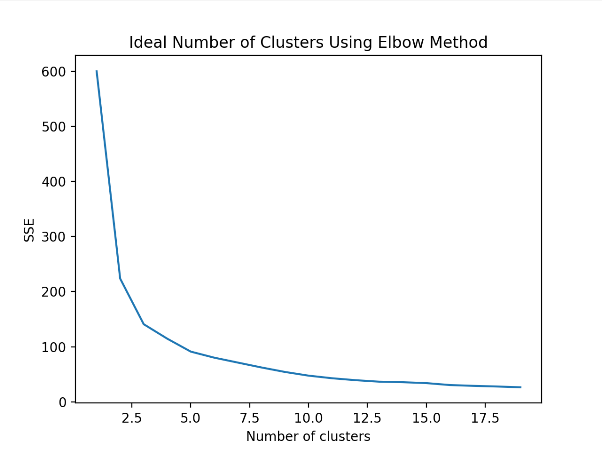
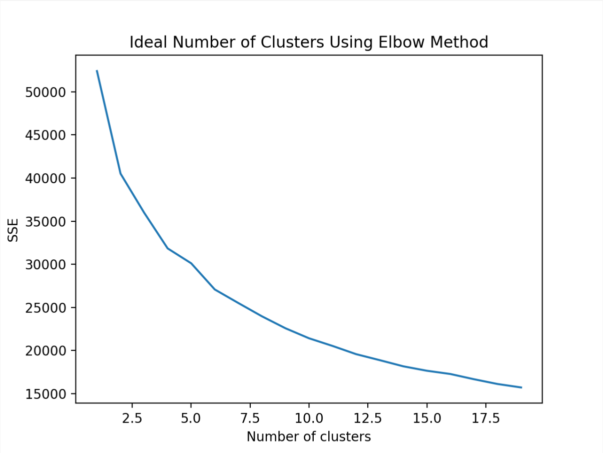
CS 487 - HW 7

Cyrus Baker
April 24, 2020

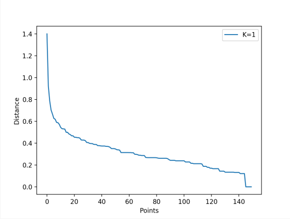
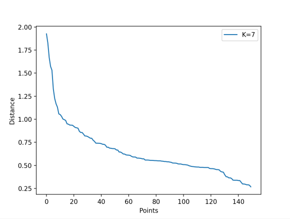
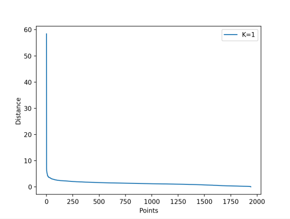
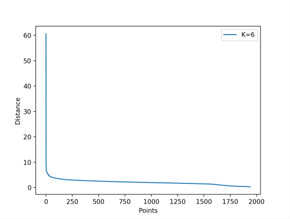
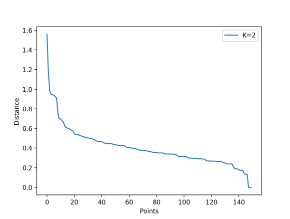
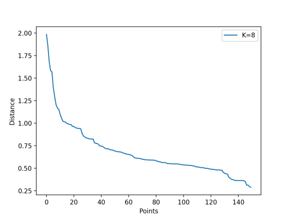
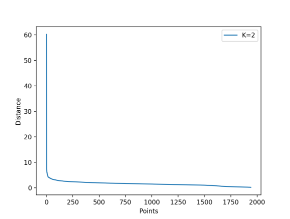
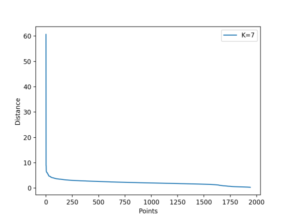
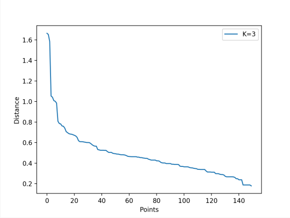
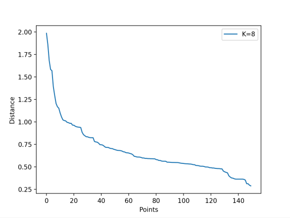
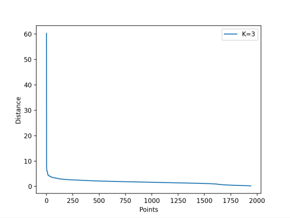
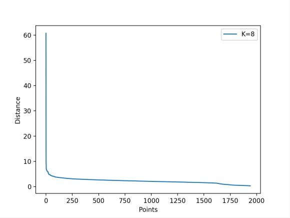
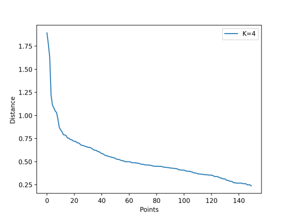
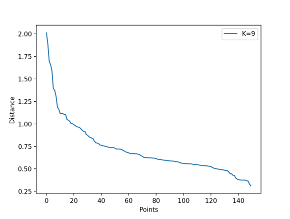
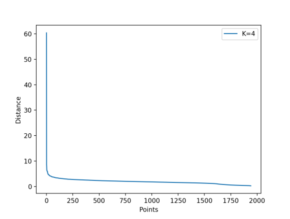
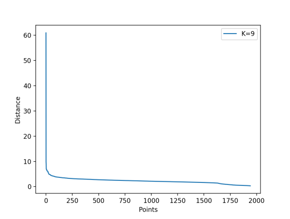
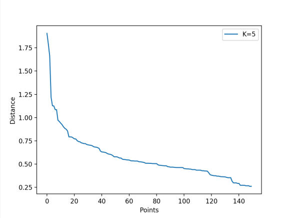
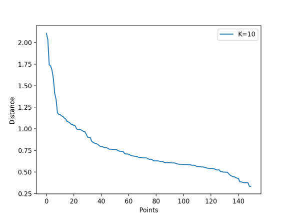
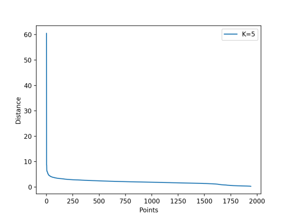
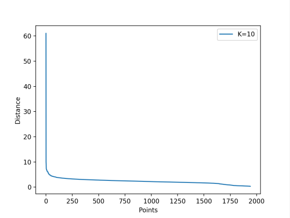
Datasets:

- Iris - UCI Repository
- faulty-steel-plates - UCI Repository

Charts

Iris	Faulty
	

These charts are used to determine the optimal number of clusters for the KMEANS clusterer.

Iris		Faulty	
			
			
			
			
			

These charts are used to determine the optimal number of min_samples and eps for the DBSCAN clusterer.

Results

Clusterer	Dataset	Iris	Faulty
KMEANS	**Error Score**	561	16348
	Homogeneity Score	0.7465	0.2933
	Completeness Score	0.5230	0.2902
	V-Measurement Score	0.6151	0.2917
SCIPY HEIRARCHICAL	**Error Score**	100	5574209
	Homogeneity Score	-4.548	0.3631
	Completeness Score	1	0.3191
	V-Measurement Score	-9.095	0.3397
SKLEARN HEIRARCHICAL	**Error Score**	297	37462
	Homogeneity Score	0.5922	0.0024
	Completeness Score	0.8431	0.1303
	V-Measurement Score	0.6958	0.0046
DBSCAN	**Error Score**	142	37677
	Homogeneity Score	0.5225	0.0013
	Completeness Score	0.6134	0.0642
	V-Measurement Score	0.5643	0.0026

Analysis

I found that my results improved for the faults dataset after I standardized the X values. However I know that something is wrong because the values for homogeneity, completeness, and v-measurement scores should range between 0 and 1 (higher is better). The erros score is the mean of the sum squared errors and a lower score is better. For the Iris dataset the Scipy clusterer can be ruled out as a candidate because of the negative scores. I would guess that the Kmeans and the Sklearn Heirarchical clusterers would work well because of the high homogeneity, completeness, and v-measurement scores.

For the faulty steel plate the most reasonable clusterer was the Scipy heirarchical clusterer. But based on my results I think that my implementation was a little off.