IDEAS WORKSHOP SUMMER 2017
WEEK 2

UTHAIPON TANTIPONGPIPAT

Some time spent setting things up. I didn't know before that package can be linked online. Particularly Plotly package needs some online set up and credential set up. Took some time on Google to figure this out.

I found a DeprecationWarning: " .ix is deprecated. Please use .loc for label based indexing or .iloc for positional indexing " Which I don't care yet. It works for now. I may look into the documentation to learn more later.

I used glass identification data. I tried with 2 components and 3 components. With 3 components, I do not see much more information - parts that are mixed by many types of glasses in 2D are still mixed in 3D. There is no clear hyperplane that is paralleled to or makes small angle to the X-Y plane that will separate colors well. For example, the part in the center where many colors are mixed together in 2-D is still mixed together in 3-D and not separated by being high or low on z-axis. Hence, adding the third dimension does not help to classify the type of glasses. Because of this, it's unlikely 4-D or more dimension is going to be useful. But another problem is also that I cannot do 4D-visualization plot.

The PCA shows that we can classify containers, tablewares, and headlamps reasonably well. There are some non-float window glasses that can be classified correctly as well. However, to distinguish between all three types of windows glasses (type 1,2, and 3 in the plot) is difficult. This motivates the study that if we were to improve the classification, we should look for some characteristics of glasses that are different mainly among these three.

Please see the file PCA+Glass+Identification.py for the code. Note that the file is exported from the notebook format, which is also included in the same folder. (For some reason, running the .py file on Spyder does not give the Plotly plot. But it runs and prints the numerical parts correctly.) The link is:

https://github.com/uthaipon/SkillsWorkshop2017/tree/master/Week02