主成分分析（PCA）及其在人脸压缩应用

一、特征值分解(实对称矩阵) v.s. 奇异值分解

定理3：

$A\in R^{n\times n}$ 为实对称矩阵，则存在**正交矩阵Q**：

$$Q^T AQ=Q^{-1}AQ=\begin{pmatrix}\lambda_1 & 0 & 0 & 0\\ 0 & \lambda_2 & 0 & 0\\ 0 & 0 & \ddots & 0\\ 0 & 0 & 0 & \lambda_n\end{pmatrix}=D$$

其中：$\lambda_1,\lambda_2,\cdots,\lambda_n$ 为矩阵A的特征值。

实对称矩阵的特征值分解：

$$A=Q\begin{pmatrix}\lambda_1 & 0 & 0 & 0\\ 0 & \lambda_2 & 0 & 0\\ 0 & 0 & \ddots & 0\\ 0 & 0 & 0 & \lambda_n\end{pmatrix}Q^T=QDQ^T$$

如果A是非方阵

奇异值分解定义：

$X\in R^{m\times n}$为任意实矩阵，则存在**正交矩阵**$U=[u_1,u_2,\cdots,u_m]\in R^{m\times m}$

和**正交矩阵** $V=[v_1,v_2,\cdots,v_n]\in R^{n\times n}$

$$X=U\Sigma V^T$$

$$\underset{m\times n}{\boxed{X}}=\underset{m\times m}{\boxed{U}}\ \underset{m\times n}{\boxed{\Sigma}}\ \underset{n\times n}{\boxed{V^T}}$$

成立。

其中：

$$\Sigma=\begin{pmatrix}\sigma_1 & 0 & 0 & 0 & 0 & 0\\ 0 & \sigma_2 & 0 & 0 & 0 & 0\\ 0 & 0 & \ddots & 0 & 0 & 0\\ 0 & 0 & 0 & \sigma_m & \cdots & 0\end{pmatrix}\in R^{m\times n},\quad \sigma_i\ge 0$$

对实对称矩阵 XX^T和 X^TX 实行特征值分解：

$$XX^T=U\Sigma V^T(U\Sigma V^T)^T=U\Sigma\Sigma^T U^T$$

- U为对称矩阵XX^T的特征值分解的正交阵

$$X^TX=(U\Sigma V^T)^T U\Sigma V^T=V\Sigma^T\Sigma V^T$$

- V为对称矩阵 X^TX的特征值分解的正交阵
- 奇异值：$\sigma_i=\sqrt{\lambda_i}$，$\lambda_i$ 为对称阵的特征值

主成分分析（PCA）及其在人脸压缩应用 （算法流程图）

- PCA是一种经典的数据降维方法（为什么要降维？）
- 能将数据如 $\mathbf{x} \in \boldsymbol{R}^{1\mathrm{x}N}$ 降到 $\boldsymbol{z} \in \boldsymbol{R}^{1\mathrm{x}n}$, 其中 $\boldsymbol{n} < \boldsymbol{N}$

举例1： 如何将 m 个 $\mathbf{x}_i \in \boldsymbol{R}^{1\mathrm{x}N}$ $(X \in R^{m\times N})$降维(压缩)至 $\mathbf{z}_i \in \boldsymbol{R}^{1\mathrm{x}n}$ $(Z \in R^{m\times n})$？

Step 1: 数据预处理（mean normalization, 为什么？）

a) $\mu_i = mean(x_i), s_i = std(x_i)$

b) $x_i = \dfrac{x_i - \mu_i}{s_i}$

Step 2: 计算协方差矩阵

$$\Sigma = \frac{1}{m} X^T X, \quad \text{where } X \in R^{m\times N}$$

Step 3: 计算协方差矩阵的奇异值分解（SVD）

$$\Sigma = USV^T$$

(Matlab命令： `[U,S,V] = svd(∑);`)

Step 4: 提取U的前n列向量构造降维矩阵（n为你想降到的维数）

```
Ureduce = U(:, 1:n)
```

Step 5: 最终得到降维(压缩)数据Z

```
Z = X*Ureduce;
```

举例2： 如何将 m 个降维(压缩)数据$\mathbf{z}_i \in \boldsymbol{R}^{1\mathrm{x}n}$ $(Z \in R^{m\times n})$ 恢复(解压)出原来数据 $\mathbf{x}_i \in \boldsymbol{R}^{1\mathrm{x}N}$ $(X \in R^{m\times N})$？

Step 6: `Ureduce = U(:, 1:n)`进行转置， 并进行如下操作即可：

$X_{recover}$ = `Z*Ureduce`^{T};

详见Project2_PCA_Face_ToStudent文件下人脸压缩与恢复代码样例：

Demo_PCA_Face_ToStudent.m

特征值&奇异值：https://blog.csdn.net/program_developer/article/details/80632779

二、课题要求 **请使用提供的软件包：Project2_PCA_Face_ToStudent**

- ✓ 实现提供代码包里 Demo_PCA_Face_ToStudent.m中的三个子函数(单独)，即

  (1) Line 39: [U, S] = pca(X_norm); [即实现算法流程图中的Step 2 + Step 3]

  (2) Line 54: Z = projectData(X_norm, U, K); [即实现算法流程图中的Step 4 + Step 5]

  (3) Line 71: X_rec = recoverData(Z, U, K); [即实现算法流程图中的Step 6]

- ✓ 请用： 设置断点+调试方式，将代码所有步骤过一遍！

- ✓ 检验代码是否正确？（如下图为正确）