Machine Learning (CS 6140)
Homework 3

Instructor: Ehsan Elhamifar

Due Date: November 27, 2017, 11:45am

1. PCA Objective Value. Show that for the optimal solution of fitting a d-dimensional subspace to data using PCA, the objective function value is equal to $\sum_{i \geq d+1} \sigma_i^2$. Here, σ_i denotes the i-th largest singular value of the mean subtracted data matrix, $\bar{\boldsymbol{Y}} = \begin{bmatrix} \boldsymbol{y}_1 - \bar{\boldsymbol{y}} & \cdots & \boldsymbol{y}_N - \bar{\boldsymbol{y}} \end{bmatrix}$, where $\bar{\boldsymbol{y}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{y}_i$.

2. Locally Linear Embedding. As discussed in the class, the first step of LLE involves solving the optimization

$$\min_{\boldsymbol{w}_i} \frac{1}{2} \|\boldsymbol{Y}_i \boldsymbol{w}_i\|_2^2 \ \ \text{s.t.} \ \ \boldsymbol{1}^\top \boldsymbol{w}_i = 1,$$

where $\boldsymbol{Y}_i = \begin{bmatrix} \cdots \boldsymbol{y}_j - \boldsymbol{y}_i \cdots \end{bmatrix}$ for all j's that are neighbors of point j and $\boldsymbol{1}$ is a vector of all 1's of appropriate dimension. a) Prove that the solution of this optimization is given by $\boldsymbol{w}_i = \frac{(\boldsymbol{Y}_i^\top \boldsymbol{Y}_i)^{-1} \boldsymbol{1}}{\boldsymbol{1}^\top (\boldsymbol{Y}_i^\top \boldsymbol{Y}_i)^{-1} \boldsymbol{1}}$. b) Assume that we have N data points in the D-dimensional ambient space. We set the number of nearest neighbors in LLE to be K. We try to find a d-dimensional representation of the data. What is the computational cost of step 1 and step 2 of LLE, in terms of big O notation (for example, $O(D^p N^q)$)? Explain your answer in details.

3. Laplacian Matrix. The Laplacian matrix of a graph on N nodes is defined by $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$, where $\boldsymbol{W}$ is the matrix of weights of the graph, where W_{ij} is the weight between nodes i and j. The degree matrix $\boldsymbol{D}$ is also defined as a diagonal matrix whose i-th diagonal entry is the sum of the weights of the nodes connected to node i, i.e., $D_{ii} = \sum_{j=1}^{N} W_{ij}$. a) Let $\boldsymbol{x} = \begin{bmatrix} x_1 & \cdots & x_N \end{bmatrix}^\top$. Compute a closed-form expression for $\boldsymbol{x}^\top \boldsymbol{L} \boldsymbol{x}$ that only depends on W_{ij}'s. b) Show that the Laplacian matrix is a positive semi-definite matrix. c) Is L an invertible matrix? Prove. d) How many zero singular values does $\boldsymbol{L}$ have?

4. PCA Implementation. Implement the PCA and the Kernel PCA algorithm. The input to the PCA algorithm must be the input data matrix $Y \in \mathbb{R}^{D \times N}$, where D is the ambient dimension and N is the number of points, as well as the dimension of the low-dimensional representation, d. The output of the PCA must be the basis of the low-dimensional subspace $U \in \mathbb{R}^{D \times d}$, the mean of the subspace $\mu \in \mathbb{R}^D$ and the low-dimensional representations $X \in \mathbb{R}^{d \times N}$. For KPCA, the input must be the Kernel matrix $K \in \mathbb{R}^{N \times N}$, where N is the number of points, as well as the dimension of the low-dimensional representation, d. The output of KPCA must be the low-dimensional representations $X \in \mathbb{R}^{d \times N}$. Apply the PCA algorithm for $d = 2$ to the provided data and plot the 2-dimensional representations of data points. Apply KPCA using Gaussian Kernel ($k_{ij} = \boldsymbol{e}^{-\|\boldsymbol{y}_i - \boldsymbol{y}_j\|_2^2/(2\sigma^2)}$) for several values of σ to the data and plot the 2-dimensional representations of data points.

5. Kmeans Implementation. Implement the Kmeans algorithm. The input to the algorithm must be the data matrix $X \in \mathbb{R}^{D\times N}$, the desired number of clusters, k, as well as the number of repetitions of kmeans with different random initializations, r. The output of the algorithm must be the clustering of the data (indices of points in each group) for the best run of the kmeans, i.e., the run of kmeans that achieves the lowest cost function among all different initializations. Apply the Kmenas algorithm for several values of k to the provided data. Plot the original data points in one graph and the clustering results in another graph, by indicating points in the same group by a different color.

6. Spectral Clustering Implementation. Implement the Spectral Clustering algorithm with Gaussian kernel similarities, i.e., $s_{ij} = \exp(\|x_i - x_j\|_2^2/\sigma)$, where $\sigma > 0$ is a positive parameter. The input to the algorithm must be the data matrix $X \in \mathbb{R}^{D\times N}$, the desired number of clusters, k, as well as σ. The output of the algorithm must be the clustering of the data (indices of points in each group). Apply the spectral clustering algorithm for $k = 4$ and several values of σ to the provided data. Plot the original data points in one graph and the clustering results in another graph, by indicating points in the same group by a different color.

Homework Submission Instructions:

– Submission of Written Part: You must submit your written report in the class BEFORE CLASS STARTS. The written part, must contain the plots and results of running your code on the provided datasets.

–Submission of Code: You must submit your Python code (.py file) via email, BEFORE CLASS STARTS. For submitting your code, please send an email to me and CC both TAs.

- The title of your email must be "CS6140: Code: HW3: Your First and Last Name".
- You must attach a single zip file to your email that contains all python codes and a readme file on how to run your files.
- The name of the zip file must be "HW3-Code: Your First and Last Name".