

Issued: May. 5, 2018
Due: May. 23, 2018

Assignment II-part2

Policy

Group study is encouraged; **however, assignment that you hand-in must be of your own work. Anyone suspected of copying others will be penalized.** The homework will take considerable amount of time so start as soon as possible..

1. **PCA (Matlab Programming Assignment):** This problem involves implementing a PCA (Principal Component Analysis) based algorithm to represent a human face in terms of principal components, and based on this representation, identify the face in a given image as one of identity in the training set. A dataset comprising of 213 human-face images of 10 individuals is provided of which 163 images are used for training and 50 are used for testing. A balanced training data is used: there are equal number of images for each identity. Each face image portrays an individual with different expression.
 - (i) Perform PCA on the 163 training images with different number of principal components (PC) then display the top 5 principal components. Also, reconstruct each image using 5, 50, 200 and 500 principal components. Compare the average reconstruction mean square error versus the number of PCs.
 - (ii) Use PCA representation to identify the 50 test face images. Write a Matlab code to identify input test images. Use the Euclidean distance as a measure of closeness.

Algorithm 1: PCA algorithm

- 1 Construct data matrix \mathbf{X}
 - 2 Subtract mean face $\bar{\mathbf{x}}$ from each image
 - 3 Construct Covariance Sample matrix Σ
 - 4 Find eigenvectors and eigenvalues of Σ
 - 5 Find principal Components, which are k eigenvectors with largest eigenvalues.
-

For implementing the PCA algorithm, a skeleton code is provided in the “PCA” folder. The main function `PCA_main.m` calls the following 5 functions:

- (1) `[train_matrix,test_matrix] = createDataset(),`
 - (2) `[project_train_img,k_eig_vec,m] = train_PCA(train_matrix,k),`
 - (3) `[recon_error] = train_recon(train_matrix,project_train_img,k_eig_vec,m),`
 - (4) `[project_test_img] = test_PCA(test_matrix,k_eig_vec,m),`
 - (5) `[id] = identify(project_train_img,project_test_img)`
- (i) `[train_matrix,test_matrix] = createDataset()` constructs both the training and test matrix in form of $N \times d$ where N is the number of images, and d is the size of vector image. training and test image sets are contained respectively in ‘training_img’ and ‘test_img’ folder.

- (ii) `[project_train_img,k_eig_vec,m] = train_PCA(train_matrix,k)` takes training matrix `train_matrix` and parameter `k`, which is the number of PCs. This function outputs 3 values, `project_train_img`, `k_eig_vec`, `m` and display the 3 largest eigen faces, which interpret the eigen vectors as image.
`project_train_img` is the training matrix represented by k PCs. `k_eig_vec` is k PCs and `m` is mean.
- (iii) `[recon_error] = train_recon(train_matrix,project_train_img,k_eig_vec,m)` takes 4 inputs: training matrix `train_matrix`, projected training matrix `project_train_img`, k principal components `k_eig_vec` and mean `m`. This function saves the reconstructed training images from the training matrix represented by k eigenvectors and outputs the average reconstruction mean square error.
- (iv) `[project_test_img] = test_PCA(test_matrix,k_eig_vec,m)` takes test matrix `test_matrix`, k principal components `k_eig_vec` and mean `m` and outputs `project_test_img`, the test matrix represented by k PCs from training matrix.
- (v) `[id] = identify(project_train_img,project_test_img)` takes training matrix and test matrix both represented by k PCs as inputs and outputs the index(e.g. 1,2,...) of training images which is most similar to the test images.

Submit Instructions for Programming Assignment

- Please submit in .zip file to KLMS named `ee488_assignment3_part2_student#.zip`, for example,
`"ee488_assignment3_part2.20181234.zip"`.
- This file should contain the following folder - **PCA** and each folder contains document file for the result with analysis.
- In matlab code, the comment explaining your code **must be** included, or you will not get a full grade even if your code works fine. Please also include all the files that are required to run the code in the zip file. Do not change the name of the folder and comments should be written in English. Additionally submitting unexecutable code will receive no points.