Multidimensional Scaling

Download the associated data set, named "**US_Distance_Matrix.csv**" and perform multidimensional scaling on it. It contains a matrix of pairwise Euclidean distances between points.

(i) Build the matrix B we constructed in class, and explore its spectrum. Show a barplot of the top 10 eigenvalues. What do you observe, and what can you conclude about the space in which the point $x_1, \ldots, x_n$ live?

(ii) Compute a 2-dimensional embedding, by using the spectral decomposition of B, as shown in class (You may want to reflect one of the axes in your plot, to make the end result more visually appealing).

(iii) Look into Procrustes analysis to find the best alignment between the ground truth embedding, and the one recovered by cMDS.

https://www.mathworks.com/help/stats/procrustes.html

Procrustes analysis is a useful tool to know about, that comes handy in various data analysis problems. See also some recent theoretical work on generalized Procrustes

https://arxiv.org/pdf/1907.01145.pdf

(iv) You could repeat the experiment, and make the problem more difficult by either removing some (or most of) the entries in the initial distance matrix, at random (independently with some probability p), and/or by perturbing the given distances by noise (eg., either by additive Gaussian noise or outliers). Evaluate the performance of cMDS as you increase sparsity and/or the noise level, by computing a goodness of fit given by RMSE error, after factoring out the optimal rigid transformation between the true embedding and your reconstruction using Procrustes.

You may want to glance at the experimental setting in this paper and the ANE error measure used there (Section 8, top of page 32, equation (44))

http://www.stats.ox.ac.uk/~cucuring/Sensors_ASAP_TOSN_final.pdf