Machine Learning Homework 7

Due Date 23:55 4th Jan

I. Kernel Eigenfaces

In this section, you are going to do face recognition using eigenface and fisherface.
Reference: https://www.csie.ntu.edu.tw/~mhyang/papers/fg02.pdf

- Data
  - The **Yale Face Database.zip** contains 165 images of 15 subjects (subject01, subject02, etc.). There are 11 images per subject, one for each of the following facial expressions or configurations: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink.
  - These data are separated into training dataset(135 images) and testing dataset(30 images). You can resize the images for easier implementation.
- What you are going to do
  - Part1: Use PCA and LDA to show the first 25 eigenfaces and fisherfaces, and randomly pick 10 images to show their reconstruction. (please refer to the lecture slides).
  - Part2: Use PCA and LDA to do face recognition, and compute the performance. You should use **k nearest neighbor** to classify which subject the testing image belongs to.
  - Part3: Use kernel PCA and kernel LDA to do face recognition, and compute the performance. (You can choose whatever kernel you want, but you should try **at least two different kernels** in your implementation.) Then compare the difference between simple LDA/PCA and kernel LDA/PCA, and the difference between different kernels.

II. t-SNE

Here are nice implementations of t-SNE in different programming languages: https://lvdmaaten.github.io/tsne/

- Data & reference code
  - Download link: https://lvdmaaten.github.io/tsne/code/tsne_python.zip,
  - **mnist2500_X.txt**: contains 2500 feature vectors with length 784, for describing 2500 mnist images.
  - **mnist2500_labels.txt**: provides corresponding labels
  - **tsne.py:** reference code
- What you are going to do
  - Part1: Try to modify the code a little bit and make it back to symmetric SNE. You need to first understand how to implement t-SNE and find out the specific code piece to modify. You have to explain the difference between symmetric SNE and t-SNE in the report.
  - Part2: Visualize the embedding of **both t-SNE and symmetric SNE**. Details of the visualization:
    - Project all your data onto 2D space and mark the data points into different colors respectively. The color of the data points depends on the label.
    - Use videos or GIF images to show the optimized procedure.
  - Part3: Visualize the distribution of pairwise similarities in both high-dimensional space and low-dimensional space, based on **both t-SNE and symmetric SNE**.
  - Part4: Try to play with different perplexity values. Observe the change in visualization and explain it in the report.

III. Report

- Submit a report in pdf format. The report should be written in English.
- Please strictly follow the report format. We will deduct some points according to the situation if you don't follow it.
- Since this homework is mainly graded by report, please spend more time on it. (e.g. well explained & organized) We won't give you any point if you just finish the code.
- Please don't explain the code line by line. You need to explain it clearly and well structured. For example, explain which part you have done in the function and how.
- Report format:
    - a. code with detailed explanations (40%)
        - Paste the screenshot of your functions with comments and explain your code. Explain the process to clustering and show different kernels.
        - Note that if you don't explain your code clearly, you cannot get any points in section b and c either.
            - Kernel Eigenfaces
                - Part1 (10%) Also, simply explain how you do PCA & LDA (what is the step of it?)
                - Part2 (5%)
                - Part3 (10%) Also, simply explain how you do Kernel PCA & kernel LDA (what is the step of it?)
            - t-SNE
                - Part1 (10%) (Also, show the formula of t-SNE & symmetric SNE)
                - Part2 (2%)
                - Part3 (2%)
                - Part4 (1%)
    - b. experiments settings and results (33%) & discussion (17%)
        - Show everything we asked you to show
            - Kernel Eigenfaces
                - Part1 (6%)
                - Part2 (4%)
                - Part3 (4%) & (6%) Please discuss the observation in this part (You can compare the result with PCA/LDA)
            - t-SNE

        - Part1 (5%) & (5%) Please discuss the observation in this part (You can compare the performance between t-SNE and symmetric SNE)
        - Part2 (5%)
        - Part3 (4%)
        - Part4 (5%) & (6%) Please discuss the observation in this part
    - c. observations and discussion (10%)
        - Try to explain the meaning of eigenfaces (4%)
        - Explain what is the crowded problem of symmetric SNE (3%)
        - Anything you want to discuss (3%)

IV. Turn in

1. Report (.pdf)
2. Source code
3. Videos or GIF images of optimize procedure

You should zip source code and report in one file and name it like ML_HW7_yourstudentID_name.zip, e.g. ML_HW7_0856XXX_王小明.zip.

P.S. If the zip file name has format error or the report is not in pdf format, there will be a penalty (-10). Please submit your homework before the deadline, late submission is not allowed.

Note that if you miss any one of the requirements (report, or source code), you cannot get any score!

♦ Packages allowed in this assignment:

You are only allowed to use numpy, scipy.spatial.distance, and I/O related functions (like cv2.imread(), csv, matplotlib etc.). Official introductions can be found online.
Important: scikit-learn and SciPy are not allowed.