

Assignment 1 Q)2

ISOMAP: The Isomap algorithm takes as input the distances $d_X(i, j)$ between all pairs i, j from N data points in the high-dimensional input space X , measured either in the standard Euclidean metric or in some domain-specific metric (as in Fig. 1B). The algorithm outputs coordinate vectors y_i in a d -dimensional Euclidean space Y that (according to Eq. 1) best represent the intrinsic geometry of the data. The only free parameter (ϵ or K) appears in Step 1.

Step:

1 Construct neighborhood graph: Define the graph G over all data points by connecting points i and j if [as measured by $d_X(i, j)$] they are closer than ϵ (ϵ -Isomap), or if i is one of the K nearest neighbors of j (K -Isomap). Set edge lengths equal to $d_X(i, j)$.

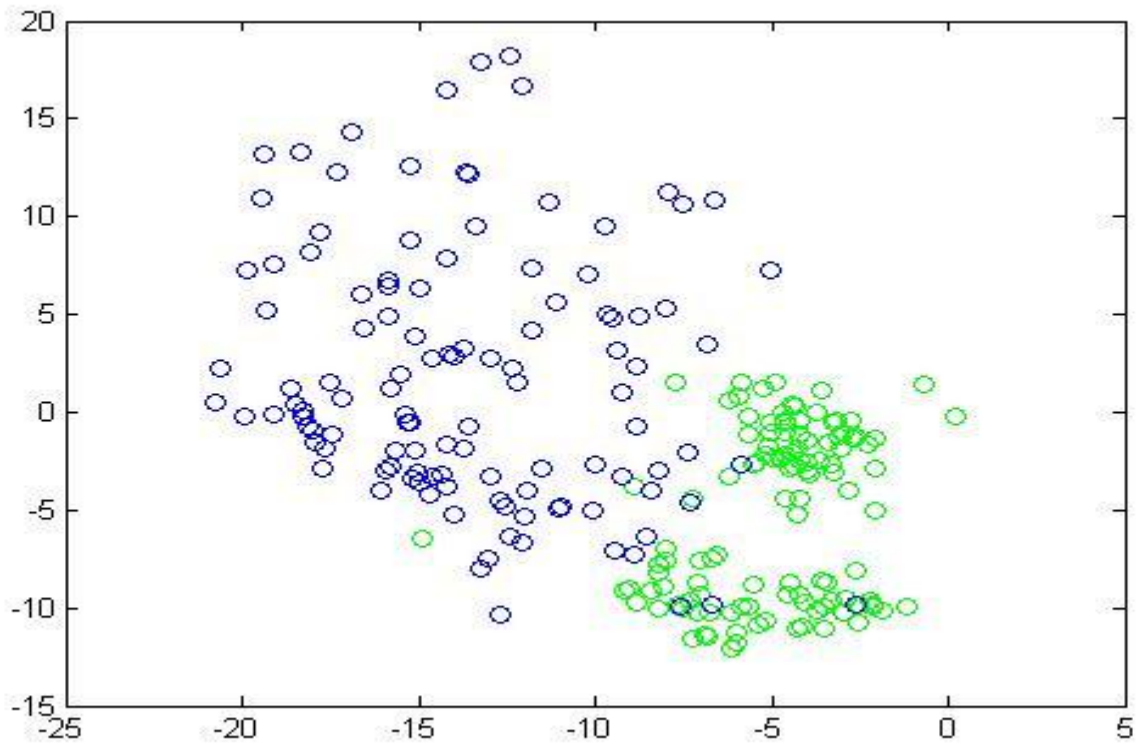
2 Compute shortest paths: Initialize $d_G(i, j) = d_X(i, j)$ if i, j are linked by an edge; $d_G(i, j) = \infty$ otherwise. Then for each value of $k = 1, 2, \dots, N$ in turn, replace all entries $d_G(i, j)$ by $\min\{d_G(i, j), d_G(i, k) + d_G(k, j)\}$. The matrix of N values $d_G(i, j)$ will contain the shortest path distances between all pairs of points in G (16, 19).

3 Construct d -dimensional embedding: Let λ_p be the p -th eigenvalue (in decreasing order) of the matrix $t(DG)$ (17), and $(V_p)^i$ be the i -th component of the p -th eigenvector. Then set the p -th component of the d -dimensional coordinate vector Y_i equal to $\sqrt{\lambda_p} (V_p)^i$.

Q2.1)

1:GREEN 7:BLUE

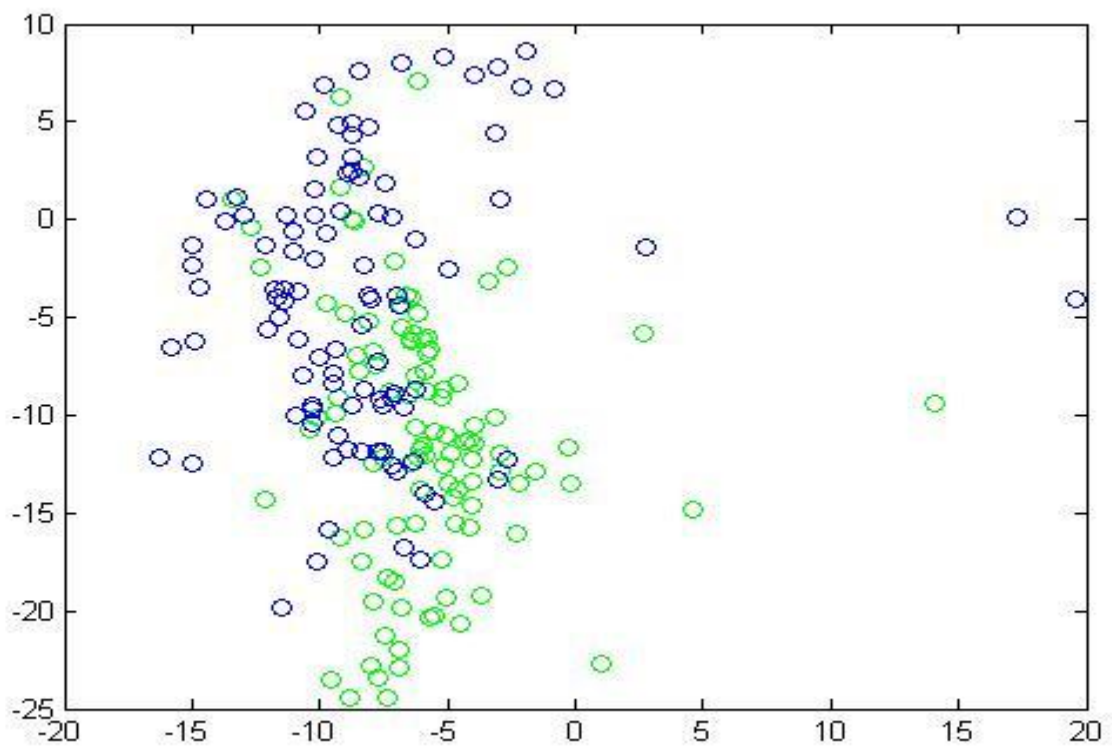
Q2.1.1



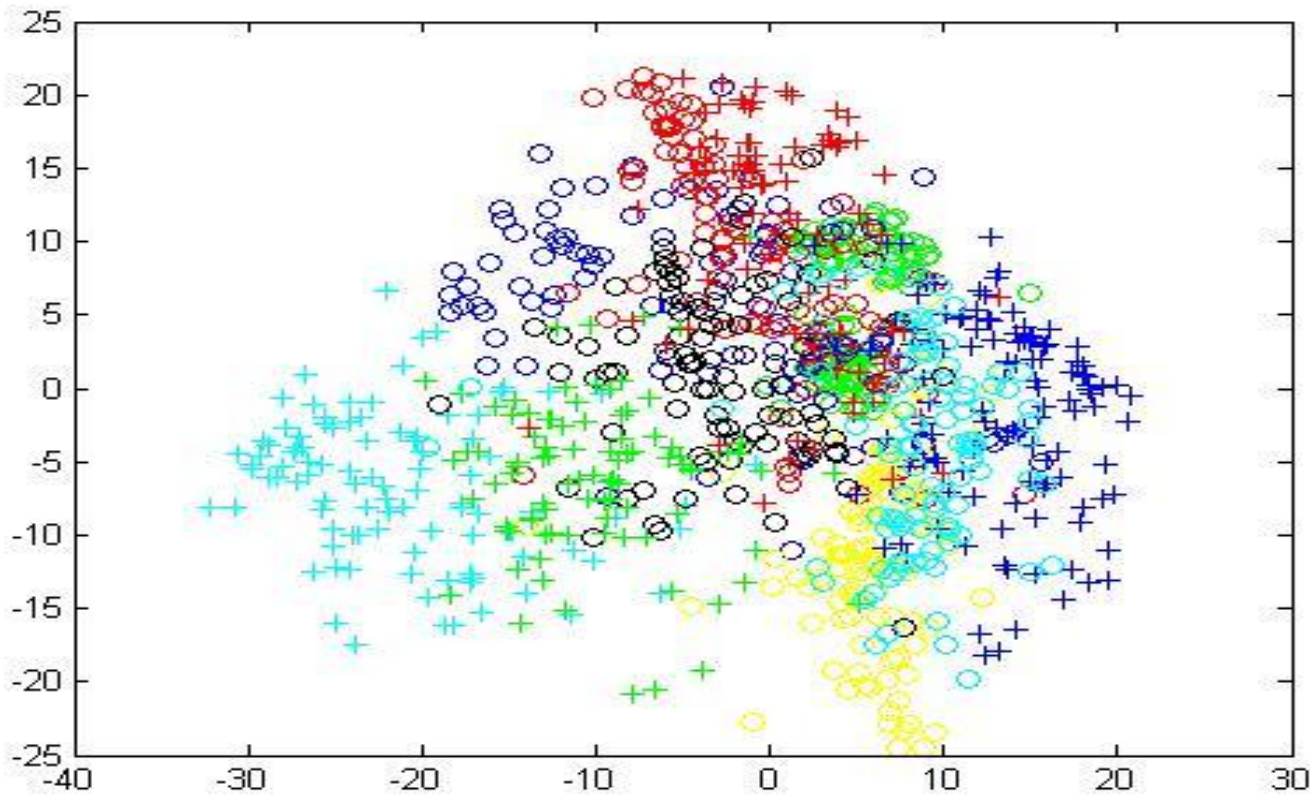
Q2.1.2)

4:GREEN

9:BLUE

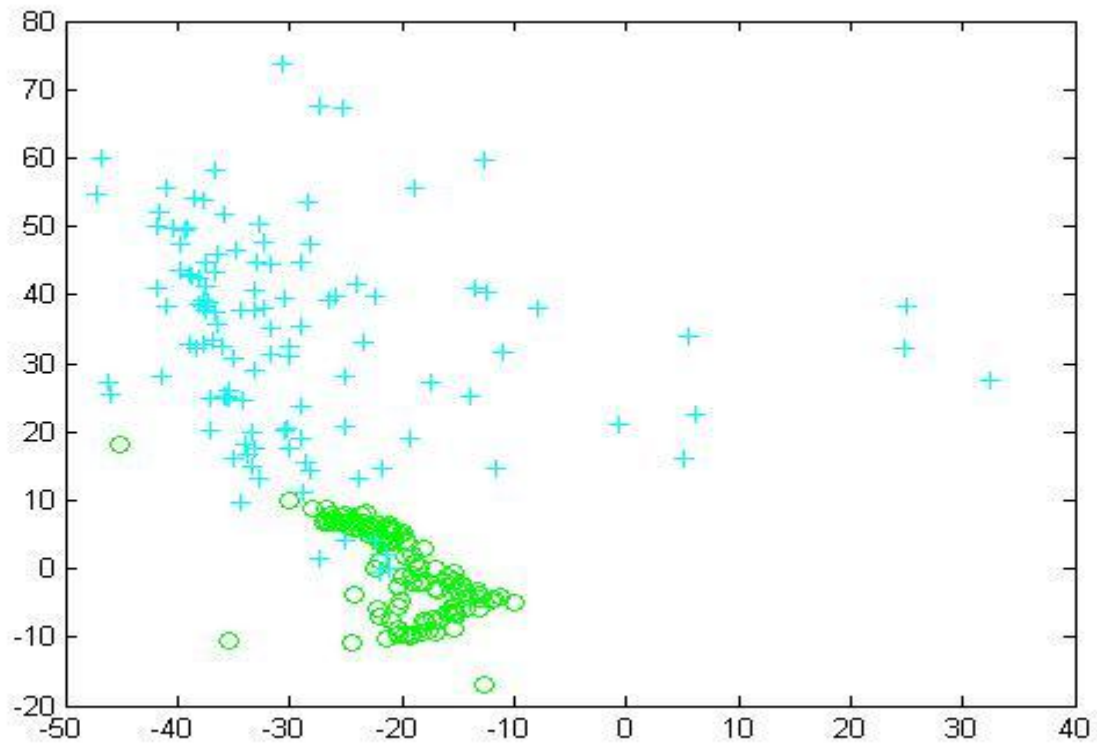


Q2.1.3) 0=cyan+ 1=green0 2= blue0 3=red0 4=yellow0 5=black0
6=green+ 7=blue+ 8=red+ 9=cyane0

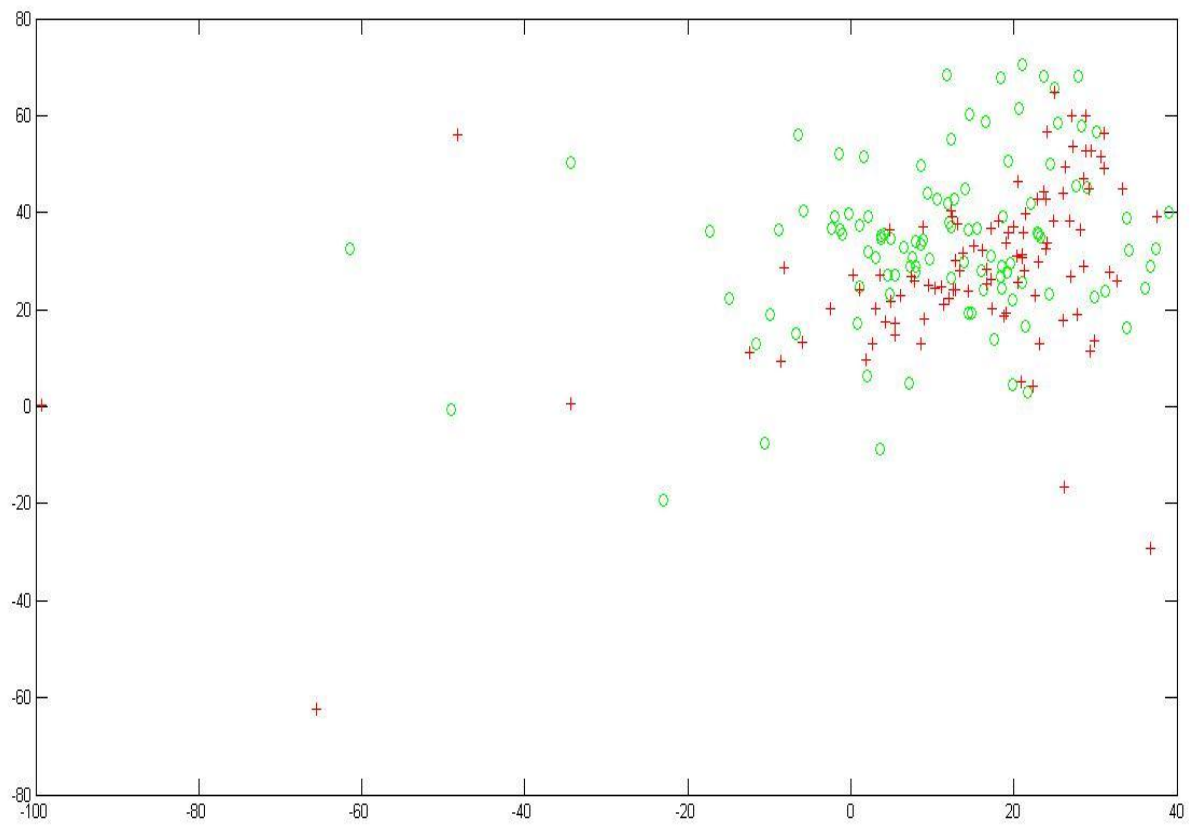


Q2.2.1)

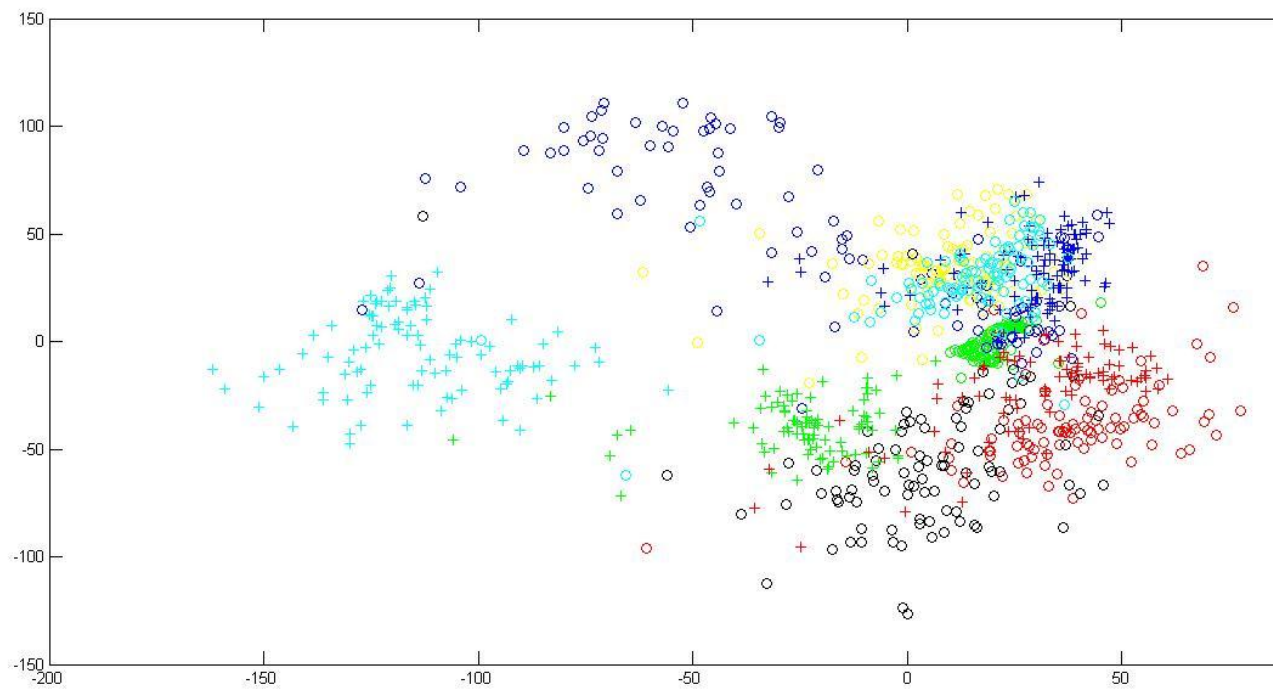
1: GREEN 7:cyane+



Q2.2.2) 4:Green 9:red+



Q2.2.3) 0=cyan+ 1=green0 2=blue0 3=red0 4=yellow0 5=black0 6=green+
7=blue+ 8=red+ 9=cyan0



OBSERVATION: the second part in which we use tangent distance separates the data corresponding to images of 1 are more dense in some area. The mixing

of data is less than the first one in which we used euclidean distance. the boundary of different label is more clear in second part than

first One.