Machine Learning Homework 7

Dimension Reductions

Due day 23:59PM 10th June

I. Different ways to do dimension reductions (60% in total):
In this assignment, you need to use different ways to do dimension reduction on **mnist_X.csv** and **mnist_label.csv**, including PCA, LDA, S-SNE and T-SNE.

A. PCA (15%)
Use PCA to project all your data mnist_X.csv onto 2D space and mark the data points into different colors. The color of the data points depends on which cluster it belongs to (mnist_label.csv).

B. LDA (15%)
Use LDA to project all your data mnist_X.csv onto 2D space and mark the data points into different colors. The color of the data points depends on which cluster it belongs to (mnist_label.csv).

C. Symmetric SNE and T-SNE (30%)
Use T-SNE and symmetric SNE to project all your data mnist_X.csv onto 2D space and mark the data points into different colors respectively. The color of the data points depends on which cluster it belongs to (mnist_label.csv). The T-SNE code is provided. You need to modify it to symmetric SNE and discuss their differences. You also have to visualize the distribution of pairwise similarities in both high-dimensional space and low-dimensional space, based on both T-SNE and symmetric SNE.

II. Eigenface (20% in total):
att_faces dataset contains in total 400 images (40 distinct subjects, 10 images per subject). You should use PCA to show the first 25 eigenfaces, and randomly pick 10 images to show their reconstruction (please refer to the lecture slides p.125).
download link:
http://www.cl.cam.ac.uk/Research/DTG/attarchive:pub/data/att_faces.zip

III. Report (20% in total):

Submit a report in **pdf** format for showing your **code with detailed explanations and results**, giving **detailed discussion** on experiments as well as your observations. The report should be written in **English**.

You should zip source code and report in one file and name it like ML_HW7_yourstudentID_name.zip, e.g. ML_HW7_0756005_鄭家期.zip.

If the zip file name has format error, it will be given penalty (-10).

IV. Packages allowed in this assignment:

You are only allowed to use numpy and I/O related functions (like cv2.imread(), csv, matplotlib etc.).

Important: scikit-learn is not allowed.