Question 2

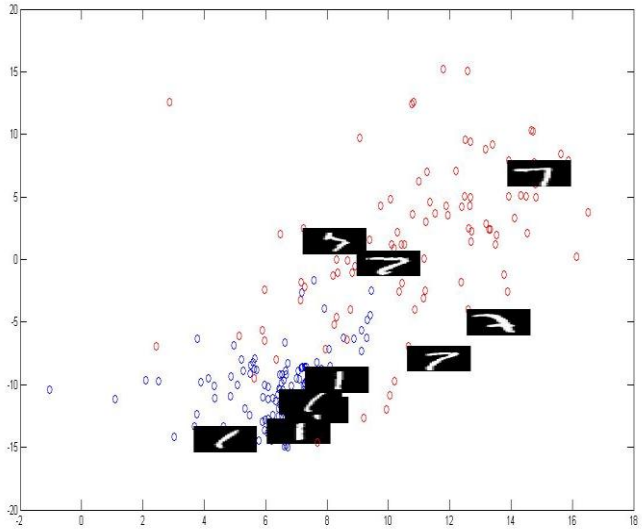
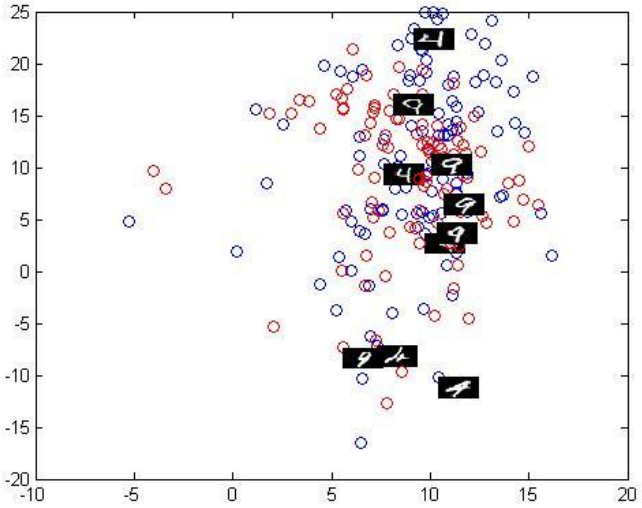
A. MANIFOLD BASED MODELING OF MNIST DIGITS

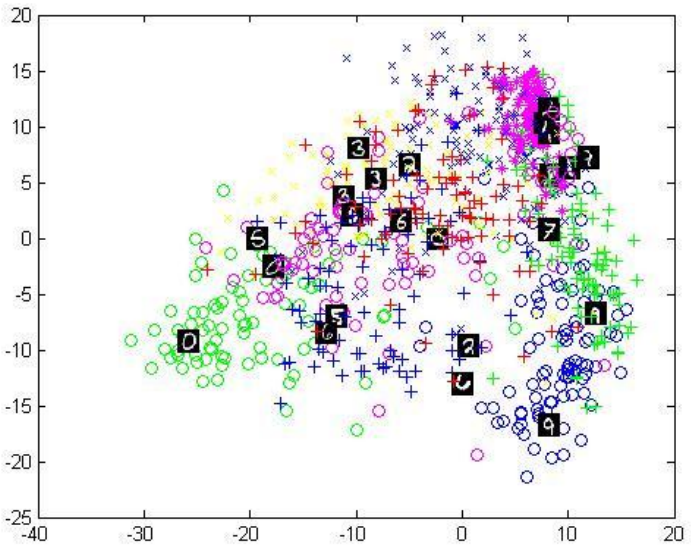
Isomap is low-dimensional embedding methods, where geodesic distances on a weighted graph are incorporated with the classical scaling. The algorithm provides a simple method for estimating the intrinsic geometry of a data manifold based on a rough estimate of each data point's neighbors on the manifold. Unlike classical techniques such as principal component analysis (PCA) and multidimensional scaling (MDS), this approach is capable of discovering the nonlinear degrees of freedom that underlie complex natural observations, such as human handwriting or images of a face under different viewing conditions. In contrast to previous algorithms for nonlinear dimensionality reduction, this algorithm efficiently computes a globally optimal solution, and, for an important class of data manifolds, is guaranteed to converge asymptotically to the true structure.

Steps

1 **Construct neighborhood graph**	Define the graph G over all data points by connecting points i and j if [as measured by dX(i, j)] they are closer than e (e-Isomap), or if i is one of the K nearest neighbors of j (K-Isomap). Set edge lengths equal to dX(i, j).
2 **Compute shortest paths**	Initialize dG(i, j) 5 dX(i, j) if i, j are linked by an edge; dG(i, j) 5 ` otherwise. Then for each value of k 5 1, 2, . . ., N in turn, replace all entries dG(i, j) by min{dG(i, j), dG(i,k) 1 dG(k, j)}. The matrix of Þnal values DG 5 {dG(i, j)} will contain the shortest path distances between all pairs of points in G (16, 19).

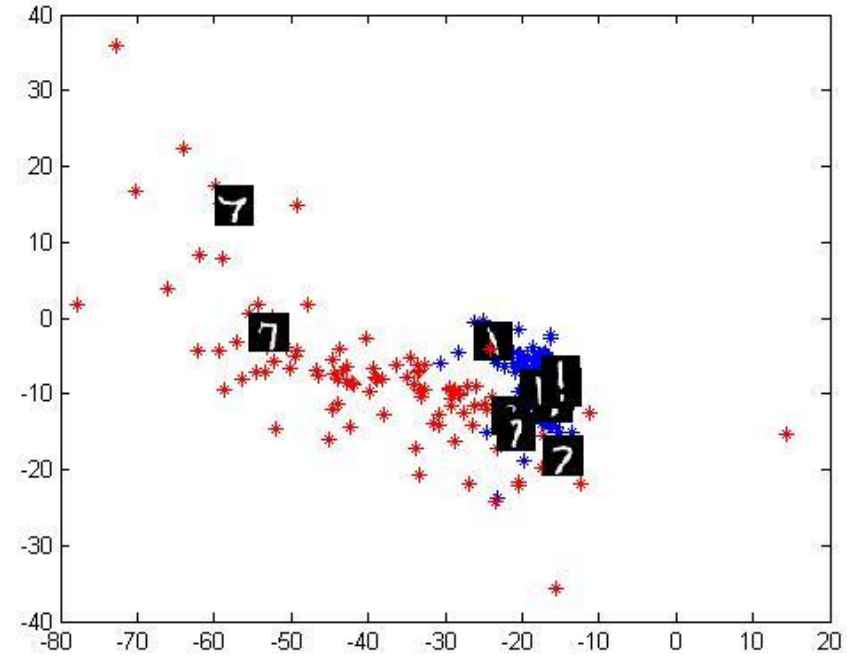
2-D isomap with Euclidean distance

	Projection of 1's and 7's on a 2-D isomap. The Red o's represent 7's while the Blue o's represent the 1's.
	Projection of 4's and 9's on a 2-D isomap. The Red o's represent 9's while the Blue o's represent the 4's.

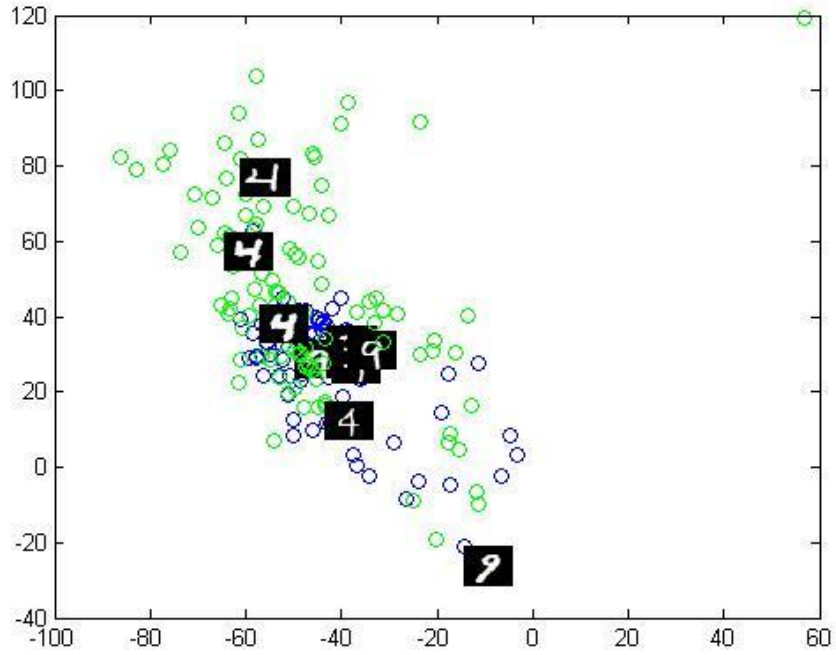


Projection of all the ten digits on a 2-D isomap. Each of the digits is represented by a different type of symbol. Also shown some of the digits on the maps, as in the **tenenbaum paper**

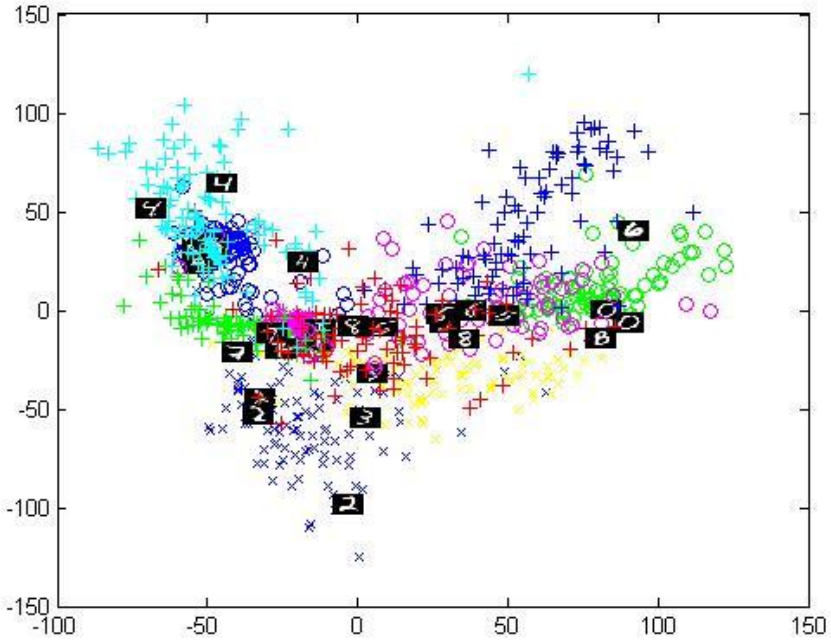
2-D isomap with Tangent distance



Projection of 1's and 7's on a 2-D isomap. The Red *'s represent 7's while the Blue *'s represent the 1's.



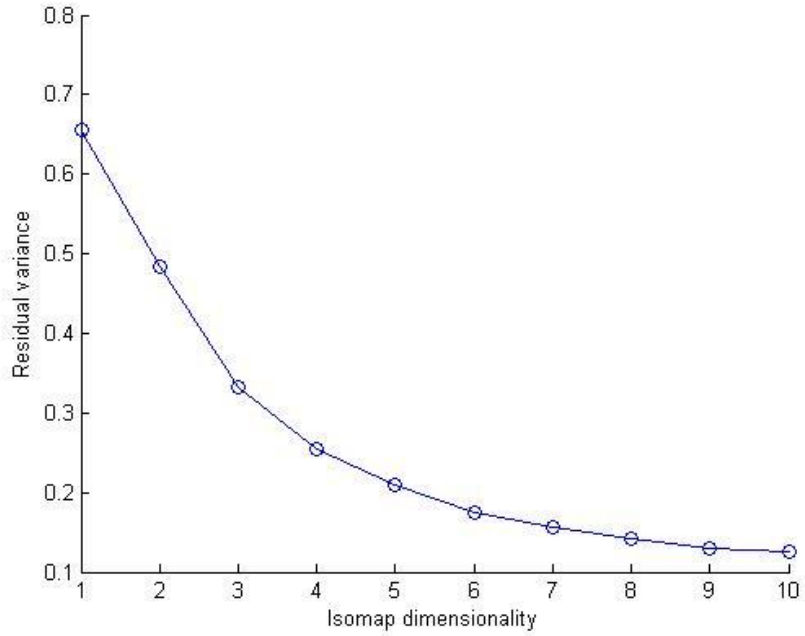
Projection of 4's and 9's on a 2-D isomap. The blue o's represent 9's while the green o's represent the 4's.



Projection of all the ten digits on a 2-D isomap. Each of the digits is represented by a different type of symbol. Also shown some of the digits on the maps, as in the **tenenbaum paper**

OBSERVATIONS- When tangent distances are used, they lead to more distinguishable clustering. This could be easily seen in the case of 4's and 9's which were hardly distinguishable with Euclidian distance but formed distinct clusters when tangent distances were used. Tangent

distance maps the digits on a larger scale [-150 150] and the variance among instances of same digits is lesser.



As we could observe in the residual variance graph there is lot of residual variance in 2 dimensional isomap. In about 6-7 dimension the variance decreases significantly and we could hope for better distinguishability among different digits.