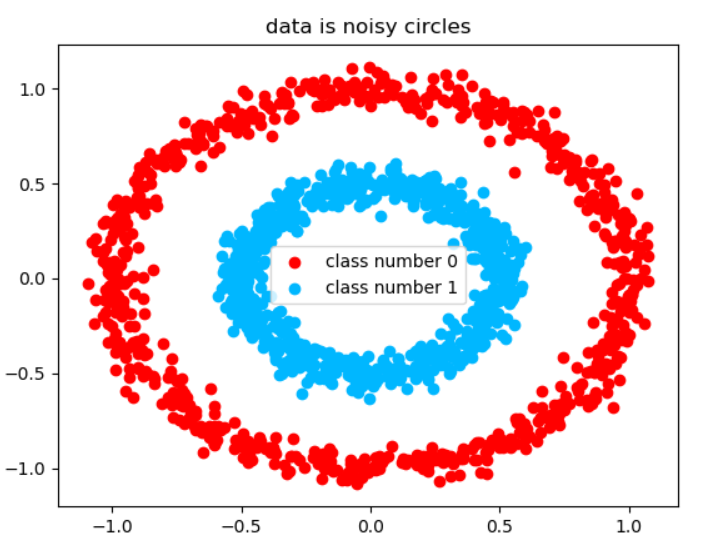
I used Agnes on three toy dataset and two real datasets. The toy data includes, noisy circles, noisy moons, and Blobs. The real data includes, iris and handwritten digits.

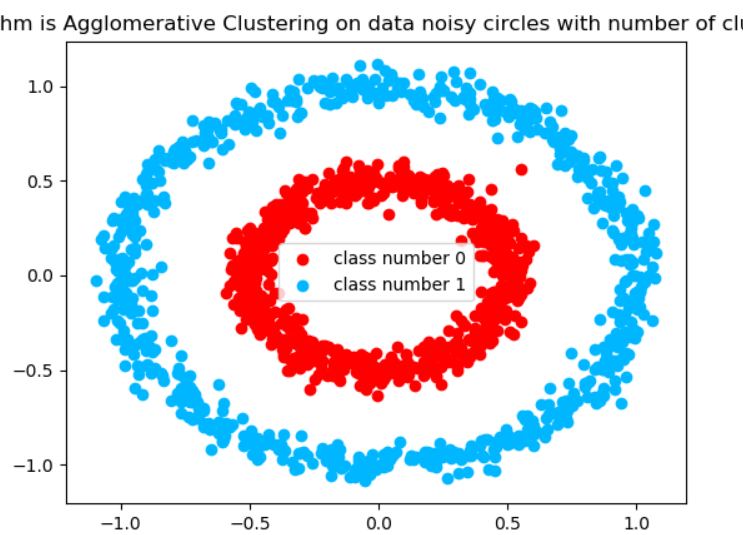
For Agnes I used average linkage algorithm which means it uses average of points to merge the data. With Euclidean distance as a metric to compute the distance between two points. I used number of classes of real data as number of clusters that we are going to find, algorithm uses this number to make a cut in dendogram. We also used a Connectivity matrix to make a neighboring graph which it uses to merge the data better.

DATA:

Noisy Circle is consists of two inner and outer circle:



Which Agnes makes it:



algorithm: Agglomerative Clustering

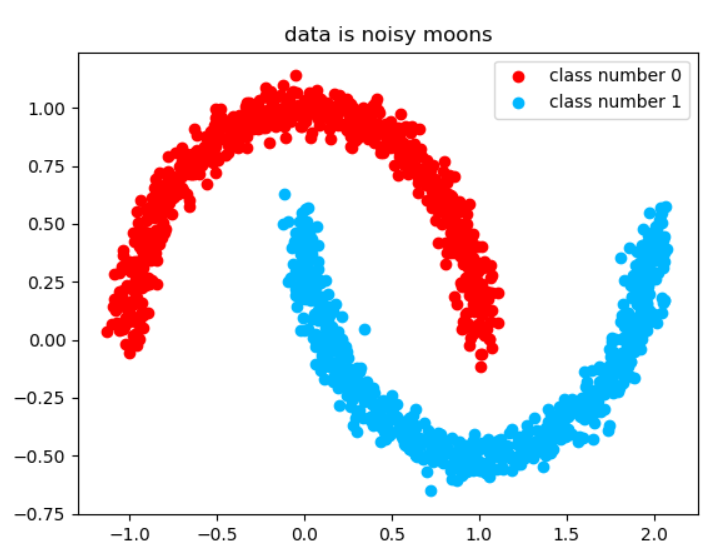
homogeneity score: 0.99

completeness score: 0.99

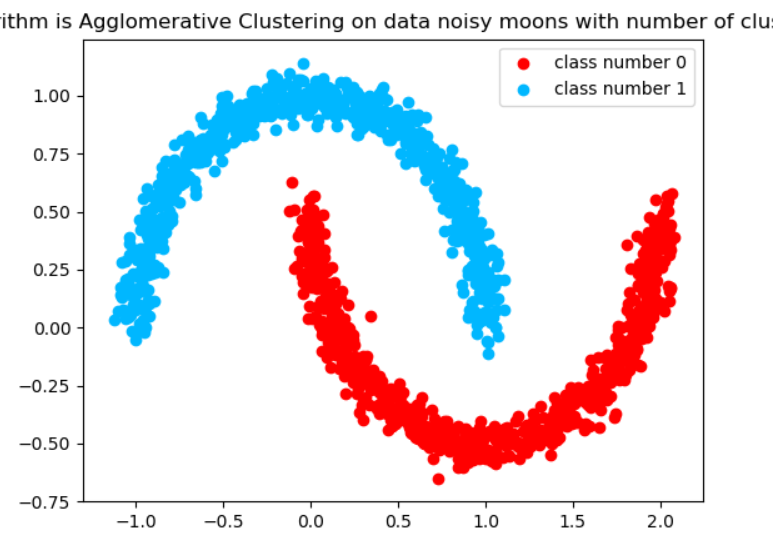
jaccard score: 0.00

normalized mutual information score: 0.99

Noisy moon is two half circle which are separated from each other:



Agnes says:



homogeneity score: 1.00

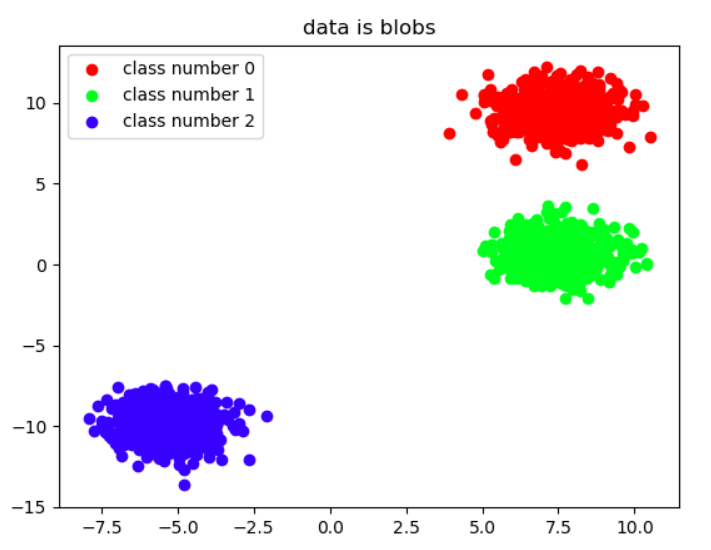
completeness score: 1.00

jaccard score: 0.00

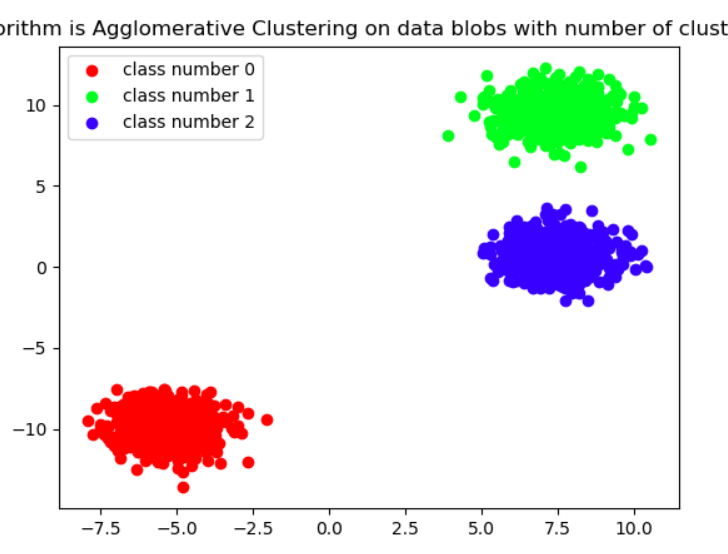
normalized mutual information score: 1.00

Has Jaccard 0 because although it is complete and homogeneous unfortunately our first class is real second class (the label mismatch) so that’s why we need homogeneity and completeness as our measurements.

Next data is Blob which is three different Gaussian mixture data and since they are separated it’s the easiest possible classification:



For Agnes:



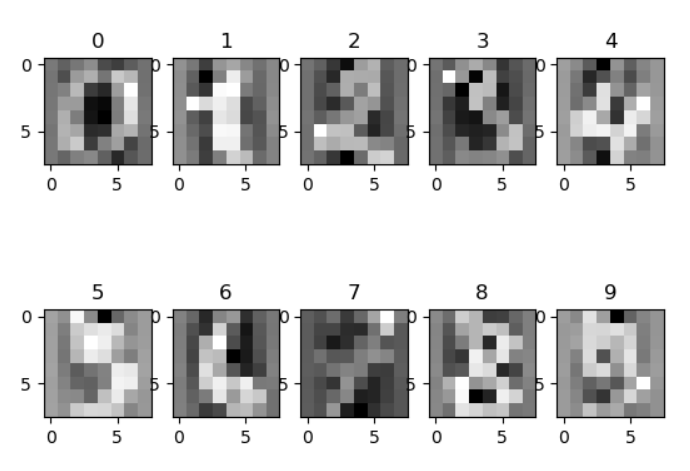
homogeneity score: 1.00

completeness score: 1.00

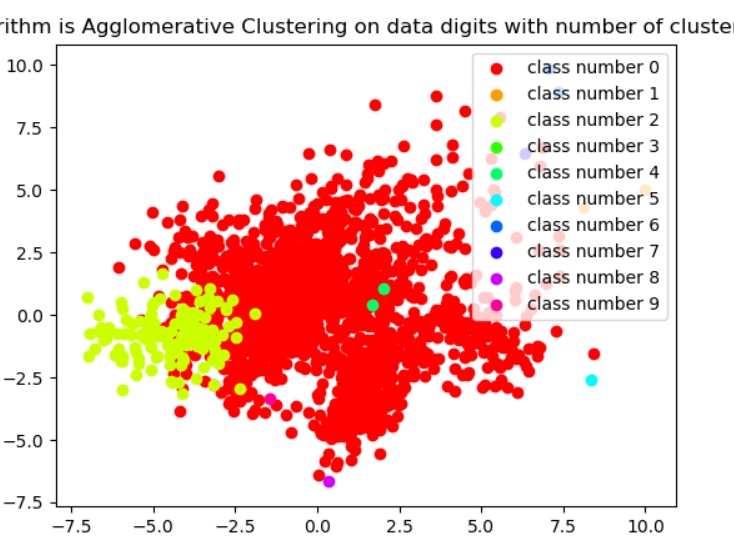
jaccard score: 0.00

normalized mutual information score: 1.00

Next data which is real life data is Handwritten digits:



For showing this data since we cannot use our whole feature data to show in two dimension we use PCA to map data in two dimension to show it there but use whole reduced feature data using PCA to number of classes to cluster them:



algorithm: Agglomerative Clustering

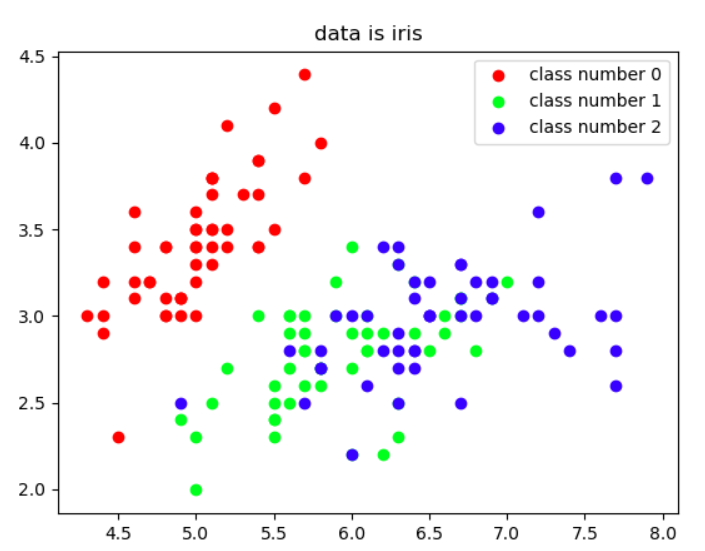
homogeneity score: 0.08

completeness score: 0.58

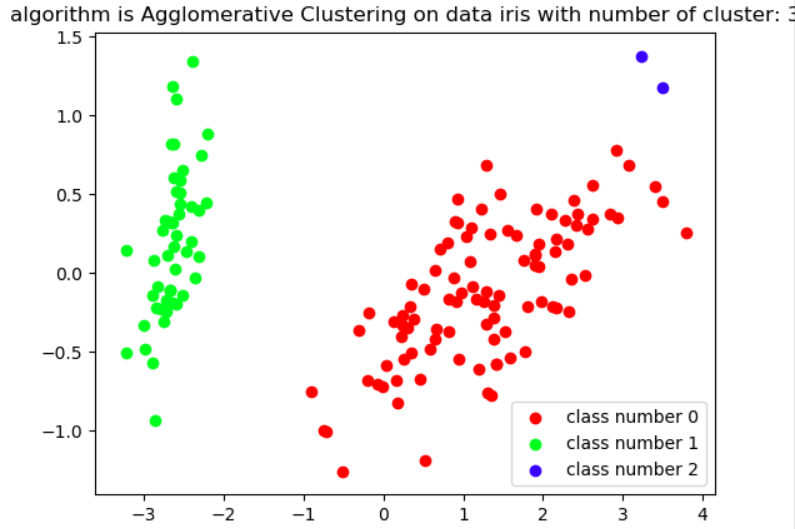
jaccard score: 0.17

normalized mutual information score: 0.21

Lastly the data is IRIS which is about flowers petal width and length and two more feature containing 3 classes:



Which Agnes:



algorithm: Agglomerative Clustering

homogeneity score: 0.59

completeness score: 0.92

jaccard score: 0.01

normalized mutual information score: 0.74

We can further improve on each results if we start tuning the hyper parameters but we decided to use same hyper for all of them.

You can see the results of fine tuning hyper parameters in the first two toy data.