**Problem Set #8**

**28 points + 2 bonus (<3 hours) counts for 7.5 points on your final grade.**

**Due: Monday, 11:59 PM, Mar 24/2025**

**Administrative comments:**

1. **Please turn in your answers in the following way: written parts of your answer together with plotted figures as a single pdf + your code as a single .m file containing code for all the problems, appropriately commented)** **.**
2. **When you name your files, use the following convention:**
   1. **Name the document that contains all your written parts as: ps#\_name.pdf (e.g., ps2\_shreesh.pdf). Make sure it includes the table below.**
   2. **Name the script that contains all your code as: ps#\_name.m (e.g., ps2\_shreesh.m)**
      * **Within this code, separate the code for each problem with a “section break”, which in MATLAB is obtained by inserting a line with %% followed by the problem number. (e.g., %% Problem 2)**
3. **Upload your answer and figures in one PDF to Gradescope Problem Set 8 (PDF), and code in one .m file to Problem Set 8 (code).**
4. **Manifold learning (5 points) (45 min).**

Read through the Seung et al. (2000) piece and write a short paragraph (100-200 words) summarizing the key points in that piece.

1. **Nonlinear dimensionality reduction (8 points) (45 min).**

In addition to LLE, other popular nonlinear dimensionality reduction techniques are **Isomap**, Stochastic Neighbor Embedding (**SNE**), and Multi-Dimensional Scaling (**MDS**). From either the papers below, or Google searches, answer succinctly (in a sentence or three), the following questions about **Isomap**. *The goal of this question is for you to develop a general sense for how this technique is similar or different from PCA and LLE. You don’t need to learn the gory math and proofs, and you don’t have to read all the papers; just develop a conceptual sense for Isomap.*

* 1. [2 points] What is the point of the technique?

*For instance, for LLE, one could say that LLE is a dimensionality reduction technique that attempts to learn the lower-dimensional (nonlinear) manifold on which the given dataset lies.*

* 1. [2 points] What is the key concept of the technique?

*For instance, for LLE: “LLE does so by retaining the local neighborhood structure (relative distances) among points in the original high-dimensional space in such a way that nearby points in the original dataset stay nearby in the lower-dimensional approximation. ”*

* 1. [4 points] What are the key steps of the technique, in words (as best as you can describe them)?

*For instance, “In LLE, each point in the original D-dimensional dataset is first reconstructed as the best possible linear combination of a fixed number of neighbors. Then, using the coefficients of this linear combination (weights), each point is approximated as a point in lower dimensional space while making sure that the relationships between the point and its neighbors in the low-dimensional space match, as closely as possible, the corresponding relationships in the original high-dimensional space. ”*

* 1. **[2 bonus points**] How is Isomap similar / different from LLE?

**Support material**

* [**References for isomap**] Carlotta Slides, Tenenbaum1998, Tenenbaum2000, vanDerMaaten 2009 (review)
* **[For your future reference**] MATLAB toolbox containing functions for these and 29 other dimensionality reduction techniques: <http://lvdmaaten.github.io/drtoolbox/>

1. **LLE (15 points) (1 hr)**. On the given dataset, perform dimensionality reduction (by manifold learning) using LLE. *[Feel free to discuss with others or with us if you have questions.] (For question a-e, please put your answer in .m file. And for question f, please put your answer in the pdf summary file.)*
   1. (1 point) Plot the original data in 3-D space using the following commands (the plot should look like an S-curve).

load **ps7\_LLE\_datasetA**

figure; subplot(3,3,1); scatter3(X(1,:),X(2,:),X(3,:),20,colorMat)

Make sure to say “help scatter3” and see what it is doing.

* 1. (1 point) Apply LLE (using k=20) to reduce the data to a 2-D space (d=2) and plot the resulting lower-dimensional data in the next “subplot”. (*Hints*: (a) use lle.m and (b) use scatter instead of scatter3.)
  2. (2 points) Repeat with k= 4, 6, 8, 12, 16, 50, 200 (and plot the resulting lower dimensional data in each case in successive subplots).
  3. (2 points) Compare the results across k and comment on what the optimal k-value may be (or a good range of k values).
  4. (6 points) (15 min) Repeat a-d for **ps8\_LLE\_datasetB.mat** (this is a more densely sampled version of datasetA).
  5. (3 points) (15 min) In a sentence or two, answer each of the following questions:
  6. Does it make intuitive sense why the results are not good when k is either too low or too high? (Think about neighborhoods in the original, high dimensional space.)
  7. Are the “best” k’s different for the two datasets?
  8. Given a manifold on which data lie, can you describe in words the approximate relationship between density of sampling and “best k” (describe in a sentence or two)?

**(Optional: For your future reference; nothing to submit) Real world, brain-sciences applications***.* Examples of the application of nonlinear dimensionality reduction in the literature.

* Words in semantic space: Roweis 2000 (Science), Fig. 4
* Visual perception: Tanenbaum2000 (Science), Fig. 1
* fMRI: Anderson (NIPS), Qiu 2015, Liu 2013.
* Spiking data: Stopfer 2003 (Neuron), Fig. 5A and Experimental Procedures