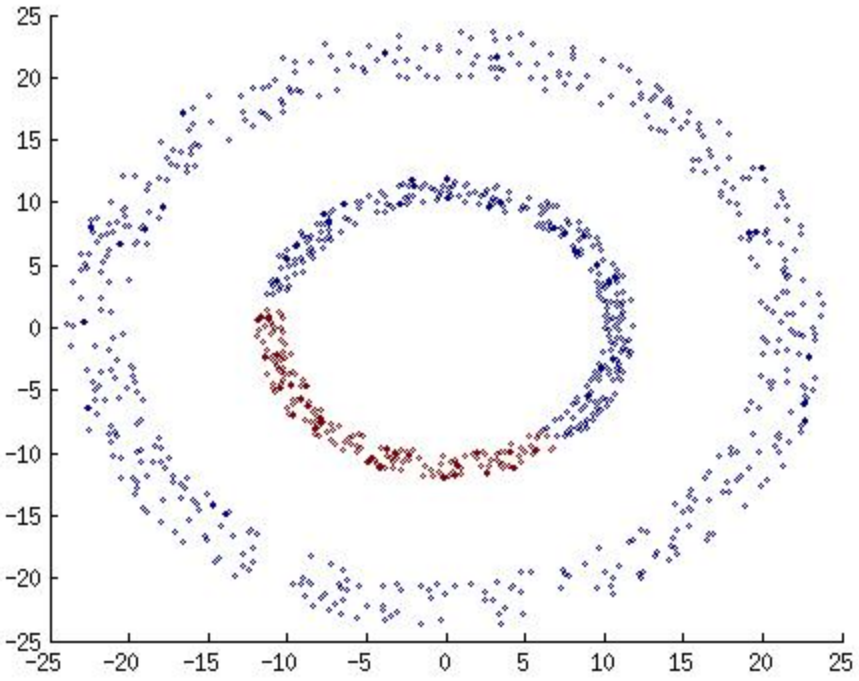
Spectral Clustering Lab/Homework

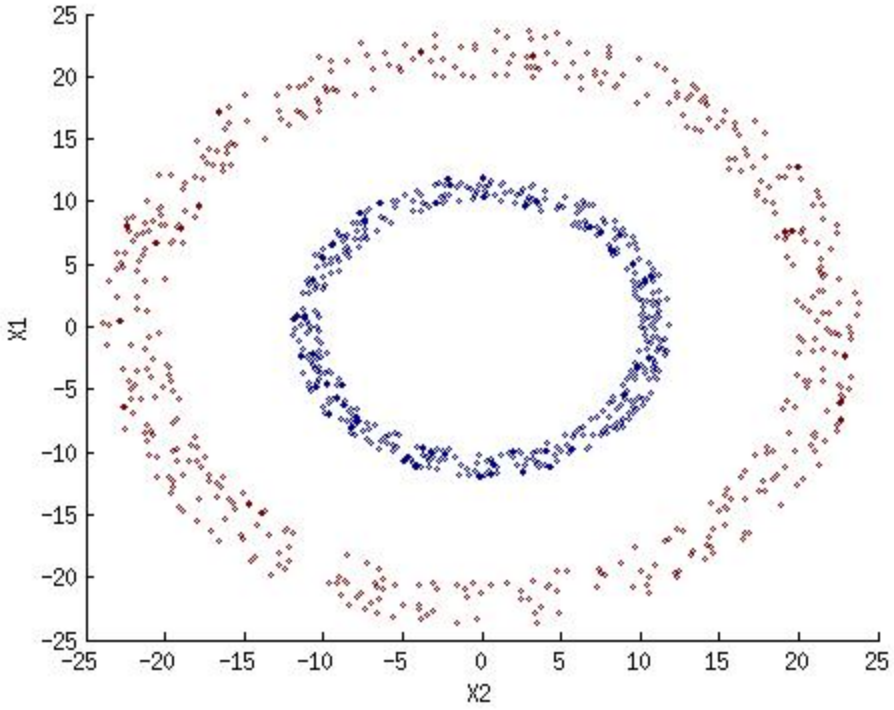
Peter Sujan, Michael Ball, Alex Caravan, Alec Guertin

3. Clustering with default parameters gives poor clusters.



4. Eigenvalues: 1.0000, 0.9998

5. The heuristic value for k was 7. We looped from 7 - 5 to 7 + 5 found that using 9 for the nearest neighbor number gave us the best clustering (100% accuracy).



6. See parameters for Dataset 1 below.

7. For all of the datasets below, we looped over k values as in part 5.

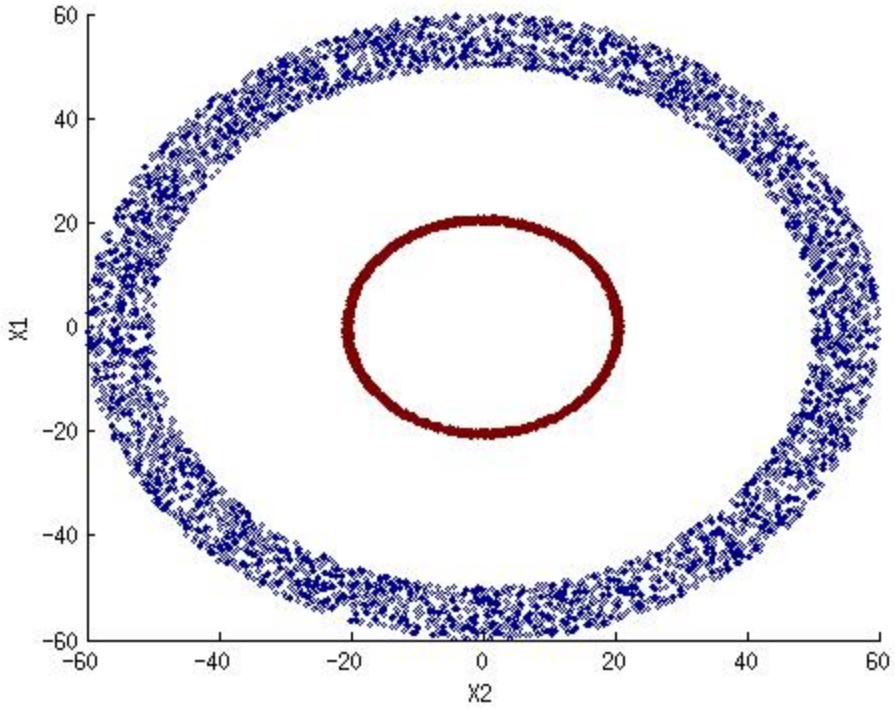
Dataset 1:

Explanation:

Distribution 1 - mean: 50, number points: 5000, variance: 0.2

Distribution 2 - mean: 50, number points: 4000, variance: 0.05

Accuracy: 100%



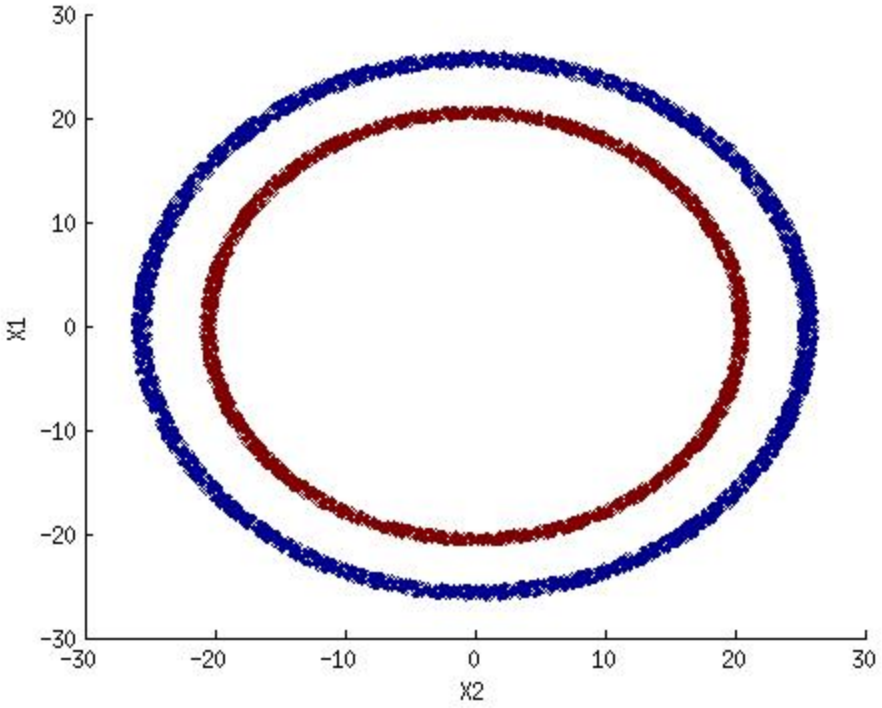
Dataset 2:

Explanation:

Distribution 1 - mean: 25, number points: 5000, variance: 0.05

Distribution 2 - mean: 20, number points: 4000, variance: 0.05

Accuracy: 100%



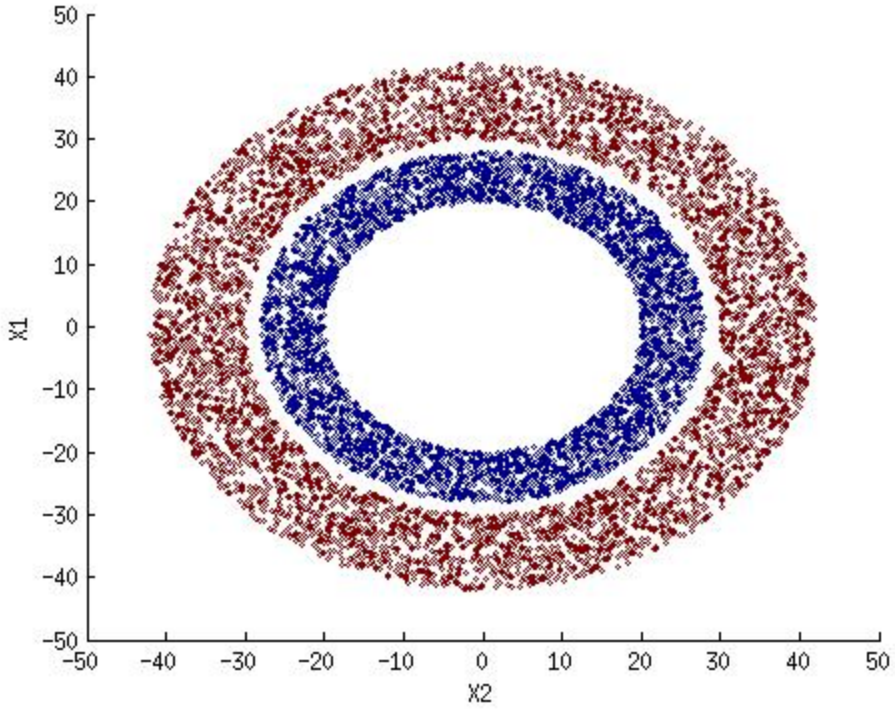
Dataset 3:

Explanation:

Distribution 1 - mean: 30, number points: 5000, variance: 0.4

Distribution 2 - mean: 20, number points: 4000, variance: 0.4

Accuracy: 100%

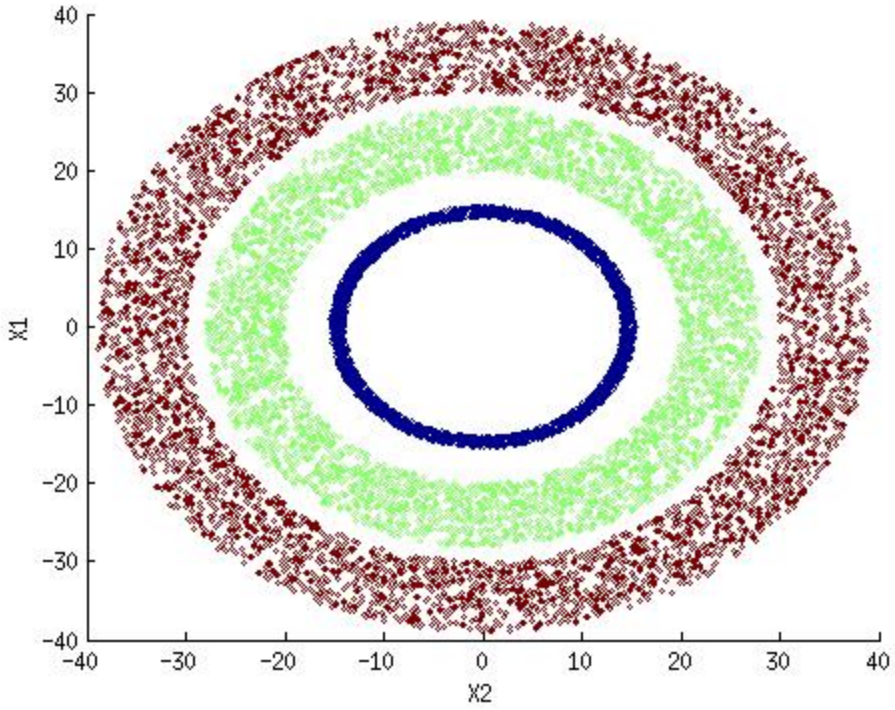


Dataset 4:

Explanation:

Distribution 1 - mean: 30, number points: 5000, variance: 0.3
Distribution 2 - mean: 20, number points: 4000, variance: 0.4
Distribution 3 - mean: 14, number points: 3000, variance: 0.1

Accuracy: 100%


40
30
20
10
0
−10
−20
−30
−40
X1
−40
−30
−20
−10
0
10
20
30
40
X2