GEOMETRIC DATA ANALYSIS: PROBLEM SET 2

1. Problems

For these problems, I would like you to begin by writing your own implementation of IsoMap. Please write the MDS algorithm yourself but you can use a library to compute the graph metric if you wish.

(1) Recall that LLE is comprised of two steps:
 (a) For each data point x_i, find weights w_{ij} such that we minimize the expression $||x_i - \sum_j w_{ij} x_j||$, where j varies over the k-nearest neighbors of x_i.
 (b) Solve for points $y_i \in \mathbb{R}^m$ such that the expression $\sum_i ||y_i - \sum_j w_{ij} y_j||$ is minimized.

 Explain how to solve these minimization problems. Then implement your solution.

(2) Give examples of spaces (defined as the solutions to equations) that are and are not manifolds. Explain.

(3) Show that given any point $m \in M$, where M is a topological manifold, we can choose a chart $\theta\colon U \to \mathbb{R}^n$ that contains m such that $\theta(m) = 0$.

(4) Write down explicit coordinate patches and transition functions for the circle using charts defined by (a) the angle and (b) the projections to the x and y axes.

(5) **(Extra credit)** Show that the tangent plane provides a functor from the category of manifolds with a distinguished point $m \in M$ to the category of vector spaces and linear transformations.

(6) **(Extra credit)** Prove that any smooth connected one-dimensional manifold is diffeomorphic to either the circle or an interval in $\mathbb{R}$.

(7) Classical MDS involves solving a matrix factorization problem of the form $A = U \Lambda U^T$ where Λ is the diagonal matrix of eigenvalues and U has the eigenvectors in the columns. A natural way to consider approximating this is to subsample the points. So imagine that you have chosen $m << n$ points from the data set. Prove that solving the matrix factorization problem for those m points approximates the actual solution. (Recall that you can assume that A is positive semidefinite.) Hint: Use a block decomposition of the matrix.

(8) **(Extra credit)** Let M be a compact Riemannian manifold embedded in $\mathbb{R}^N$. Prove that the IsoMap approximation to the Riemannian metric on M converges as the number of points sampled goes to ∞. You will need to make assumptions about the curvature.

(9) Do a bakeoff of MDS, IsoMap, LLE, and Laplacian eigenmaps using the two data sets from last time plus the additional included data sets. Explore various parameter settings and the subsampling approximations from the problem above. Discuss the results.