Notes On Accuracy and Results with Higher Dimensions

I have worked on testing the accuracy using the current existing methods with both the 3d and 4d dimensional data. In addition to those I attempted checking the accuracy using spatial.distance.cosine to compute the distance and similarity of the labels, along with homogeneity\_score: A clustering result satisfies homogeneity if all of its clusters contain only data points which are members of a single class, and completeness\_score: A clustering result satisfies completeness if all the data points that are members of a given class are elements of the same cluster.

With 4d date the time to perform the experiment substantially increases to about 45-50 minutes if we are to calculate accuracy using the existing methods: ### Normalized Mutual Information (NMI) is a normalization of the Mutual Information (MI) score to scale the results between 0 (no mutual information) and 1 (perfect correlation). In this function, mutual information is normalized by some generalized mean of H(labels\_true) and H(labels\_pred)), defined by the average\_method ### Silhouette Score Compute the mean Silhouette Coefficient of all samples. The Silhouette Coefficient is calculated using the mean intra-cluster distance (a) and the mean nearest-cluster distance (b) for each sample. The Silhouette Coefficient for a sample is (b - a) / max(a, b). To clarify, b is the distance between a sample and the nearest cluster that the sample is not a part of. Note that Silhouette Coefficient is only defined if number of labels is 2 <= n\_labels <= n\_samples - 1. ### Fowlkes\_mallows\_score Measure the similarity of two clusterings of a set of points. The Fowlkes-Mallows index (FMI) is defined as the geometric mean between of the precision and recall: FMI = TP / sqrt((TP + FP) \* (TP + FN)) This will most like be a something to consider, using a subset of the points could improve the time for 4d data. In terms of accuracy using the existing tests seen above 2d, 3d, and 4d data are showing similar results. Further more homogeneity\_score and completeness\_score are demonstrating similar results also at 99 - 100 %, similar to Normalized Mutual Information (NMI) and Fowlkes\_Mallows\_Score.