Clustering Methods:

I ran K-Mean, Mean-Shift, Spectral Clustering and Affinity Propagation algorithms for 2,3,5,10 blob datasets at 3 different separation ranges each. For each combination, the algorithm was timed and the labels assigned by the algorithm compared to the y value in the original blob. The output shows how well the algorithm formed clusters based on the features input. The color of the dots in the scatter plot (see jupyter notebook) represents the predicted group for the algorithm. When the colors are grouped together, the algorithm is fairly good at forming clusters.

Jupyter Notebook:

https://github.com/spaceport729/thinkful/blob/master/Clustering%20Methods.ipynb

Conclusions: K-Mean was by far the fastest of the algorithms and also one of the most successful at recreating the original blob sets. Mean-shift was the second fastest but mostly underestimated the number of blobs. This may have been related to the assigned bandwidth. Spectral Clustering took much longer than the first two and was less good at assigning clusters when their centroids were close. Affinity Propagation almost always way over-predicted the number of clusters and took a long time. It was a lot closer (9) for the 10 blob, furthest-spaced-centroid group.

Results Table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Algorithm** | **# Blobs in Data** | **# Blobs Assigned** | **Centroid Distance** | **Run Time (ms)** | **Percent Corresponding to Original Blobs** |
| K-Means | 2 | n/a | close | 29 | 92 |
|  | 2 | n/a | medium | 13 | 100 |
|  | 2 | n/a | far | 13 | 100 |
|  | 3 | n/a | close | 51 | 77 |
|  | 3 | n/a | medium | 21 | 97 |
|  | 3 | n/a | far | 18 | 100 |
|  | 5 | n/a | close | 63 | 100 |
|  | 5 | n/a | medium | 48 | 93 |
|  | 5 | n/a | far | 24 | 65 |
|  | 10 | n/a | close | 29 | 83 |
|  | 10 | n/a | medium | 31 | 96 |
|  | 10 | n/a | far | 18 | 100 |
| Mean-Shift | 2 | 2 | close | 293 | 92 |
|  | 2 | 2 | medium | 136 | 100 |
|  | 2 | 2 | far | 140 | 100 |
|  | 3 | 1 | close | 301 | 0 |
|  | 3 | 3 | medium | 89 | 97 |
|  | 3 | 3 | far | 115 | 100 |
|  | 5 | 1 | close | 439 | ? |
|  | 5 | 4 | medium | 116 | ? |
|  | 5 | 6 | far | 312 | ? |
|  | 10 | 1 | close | 18 | ? |
|  | 10 | 2 | medium | 81 | ? |
|  | 10 | 2 | far | 88 | ? |
| Spectral Clustering | 2 | n/a | close | 472 | 91 |
|  | 2 | n/a | medium | 464 | 100 |
|  | 2 | n/a | far | 527 | 100 |
|  | 3 | n/a | close | 532 | 77 |
|  | 3 | n/a | medium | 565 | 99 |
|  | 3 | n/a | far | 582 | 99 |
|  | 5 | n/a | close | 616 | 85 |
|  | 5 | n/a | medium | 594 | 87 |
|  | 5 | n/a | far | 498 | 91 |
|  | 10 | n/a | close | 379 | 53 |
|  | 10 | n/a | medium | 330 | 84 |
|  | 10 | n/a | far | 478 | 82 |
| Affinity Propagation | 2 | 36 | close | 7900 | ? |
|  | 2 | 27 | medium | 8400 | ? |
|  | 2 | 106 | far | 8400 | ? |
|  | 3 | 36 | close | 5400 | ? |
|  | 3 | 93 | medium | 8000 | ? |
|  | 3 | 203 | far | 8700 | ? |
|  | 5 | 527 | close | 3690 | ? |
|  | 5 | 28 | medium | 2200 | ? |
|  | 5 | 37 | far | 2980 | ? |
|  | 10 | 32 | close | 2150 | ? |
|  | 10 | 23 | medium | 2640 | ? |
|  | 10 | 9 | far | 2500 | ? |