Machine Learning (CS 6140)
Homework 3

Instructor: Ehsan Elhamifar

Due Date: December 8, 2016, 11:45am

1. PCA Implementation. Implement the PCA algorithm. The input to the algorithm must be the input data matrix $X \in \mathbb{R}^{D \times N}$, 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 algorithm must be the base 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 $Y \in \mathbb{R}^{d \times N}$. Apply the PCA algorithm for $d = 2$ to the provided data and plot the 2-dimensional data points in $\mathbb{R}^2$.

2. 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 $k = 4$ 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.

3. 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 $\sigma \in \{0.001, 0.01, 0.1, 1\}$ 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.