MATH 637: Mathematical Techniques in Data Science
Clustering Lab

Dominique Guillot

Departments of Mathematical Sciences
University of Delaware

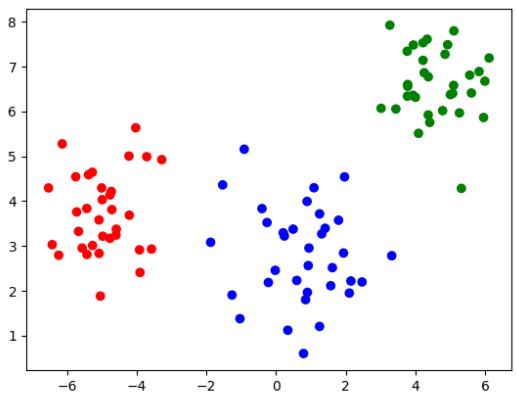
May 11, 2020

K-means

1. Use the following command from `sklearn.datasets` to general a dataset with 3 "blobs":

```
data, labels = datasets.make_blobs(n_samples=100,
                    n_features=2, centers=3)
```

2. Use `scatter` from `matplotlib` do display the data.
3. Use the K-means algorithm to cluster the data.
4. Make a scatter plot of the data, where the points in each cluster are displayed with a different color.



5. Repeat the same experiment with the following data

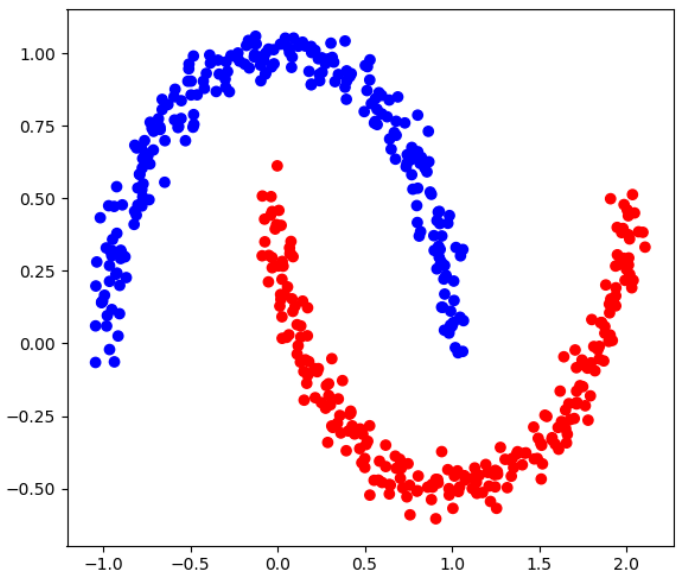
```
data, labels = datasets.make_blobs(n_samples=100,
 n_features=2, centers=[[0,0],[0,0.5],[0,-0.5]])
```

Spectral Clustering 1

1. Use the following code to generate data

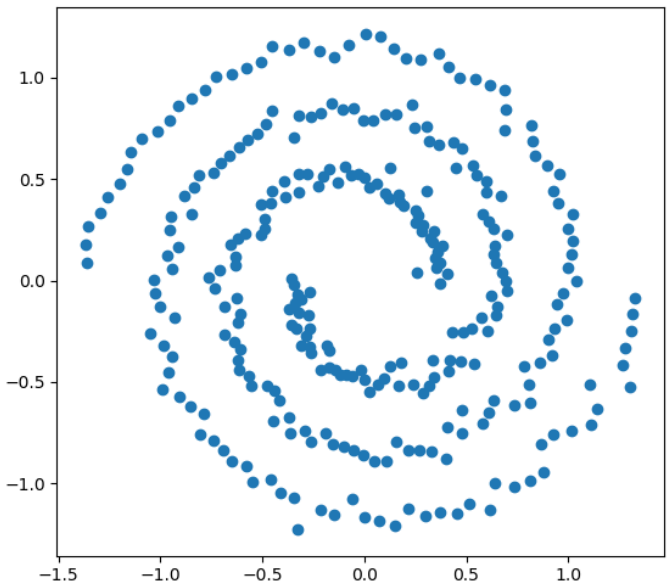
```
data, labels = datasets.make_moons(n_samples=500,
                                   noise=.05)
```

2. Use a scatter plot to display the data.
3. Use `SpectralClustering` to look for 2 clusters (which `affinity` should we use?).
4. Make a scatter plot of the data, where the points in each cluster are displayed with a different color.



Spectral clustering 2

1. Load the `spirals` dataset (available on Canvas).
2. Use the `scatter` function to display the data.



3. Use `SpectralClustering` to discover an appropriate number of clusters in the data.
4. Make a scatter plot of the data, where the points in each cluster are displayed with a different color.

Submit your work

Please submit your work on Canvas by
Friday, May 15th, 11:59 PM