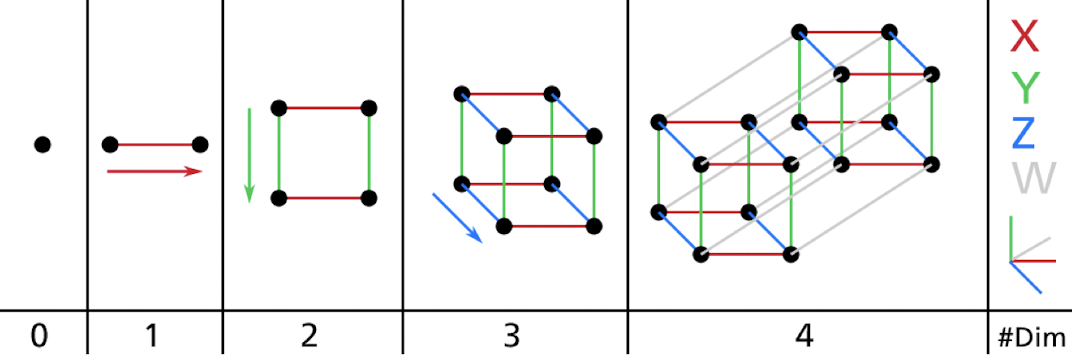
# 2. 实验一

## 2.1 算法实现及参数调节说明

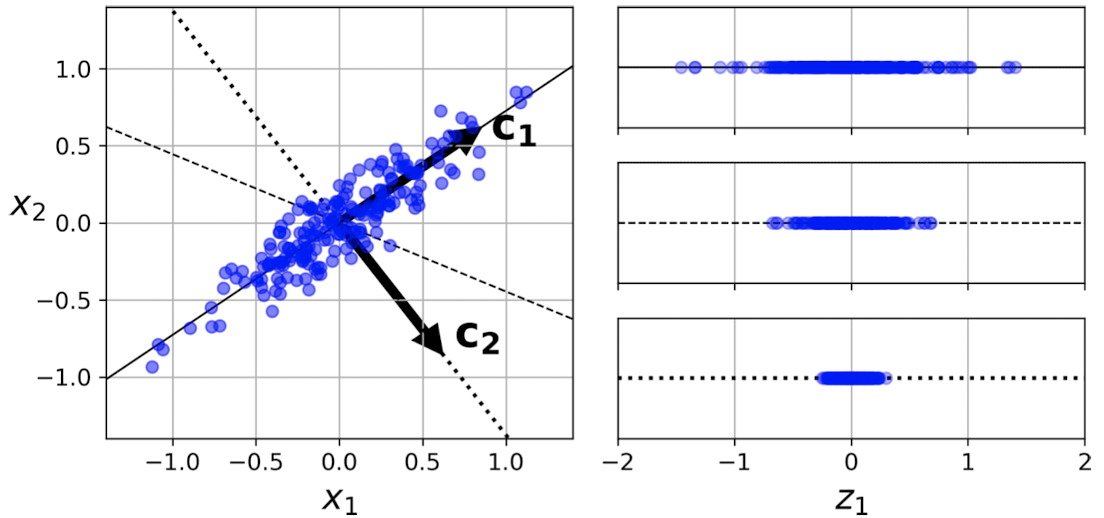
应用PCA降维模型对灰度化处理后的图片进行降维，并利用PCA重建模型和降维后的主成分向量进行人脸的重建。

PCA的引入：

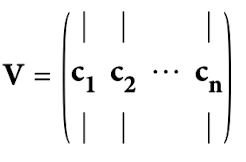
1. 随着维度的增长，点间的距离增大。维度越大，过拟合的风险越高。除此以外，空间的复杂度也会如下图一般增大。



1. 在每个训练实例涉及到成千上万的特征时，机器学习相关的训练会耗时巨大，并且不容易获得优解，导致维度灾难，采用诸如PCA（映射）或LLE（流形学习）这样的算法可以对高维数据维数约简、降维处理。
2. PCA定义一个超平面并将数据映射到上面。PCA可能会通过找到保存最多训练变量的轴或是最小化卷方差距离的轴，选取为需要引射到的轴。这些轴就是Principal Component (PC)。



通过SVD（Singular Value Decomposition）来将训练集矩阵X分解为三个矩阵乘积。其中V是主成分矩阵。



1. PCA用来压缩（映射到d维度），解压可以有如下的等式：

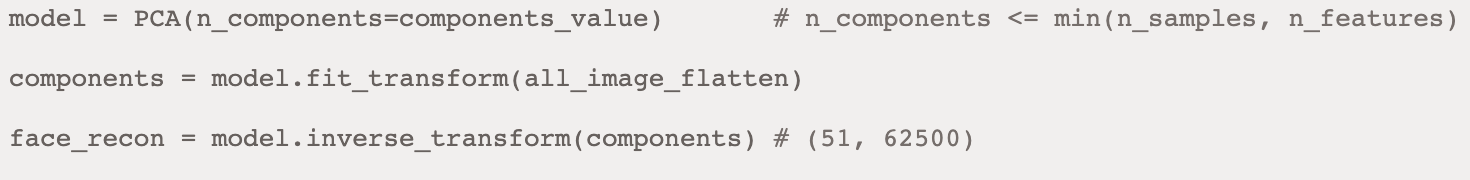
图像重建也是基于这样的思想。

算法实现的核心：

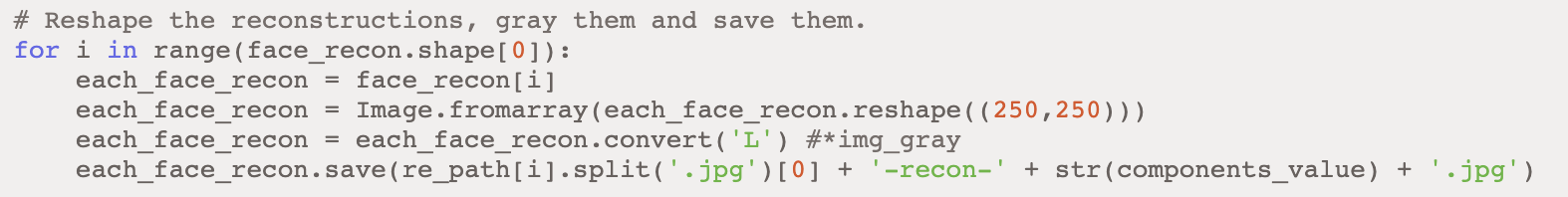
1. 数据导入、预处理以及灰度化：



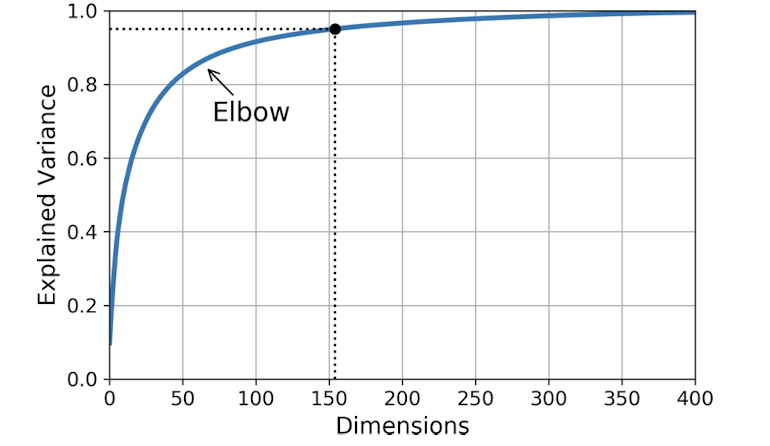
1. PCA降维以及图片重建：



1. 数据灰度化图像保存：



1. 使用多组n\_component 作为传参比较实验。其中PCA的n\_components与可解释变量的比例有一定的关系，如下图。（1）当它被设置为0-1之间的浮点数，这就表示希望保留的信息量的比例。（2）也可以设置其参数为"mle"，即让PCA用最大似然估计自己选择超参数。（3）表示将至的维数。



1. 实验结果整理：

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Reconstruction | Original Photo | n\_component = 1 | n\_component = 5 | n\_component = 10 | n\_component = 20 | n\_component = 30 | n\_component = 40 | n\_component = 50 |
| |  | | --- | | Gao Han | |  |  |  |  |  |  |  |  |
| Gao Leyu |  |  |  |  |  |  |  |  |
| Guo Siqi |  |  |  |  |  |  |  |  |
| Guo YongXu |  |  |  |  |  |  |  |  |
| Hua Yi |  |  |  |  |  |  |  |  |

## 2.2 结果分析

使用的n\_components越大，在PCA映射的时候维度就越多，相当于在压缩数据的时候，丢失的信息就越少，重建后表示的维度就越高，恢复后可表示的信息也越多，和原图相比就越相似。

其次，对比不同的同学，可以看出来背景色与肤色反差度大的同学面貌经PCA映射的时候，信息重建时更完整。而本身图像不够清晰或者饱和度比较弱的图像自然很容易就丢失掉特有的信息，PCA重建的时候会更困难。

# 3. 实验二

## 3.1 算法实现说明

补全给定的代码实现本班级人脸数据的pca降维svm分类实验，记录不同参数的实验结果，结果通过表格进行展示，请注意排版美观。

## 3.2 结果分析

根据实验结果分析哪些参数对新数据实验结果影响比较大，并尝试解释原因。

# 

# 4. 总结

实验二相较实验一更进一步，除了应用到SVM分类算法以外，加入了PCA主成分降维算法，以及对降维后的数据进行SVM的分类实际应用。

在这一实验中，深入学习了使用LFW数据集进行PCA降维重建的工程案例，并迁移学习，自己动手对班级内同学们的人脸图像进行PCA降维重建，以及运用之前所学的SVM支持向量机算法对降维后的SVM分类。