Homework 5. Due Thursday, Nov. 19

1. **(5 pts)** Show that the Laplacian eigenmap to $\mathbb{R}^m$ is the solution to the following optimization problem:

$$\min \sum_{i,j} k_{ij} \|y_i - y_j\|_2^2 \quad \text{subject to} \quad Y^\top Q Y = I, \quad Y^\top Q 1_{n\times 1} = 0. \tag{1}$$

   Here, y_i's are columns of Y, and Y is $n \times m$, and the rest of notation as in Section 7.3 of `4-DimReduction.pdf`.

2. **(20 pts) The goal** of this problem is to practice and compare various methods for dimensional reduction.

   - **Methods:**
     (a) PCA;
     (b) Isomap;
     (c) LLE;
     (d) t-SNE;
     (e) Diffusion map.

     Diffusion map should be programmed from scratch. Readily available codes can be used for the rest. For example, the built-in Matlab function can be used for t-SNE; S. Roweis's code can be used for LLE; my code for isomap is in the lecture notes. If you use some standard code, specify its source, read its description, and be ready to adjust parameters in it.

   - **Dataset 1:** `Scurve` generated by `MakeScurveData.m`: 352 data points in 3D forming a uniform grid on the manifold.

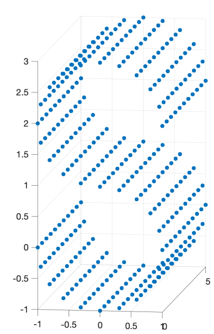

Figure 1: Scurve

   - **Dataset 2:** `Scurve` generated by `MakeScurveData.m` and perturbed by Gaussian noise. Try various intensities, push each method to its limit.

   - **Dataset 3:** "Emoji" dataset generated by `MakeEmojiData.m`: a set of 1024 images each one is 40×40 pixels. Images vary from a smiley face to an angry face and in the degree of blurring. Its subsampled set is shown in Fig. 2. *Note* *that picking a good value of ϵ for the diffusion map might require some effort as the distances between the nearest neighbors are very nonuniform. You should be able to get a nice 2D surface embedded into 3D with a right ϵ. Using $\alpha = 0$ or $\alpha = 1$ is up to you.*

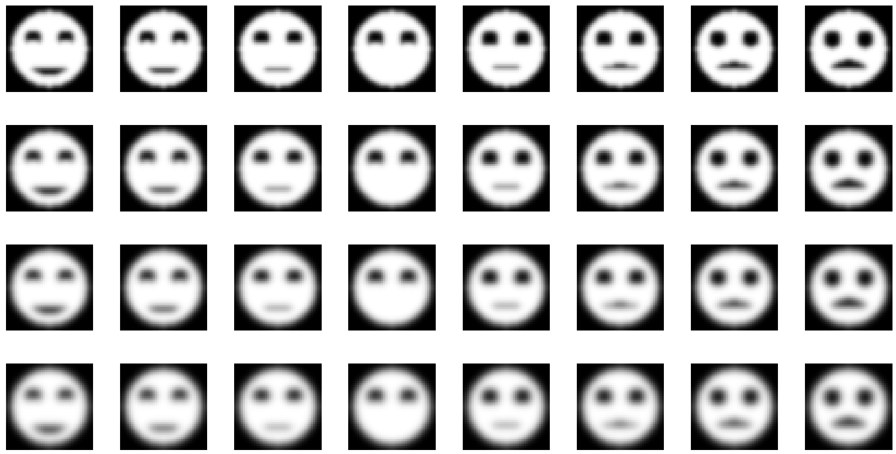

Figure 2: The subsampled "Emoji" dataset.

- **Submit** a report on the performance of these methods on each dataset. Include all necessary figures.