HW IV

EE 546: Mathematics of High-Dimensional Data

University of Southern California

Assigned on: November 8, 2020 Due date: 11:59PM November 23, 2020

The purpose of this homework is for you to gain some experience with implementing the spectral clustering algorithm.

1- **Spectral Clustering** In this problem we are going to run the spectral clustering algorithm of the paper On Spectral Clustering: Analysis and an algorithm by NG, Jordan, and Weiss on motion capture data. We use the Carnegie Mellon Motion Capture dataset (available at www.mocap.cs.cmu.edu/. For your convenience, I have prepared the data for trials 2 and 5 of this dataset for you (available under the names aca2.mat and aca5.mat).

(i) Minor data preparation. Load aca2.mat and aca5.mat. The matrix $\boldsymbol{X}$ contains the data points as columns. The vector s contains the true class of each data point. To help reduce the size of the data pick every other column of $\boldsymbol{X}$ and s. That is run your algorithm using X=X(:,1:2:end) and s=s(1:2:end). As usual, normalize the columns of $\boldsymbol{X}$ so that all columns have unit Euclidean norm.

(ii) Build the following kernel

$$\boldsymbol{K}_{ij} := k(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\gamma \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_{\ell_2}^2}.$$

Use this as the weight matrix but only pick the top k entries in each column of the matrix $\boldsymbol{K}$. The weight matrix $\boldsymbol{W}$ picked in this way is not symmetric so symmetrize it by using $\boldsymbol{W} = \frac{\boldsymbol{W}+\boldsymbol{W}^T}{2}$.

(iii) Run the spectral clustering algorithm (the variation proposed in the paper On Spectral Clustering: Analysis and an algorithm by NG, Jordan, and Weiss) on these two datasets using the weight matrix $\boldsymbol{W}$ as defined above. Use the following values for $\gamma = 0.1, 0.2, \ldots, 0.9, 1, 2, \ldots, 100$ and $k = 2, 3, 4, \ldots, 50$. Record the minimum misclassification error for all of these parameters. Misclassification error is the average number of point misclassified. I have provided this function for you in a file Misclassification.m. (You would also need the function missclassGroups.m to run this which also provided).