

模式识别作业3：在人脸数据库上应用PCA

- 给定了face文件夹，其中有train和test两个文件夹。利用train中的人脸数据训练主成分分量，并完成以下练习。
- 训练集（19X19 500张图片）



(a) 从train文件夹里随意取出一张图片向量 x ，将 x 投影到前 K 个主成分中，然后利用这些投影分量来重建人脸 x' ，并计算重建误差 $\|x' - x\|^2$ 。从 $K=1$ 开始，不断的增加 K 。给出重建误差随 K 的增长的收敛曲线。重建误差能否为0？

模式识别作业3：在人脸数据库上应用PCA

- 给定了face文件夹，其中有train和test两个文件夹。利用train中的人脸数据训练主成分分量，并完成以下练习。
- 测试集（19X19 2张图片）

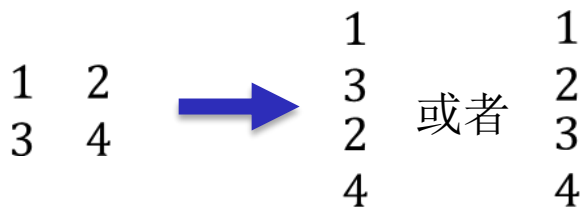


（b）从test文件夹里读取文件名为”face.jpg”的文件，按照(a)的方式来做。与(a)相比，对于相同的误差阈值，是否需要更大的K？重建误差能否为0？

（c）从test文件夹里读取文件名为”nonface.jpg”的文件，按照(a)的方式来做。与(b)相比，对于相同的误差阈值，是否需要更大的K？重建误差能否为0？

提示1：图像数据的处理

- 一幅图像转换为灰度图像后是一个二维矩阵，需要将之拉伸成为一个列向量来处理。



- 一个2x2图像转换的例子，按行或按列拉伸都可以

- 在matlab中可以通过imread()函数来读取文件，可以通过reshape()函数来将矩阵拉伸成向量，函数的具体使用方法请查看matlab的帮助文件

提示2：人脸的PCA训练

■ 计算输入训练样本的均值和协方差阵

$$C = \frac{1}{m} \sum_{k=1}^m (x_k - \mu)(x_k - \mu)^T \quad \mu \text{ 是样本均值}$$

■ 对协方差阵进行特征值分解

$$Cu_i = \lambda_i u_i$$

选取最大的前K个特征值对应的特征向量作为主成分分量，构成变换矩阵 $W = [u_1, \dots, u_d]$

提示3：投影和重构

- x 为输入图像，投影到特征空间得到特征 y

$$y = W^T(x - \mu)$$

- 利用特征脸空间的投影向量来对人脸进行重构

$$x_r = \mu + Wy$$

其中 W 为变换矩阵， μ 为训练样本的平均值