

# HW IV

EE 546: Mathematics of High-Dimensional Data

University of Southern California

Assigned on: November 1, 2018

Due date: beginning of class on November 14, 2018

---

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/](http://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  $\mathbf{X}$  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  $\mathbf{X}$  and  $s$ . That is run your algorithm using  $\mathbf{X}=\mathbf{X}(:,1:2:\text{end})$  and  $s=s(1:2:\text{end})$ . As usual, normalize the columns of  $\mathbf{X}$  so that all columns have unit Euclidean norm.
- (ii) Build the following kernel

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

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

- (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  $\mathbf{W}$  as defined above. Use the following values for  $\gamma = 1, 2, \dots, 100$  and  $k = 2, 3, 4, \dots, 50$ . 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).