Assignment8, VISULAIZATION

Mariam Ali

March 2019

1 Experience with the code

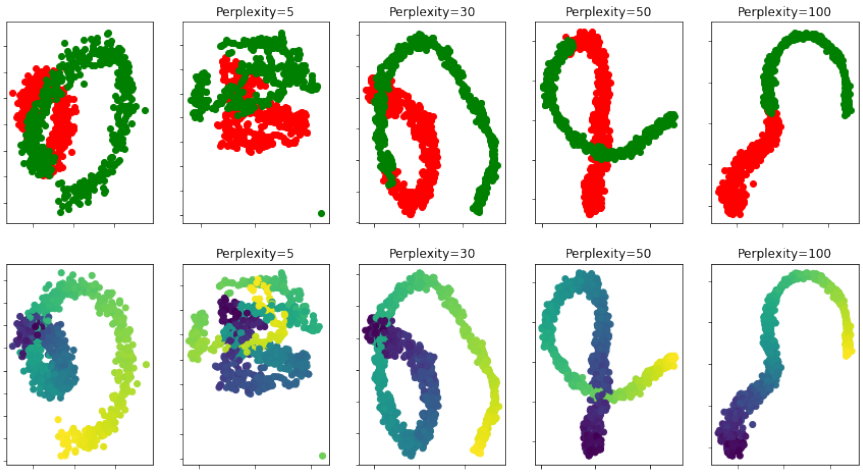
I had quiet an interesting experience trying to apply the scikit code on these data points and there corresponding "colors". In my adaptation of the code I followed the effect of change of preplexity values in the scikit TSNE function. the idea behind it is to try to figure out if the given data set is can be represented in simpler dimensions as the S-curve set or concentric circles set. in other words, this code was able to accommodated different structures to a random data set.

2 Brief explanation of abbreviations and the logic in the code

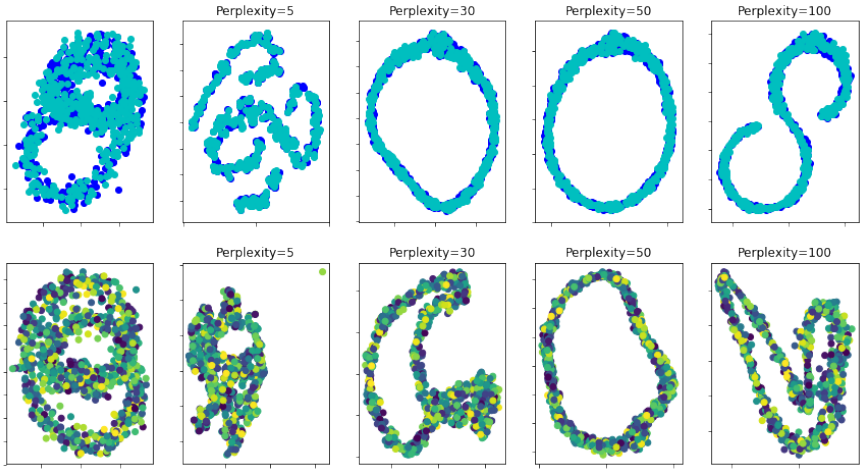
I renamed the data sets to the following letters [M, N, O, P, Q] to e easier to recall and load using pandas module. The corresponding colors (labels) are called [y_1,y_2,y_3,y_4,y_5]. Simply I tried following the concentric circles structure then the S-curve structure to find out what kind of patter does my data set would follow. I tried changing the colors as possible to be able to capture the differences in each data set.

I tried plotting a 2D grid of each data set as done in the example but my laptop ran out of memory so I tried the seaborn technique of plotting dataframes which showed some differences.

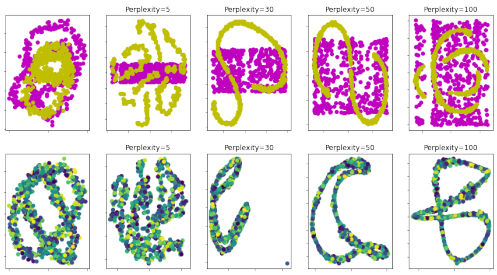
3 Results



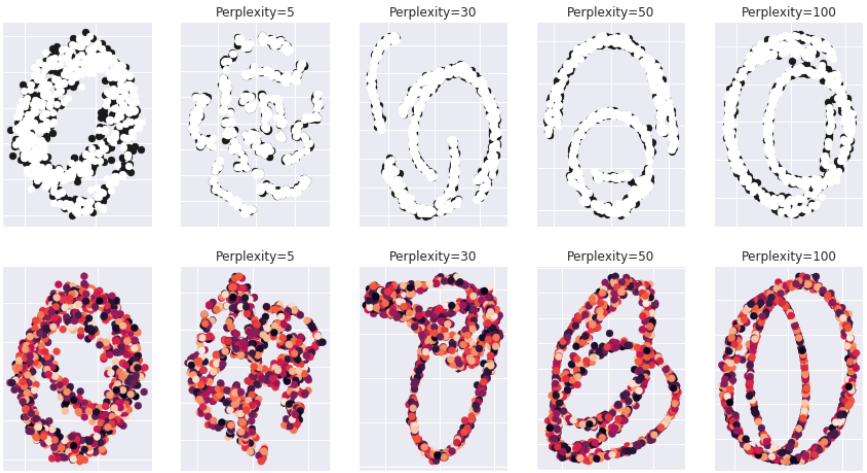
I can tell from this plot that with higher preplexity the data looses a part of its curvature or circularity. in other words, with higher preplexity the data form less concentric circles.



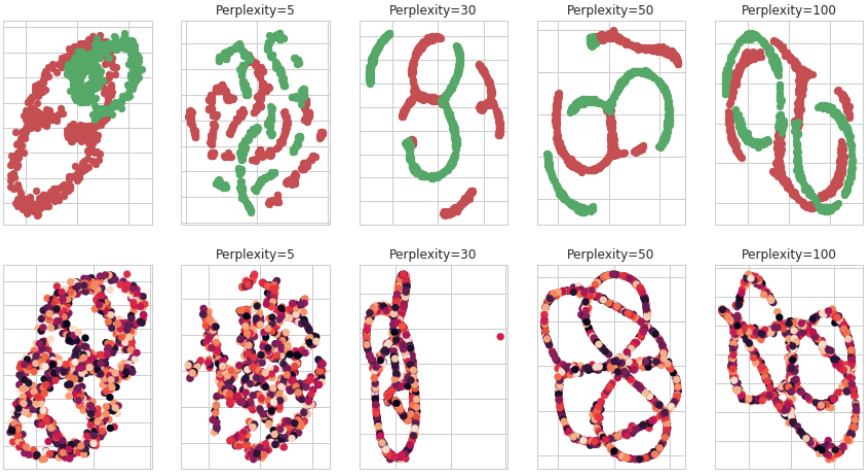
This data set has very obvious different structures, we can also see that the curvature of the data is very obvious with high values of preplexity. The S-curve example shows that only with high preplexity the looping (curving of the data shows. For the O data set similar resykts show but even stronger randomness and less alignments of circles.



We can see certain structures but this data lack much of the alignment with their labels.



The P data set of the most interesting results due to the appearance of strong curvature with high and low values of preplexity. Not only it shows curvature but it also shows very strong alignment. Finallay:



This data set show the most interesting "Butterlyish" structures and inter-connectivity. Which shows the high dimension of the given data set.