

Pattern Recognition & Deep Learning

模式识别与深度学习

实验2：PCA降维与分类实验

天津大学 智能与计算学部

PCA 原理回顾

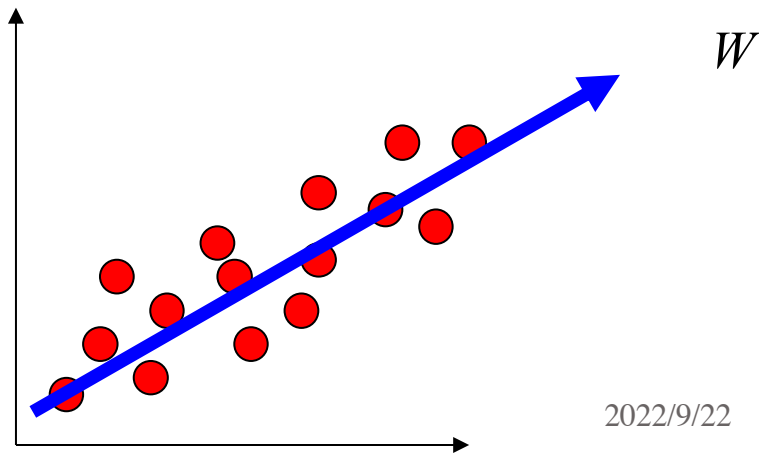
主分量分析(又称为K-L变换)的思想是**通过变化最大化样本的方差**，其可转化为特征问题求解：

$$\Sigma W = \lambda W$$

其中 Σ 为 $X - \bar{X}$ 协方差矩阵, \bar{X} 输入数据的均值。

$$\Sigma = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$$



PCA 原理回顾

- PCA算法

- 计算均值

- 计算协方差矩阵

- 计算特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ 和特征向量 $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d\}$

- 计算低维表示 $\mathbf{y} = \mathbf{U}^T \mathbf{x}$ $\mathbf{U} = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_d \\ | & | & & | \end{bmatrix}$

PCA 实验指导-人脸降维重建

□ 实验目的



灰度化



PCA 降维模型 主成分向量



PCA重建模型



PCA 实验指导-人脸降维重建

□ 数据介绍

lfw 人脸数据集：

<http://vis-www.cs.umass.edu/lfw/lfw.tgz>

5749 people

13233 images

1680 people with two or more images



Aaron_Eckhart_0001.jpg



Aaron_Guiel_0001.jpg



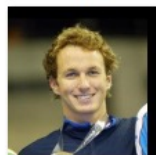
Aaron_Patterson_0001.jpg



Aaron_Peirsol_0001.jpg



Aaron_Peirsol_0002.jpg



Aaron_Peirsol_0003.jpg



Aaron_Peirsol_0004.jpg



Aaron_Pena_0001.jpg



Aaron_Sorkin_0001.jpg



Aaron_Sorkin_0002.jpg



Abba_Eban_0001.jpg



Abbas_Kiarostami_0001.jpg

PCA 实验指导-人脸降维重建

□ 调用包

```
from sklearn.decomposition import PCA  
import numpy as np  
from PIL import Image  
import os
```

pip install scikit-learn



pip install pillow

PCA 实验指导-人脸降维重建

数据读取



一级目录
path

二级目录
face_path

三级目录
Image_path

PCA 实验指导-人脸降维重建

□ 数据读取

```
def dataload(path):  
  
    all_image = [] # list, 存储所有图像  
    re_path = [] # list, 存储图像路径  
  
    for face_path_name in os.listdir(path):  
        face_path = os.path.join(path, face_path_name)  
        for image_path_name in os.listdir(face_path):  
            image_path = os.path.join(face_path, image_path_name)  
            re_path.append(os.path.join(path.split('_')[0]+'_recon', face_path_name, image_path_name))  
            if not os.path.exists(os.path.join(path.split('_')[0]+'_recon', face_path_name)):  
                os.makedirs(os.path.join(path.split('_')[0]+'_recon', face_path_name))  
            img_gray = Image.open(image_path).convert('L')  
            img_np = np.array(img_gray)  
            all_image.append(img_np)  
  
    all_image = np.array(all_image) # (98, 250, 250)  
    all_image_flatten = all_image.reshape((all_image.shape[0], -1)) # (98, 62500)  
    return all_image_flatten, re_path  
  
img_gray.save(os.path.join(os.path.join(path.split('_')[0]+'_recon', face_path_name, image_path_name)))
```


PCA 实验指导-人脸降维重建

□ PCA 模型搭建

```
model = PCA(n_components=components_value)
```

可调节参数：**n_components**

PCA 实验指导-人脸降维重建

1. `n_components`: int, float, None or str

意义 :代表返回的主成分的个数,也就是你想把数据降到几维

`n_components=2` 代表返回前2个主成分

`0 < n_components < 1`代表满足最低的主成分方差累计贡献率

`n_components=0.98`, 指返回满足主成分方差累计贡献率达到98%的主成分

`n_components=None`, 返回所有主成分

`n_components='mle'`, 将自动选取主成分个数n, 使得满足所要求的方差百分比

PCA 实验指导-人脸降维重建

□ PCA 降维

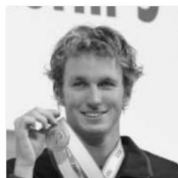
```
components = model.fit_transform(all_image_flatten)
```

□ PCA 重建样本以及重建样本保存

```
face_recon = model.inverse_transform(components)  
for i in range(face_recon.shape[0]):  
    each_face_recon = face_recon[i]  
    each_face_recon = Image.fromarray(each_face_recon.reshape((250,250)))  
    each_face_recon = each_face_recon.convert('L')  
    each_face_recon.save(re_path[i].split('.jpg')[0]+'-recon-'+str(components_value)+'.jpg')
```

PCA 实验指导-人脸降维重建

重建展示



Aaron_Peirsol_0001-gray.jpg



Aaron_Peirsol_0001-recon-1.jpg



Aaron_Peirsol_0001-recon-5.jpg



Aaron_Peirsol_0001-recon-10.jpg



Aaron_Peirsol_0001-recon-30.jpg



Aaron_Peirsol_0001-recon-50.jpg



Aaron_Peirsol_0001-recon-70.jpg



Aaron_Peirsol_0001-recon-100.jpg



Abdel_Nasser_Assidi_0001-gray.jpg



Abdel_Nasser_Assidi_0001-recon-1.jpg



Abdel_Nasser_Assidi_0001-recon-5.jpg



Abdel_Nasser_Assidi_0001-recon-10.jpg



Abdel_Nasser_Assidi_0001-recon-30.jpg



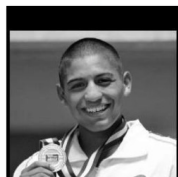
Abdel_Nasser_Assidi_0001-recon-50.jpg



Abdel_Nasser_Assidi_0001-recon-70.jpg



Abdel_Nasser_Assidi_0001-recon-100.jpg



Abner_Martinez_0001-gray.jpg



Abner_Martinez_0001-recon-1.jpg



Abner_Martinez_0001-recon-5.jpg



Abner_Martinez_0001-recon-10.jpg



Abner_Martinez_0001-recon-30.jpg



Abner_Martinez_0001-recon-50.jpg



Abner_Martinez_0001-recon-70.jpg



Abner_Martinez_0001-recon-100.jpg



Ai_Sugiyama_0001-gray.jpg



Ai_Sugiyama_0001-recon-1.jpg



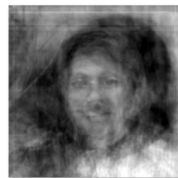
Ai_Sugiyama_0001-recon-5.jpg



Ai_Sugiyama_0001-recon-10.jpg



Ai_Sugiyama_0001-recon-30.jpg



Ai_Sugiyama_0001-recon-50.jpg



Ai_Sugiyama_0001-recon-70.jpg



Ai_Sugiyama_0001-recon-100.jpg

PCA+SVM 实验作业

□ 实验目的



PCA 模型 主成分向量 SVM 模型



51个类别

0	lihaoyu
1	chenjun
2	duchunhong
...	...

PCA+SVM 实验作业

□ 数据介绍

- 原始数据 (51) -> rgb图转灰度图->resize (250,250)
- 每位同学的人脸数据进行了20倍数据增广。包括随机旋转，上下左右翻转，加噪声，随机剪裁
- 划分了训练集和测试集，每位同学对应的训练集包含14个样本，测试集包含7个样本

PCA+SVM 实验作业

数据展示

train



lihaoyu.jpg-1.jpg



lihaoyu.jpg-2.jpg



lihaoyu.jpg-3.jpg



lihaoyu.jpg-4.jpg



lihaoyu.jpg-5.jpg



lihaoyu.jpg-6.jpg



lihaoyu.jpg-7.jpg



lihaoyu.jpg-8.jpg



lihaoyu.jpg-9.jpg



lihaoyu.jpg-10.jpg



lihaoyu.jpg-11.jpg



lihaoyu.jpg-12.jpg



lihaoyu.jpg-13.jpg



lihaoyu.jpg-14.jpg

test



lihaoyu.jpg-15.jpg



lihaoyu.jpg-16.jpg



lihaoyu.jpg-17.jpg



lihaoyu.jpg-18.jpg



lihaoyu.jpg-19.jpg



lihaoyu.jpg-20.jpg

PCA+SVM 实验要求

□ 数据获取

PCA 降维重建实验相关文件（实验数据以及实验代码）以及 PCA降维SVM分类实验相关文件（实验数据以及待补全实验代码）已上传至智慧树作业《实验2-PCA降维+SVM分类》中的 pca_recon以及pca_svm文件夹中

□ 结果分析

调节PCA以及SVM参数，得到分类准确率，结果采用表格表示

PCA	SVM			results	
n_components	C	kernel	decision_function_shape	Train acc	Test acc
			

PCA+SVM 实验要求

□ 实验要求

- 实验时间：09.22-10.07，不得晚于10月07号23.00点
- 提交文件：代码和实验报告，打包为“**实验2-学号-姓名.zip**”，其中实验报告的命名是“**实验2-实验报告-学号-姓名.docx**”（实验报告模板以及实验指导书已上传至智慧树作业《实验2-PCA降维+SVM分类》中）
- 提交方式：zip文件通过智慧树作业页面进行提交