Assignment 2 (15%): Dimensionality Reduction. (Due on 16:59:59, Oct. 19, 2016)

Please implement three dimensionality reduction methods PCA, SVD as well as ISOMAP, and then perform 1-Nearest Neighbor (1-NN) in the low-dimensional data space.

Dataset

1. Data: We adopt two binary datasets which come from here. Each dataset contains both training set and testing set. (Please use editors like Sublime or Notepad++ to open them.)
   a. Dataset 1 (Training) Download, Dataset 1 (Testing) Download
   b. Dataset 2 (Training) Download, Dataset 2 (Testing) Download
2. Data Description and Format: In each binary training and testing dataset, each row represents an example with label. Namely, the last column represents the label of the corresponding example, and the remaining columns represent the feature of the corresponding example.

Task Description

1. Please implement PCA for dimensionality reduction in Section 2.4.3.1 of the textbook.
   a. First, **only use the training data** to learn the projection matrix.
   b. Second, project the training data and the testing data to the k-dimensional space via the projection matrix learned in the first step, where k=10, 20 and 30.
   c. Third, binary classification in the low-dimensional space: perform 1-Nearest Neighbor (1-NN) **in the k-dimensional space** to predict the label of each testing example, where k=10, 20 and 30, and report the accuracy under different k. The true label in testing set is only used to compute the accuracy.
2. Please implement SVD for dimensionality reduction in Section 2.4.3.2 of the textbook.
   a. First, **only use the training data** to learn the projection matrix.
   b. Second, project the training data and the testing data to the k-dimensional space via the projection matrix learned in the first step, where k=10, 20 and 30.
   c. Third, binary classification in the low-dimensional space: perform 1-NN **in the k-dimensional space** to predict the label of each testing example, where k=10, 20 and 30, and report the accuracy under different k. The true label in testing set is only used to compute the accuracy.
3. Please implement ISOMAP for dimensionality reduction in Section 3.2.1.7 of the textbook. To implement ISOMAP, you should also read Section 2.4.4.2 and implement MDS which is a step of ISOMAP.
   a. First, **use the training data and the testing data** to learn the projection. During this step, please use 4-NN to construct a weighted graph.
   b. Second, project the training data and the testing data to the k-dimensional space via the projection learned in the first step, where k=10, 20 and 30.
   c. Third, binary classification in the low-dimensional space: perform 1-NN **in the k-dimensional space** to predict the label of each testing example, where k=10, 20 and 30, and report the accuracy under different k. The true label in testing set is only used to compute the accuracy.

You can use the existing tools/packages/functions/codes directly to compute eigenvector/eigenvalue of a matrix, conduct SVD of a matrix, and compute the shortest path

between two nodes in a graph (e.g., the Dijkstra algorithm). However, other steps in PCA/SVD/ISOMAP and 1-NN must be implemented by yourself.

The report should contain the pseudo-codes of three methods and the final results. The final results presented in the report**should be organized as a table** to show the accuracy of each method under different k on each dataset.

Reminders about Submission

1. Submission Deadline: **2016-10-19 16:59:59**.
2. Before submitting your assignment, please read **Submission Requirement and Description** section above carefully and obey it.
3. For assignment 2, please pack your **report**, **code** and **ReadMe.txt** into a zip file named with your student ID, e.g., MG1633001.zip.