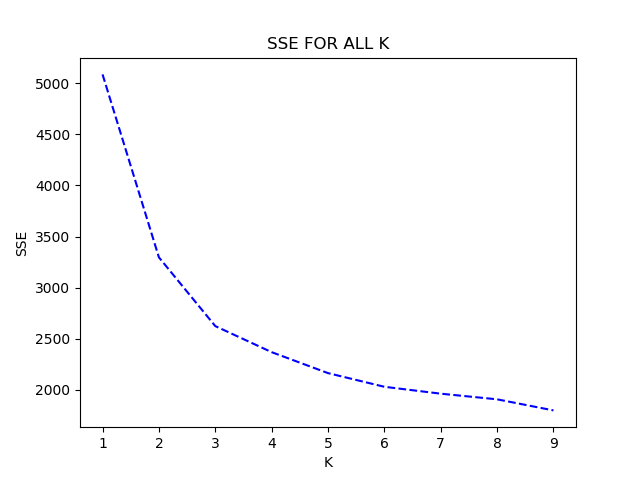
**Assignment 5**

K Means:

Dermatology:

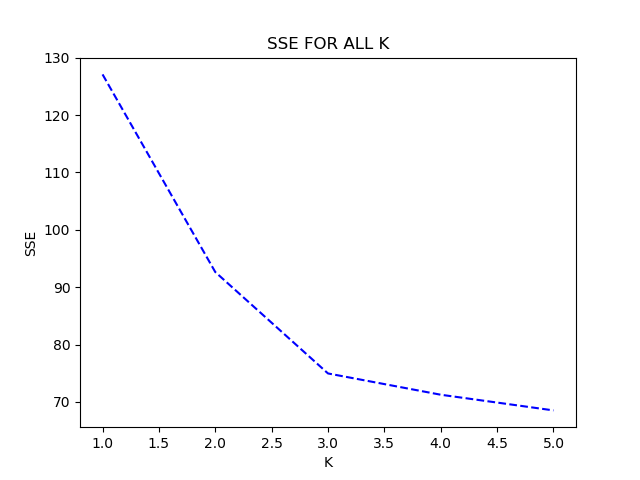
|  |  |  |
| --- | --- | --- |
| K | SSE | NMI |
| 1 | 5084.586 | -2.78E-07 |
| 2 | 3296.469 | 0.046149 |
| 3 | 2624.433 | 0.069742 |
| 4 | 2386.075 | 0.081969 |
| 5 | 2164.748 | 0.087294 |
| 6 | 2097.817 | 0.176852 |
| 7 | 1933.667 | 0.130013 |
| 8 | 1913.819 | 2.32E-01 |
| 9 | 1845.814 | 0.247762 |
| 10 | 1802.476 | 0.324336 |



Ideal k:

Actual: 6

K Means: 6

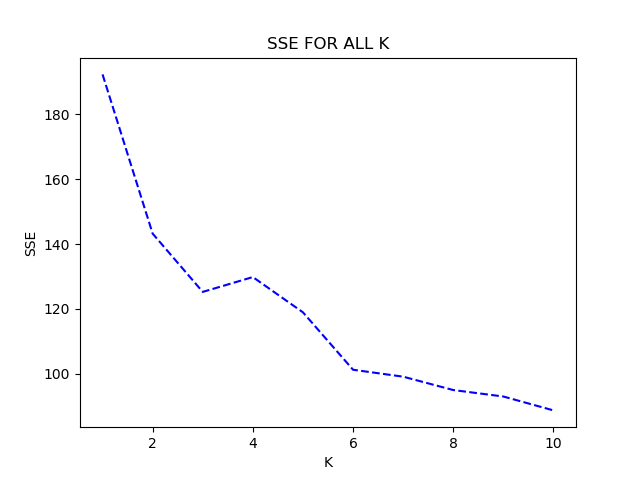
E. coli:

|  |  |  |
| --- | --- | --- |
| K | SSE | NMI |
| 1 | 127.083 | 8.33E-07 |
| 2 | 92.63649 | 0.496579 |
| 3 | 74.95997 | 0.671948 |
| 4 | 71.25152 | 0.630256 |
| 5 | 68.51955 | 0.645666 |

Ideal k:

Actual: 5

K Means: 3

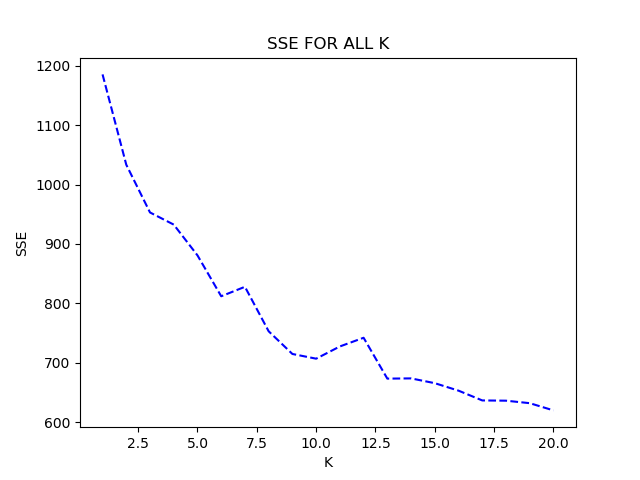
Glass:

|  |  |  |
| --- | --- | --- |
| K | SSE | NMI |
| 1 | 192.3362 | 5.27E-06 |
| 2 | 143.2293 | 0.336225 |
| 3 | 125.2594 | 0.243395 |
| 4 | 129.8375 | 0.314032 |
| 5 | 118.9576 | 0.405549 |
| 6 | 101.2529 | 0.333496 |
| 7 | 99.11394 | 0.352113 |
| 8 | 94.99398 | 3.45E-01 |
| 9 | 92.99762 | 0.37687 |
| 10 | 88.71471 | 0.31383 |

Ideal k:

Actual: 6

K Means: 6

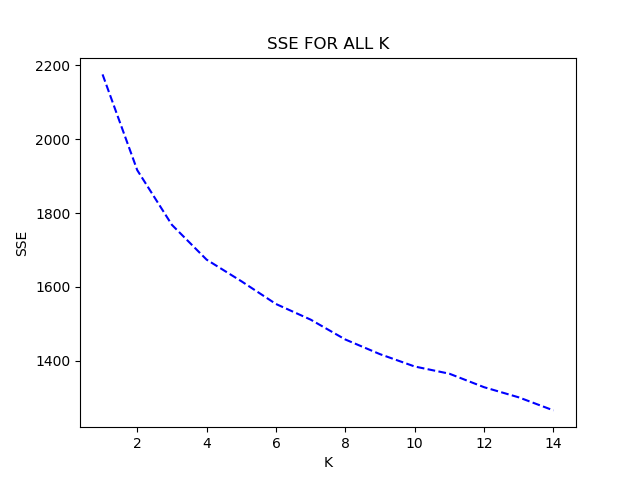
Soybean:

|  |  |  |
| --- | --- | --- |
| K | SSE | NMI |
| 1 | 1185.476 | 2.91E-06 |
| 2 | 1033.405 | 0.454256 |
| 3 | 953.0031 | 0.399262 |
| 4 | 932.5651 | 0.480289 |
| 5 | 880.7461 | 0.533287 |
| 6 | 811.9178 | 0.577902 |
| 7 | 827.8416 | 0.459402 |
| 8 | 752.9219 | 6.01E-01 |
| 9 | 714.7363 | 0.639435 |
| 10 | 706.8096 | 0.659569 |
| 11 | 727.391 | 0.630197 |
| 12 | 742.0206 | 0.624491 |
| 13 | 673.2351 | 0.679448 |
| 14 | 673.6076 | 0.679415 |
| 15 | 665.5605 | 0.633045 |
| 16 | 652.9627 | 0.666755 |
| 17 | 636.5018 | 0.652768 |
| 18 | 636.0786 | 0.670519 |
| 19 | 631.8907 | 0.666061 |
| 20 | 619.7204 | 0.65062 |

Ideal k:

Actual: 15

K Means: 15

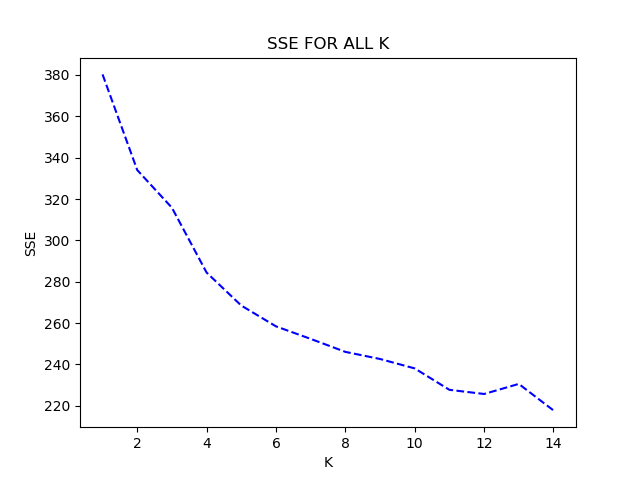
Vowels:

|  |  |  |
| --- | --- | --- |
| K | SSE | NMI |
| 1 | 2175.805 | 9.16E-06 |
| 2 | 1916.355 | 0.245468 |
| 3 | 1768.009 | 0.25742 |
| 4 | 1673.786 | 0.376221 |
| 5 | 1615.384 | 0.344678 |
| 6 | 1553.536 | 0.422278 |
| 7 | 1511.169 | 0.404531 |
| 8 | 1457.324 | 3.95E-01 |
| 9 | 1417.561 | 0.389966 |
| 10 | 1384.148 | 0.421232 |
| 11 | 1364.616 | 0.408069 |
| 12 | 1328.051 | 0.456692 |
| 13 | 1300.543 | 0.449888 |
| 14 | 1265.19 | 0.449089 |

Ideal k:

Actual: 11

K Means: 10

Yeast:

|  |  |  |
| --- | --- | --- |
| K | SSE | NMI |
| 1 | 380.2507 | -5.41E-06 |
| 2 | 334.0049 | 0.169615 |
| 3 | 315.7866 | 0.167129 |
| 4 | 284.4459 | 0.20801 |
| 5 | 268.4929 | 0.20699 |
| 6 | 258.4393 | 0.24687 |
| 7 | 252.3811 | 0.24915 |
| 8 | 246.1029 | 2.37E-01 |
| 9 | 242.6619 | 0.269087 |
| 10 | 238.1064 | 0.261163 |
| 11 | 227.7592 | 0.2698 |
| 12 | 225.7414 | 0.270948 |
| 13 | 230.6102 | 0.25985 |
| 14 | 217.8066 | 0.262554 |

Ideal k:

Actual: 9

K Means: 10

How I selected optimal number of clusters?

Answer: the point where the SSE starts getting stable in graph

GMM:

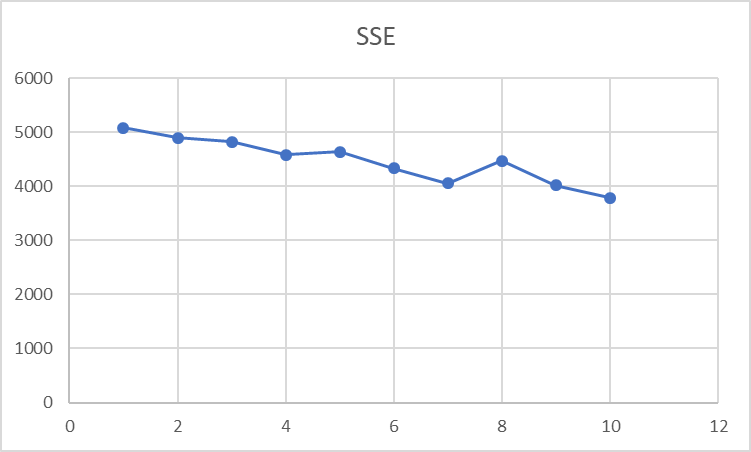
Dermatology:

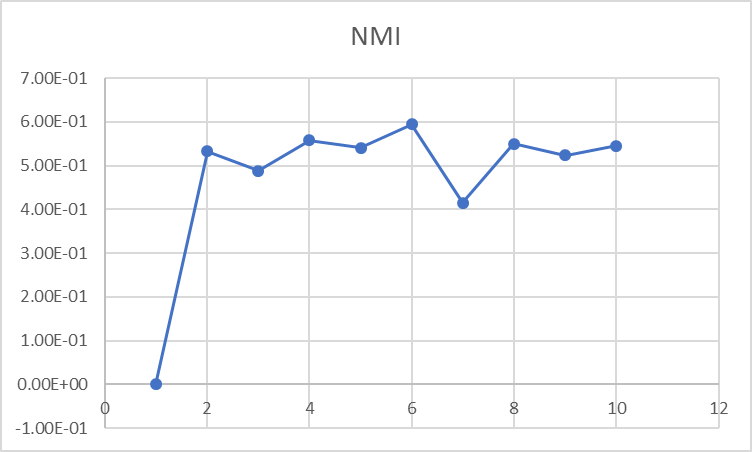
Ideal k:

Actual: 6

GMM: 6

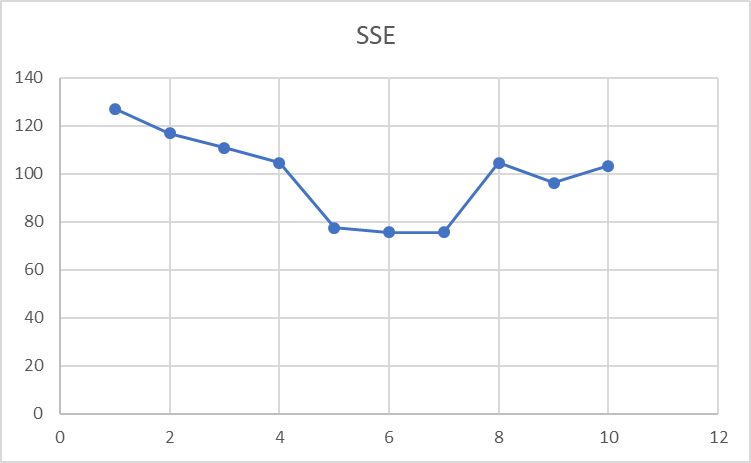
|  |  |  |
| --- | --- | --- |
| K | SSE | NMI |
| 1 | 5084.586 | -2.78E-07 |
| 2 | 4899.141 | 0.53326758 |
| 3 | 4825.165 | 0.48871424 |
| 4 | 4579.893 | 0.558028 |
| 5 | 4642.662 | 0.54045468 |
| 6 | 4332.848 | 0.594104 |
| 7 | 4055.453 | 0.41455157 |
| 8 | 4475.641 | 5.50E-01 |
| 9 | 4019.987 | 0.52346256 |
| 10 | 3785.967 | 0.54565339 |

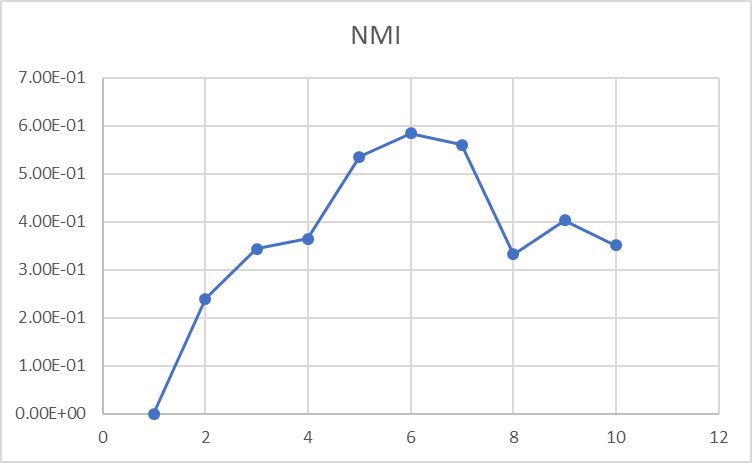




E. Coli:

|  |  |  |
| --- | --- | --- |
| K | SSE | NMI |
| 1 | 127.083 | 8.33E-07 |
| 2 | 116.8977 | 0.239046 |
| 3 | 110.962 | 0.343464 |
| 4 | 104.6582 | 0.364422 |
| 5 | 77.56856 | 0.535052 |
| 6 | 75.69719 | 0.584637 |
| 7 | 75.64225 | 0.561018 |
| 8 | 104.6715 | 3.32E-01 |
| 9 | 96.17865 | 0.403212 |
| 10 | 103.2803 | 0.351491 |





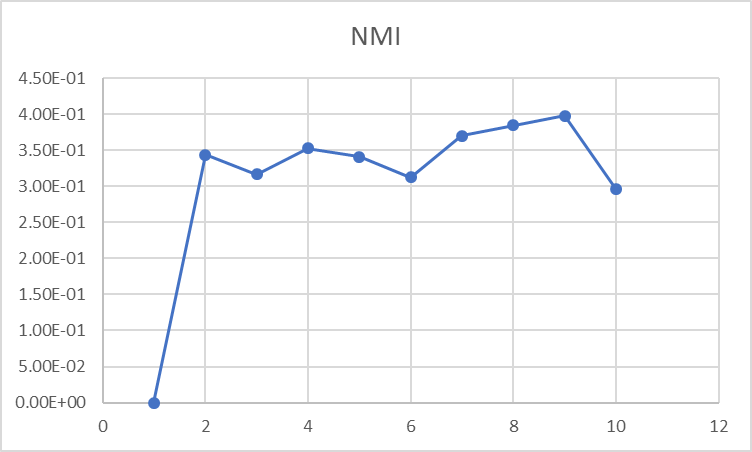
Ideal k:

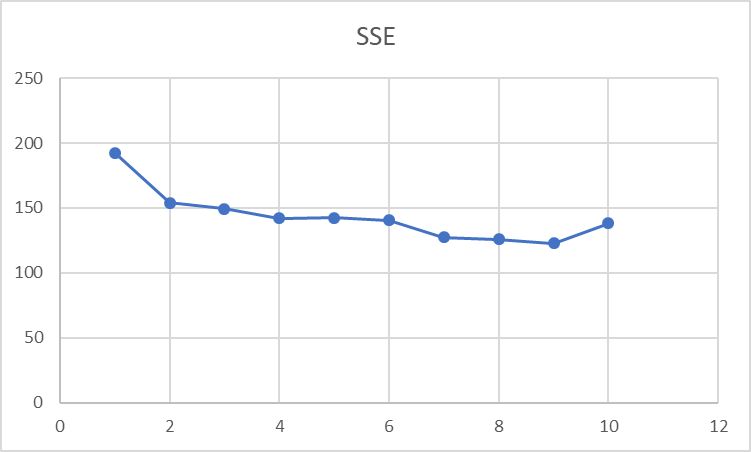
Actual: 5

GMM: 5

Glass:

|  |  |  |
| --- | --- | --- |
| K | SSE | NMI |
| 1 | 192.3362 | 5.27E-06 |
| 2 | 154.1224 | 0.343393 |
| 3 | 149.2777 | 0.316772 |
| 4 | 141.9521 | 0.352445 |
| 5 | 142.3904 | 0.341121 |
| 6 | 140.5435 | 0.312194 |
| 7 | 127.3363 | 0.370312 |
| 8 | 125.9746 | 3.85E-01 |
| 9 | 122.6888 | 0.397549 |
| 10 | 138.0302 | 0.296343 |





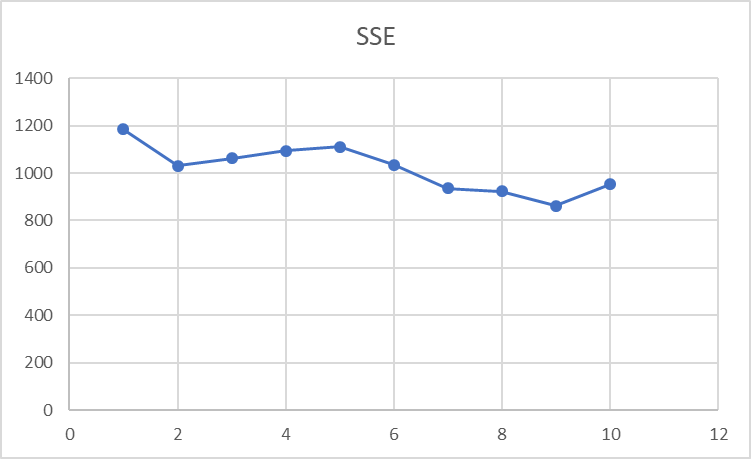
Ideal k:

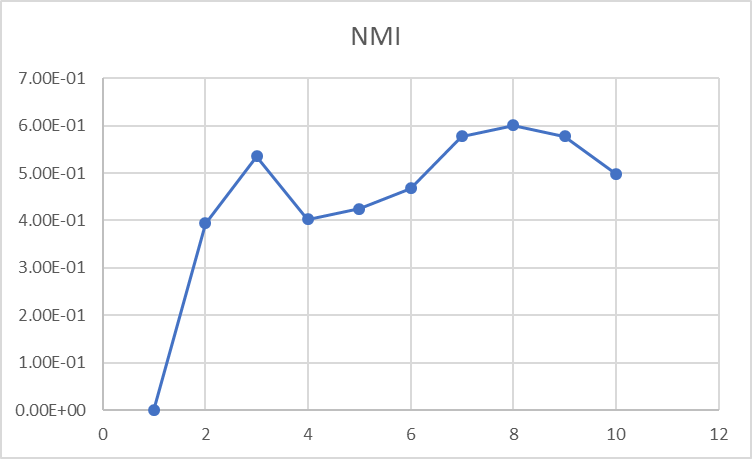
Actual: 6

GMM: 7

Soybean:

|  |  |  |
| --- | --- | --- |
| K | SSE | NMI |
| 1 | 1185.476 | 2.91E-06 |
| 2 | 1030.937 | 0.39445425 |
| 3 | 1061.294 | 0.53521485 |
| 4 | 1092.852 | 0.40220228 |
| 5 | 1111.128 | 0.42473204 |
| 6 | 1034.542 | 0.46749052 |
| 7 | 935.6541 | 0.57741373 |
| 8 | 922.0456 | 6.01E-01 |
| 9 | 861.8354 | 0.57796724 |
| 10 | 952.52 | 0.49885014 |
| 11 | 727.391 | 0.630197 |
| 12 | 742.0206 | 0.624491 |
| 13 | 673.2351 | 0.679448 |
| 14 | 673.6076 | 0.679415 |
| 15 | 665.5605 | 0.633045 |
| 16 | 652.9627 | 0.666755 |
| 17 | 636.5018 | 0.652768 |
| 18 | 636.0786 | 0.670519 |
| 19 | 631.8907 | 0.666061 |
| 20 | 619.7204 | 0.65062 |





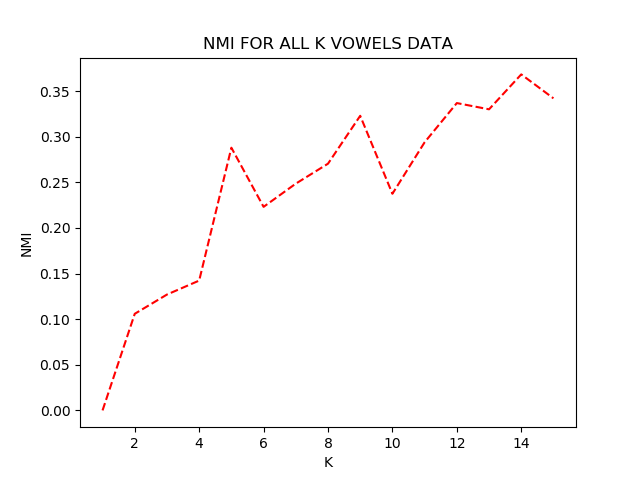
Ideal k:

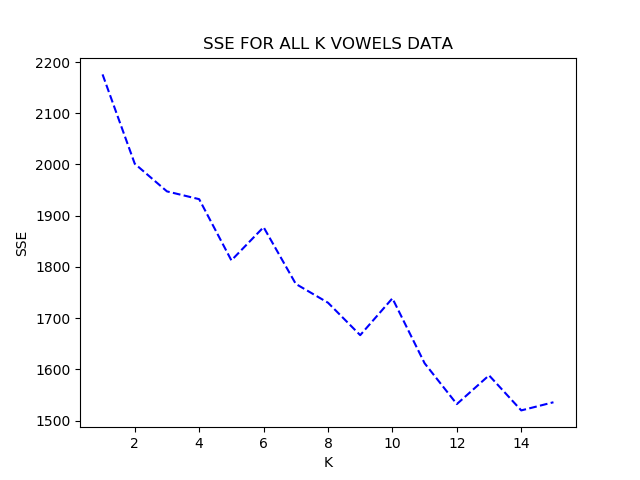
Actual: 15

GMM: 15

Vowels:

|  |  |  |
| --- | --- | --- |
| K | SSE | NMI |
| 1 | 2175.805 | 9.16E-06 |
| 2 | 2001.405 | 0.10588426 |
| 3 | 1947.429 | 0.12698033 |
| 4 | 1932.261 | 0.14226728 |
| 5 | 1812.78 | 0.28791081 |
| 6 | 1877.607 | 0.22309444 |
| 7 | 1766.834 | 0.24853037 |
| 8 | 1730.12 | 2.70E-01 |
| 9 | 1666.806 | 0.3228791 |
| 10 | 1738.705 | 0.23743486 |
| 11 | 1611.889 | 0.29386018 |
| 12 | 1532.268 | 0.33682778 |
| 13 | 1588.345 | 0.32999034 |
| 14 | 1519.935 | 0.36824954 |
| 15 | 1535.836 | 0.34203108 |





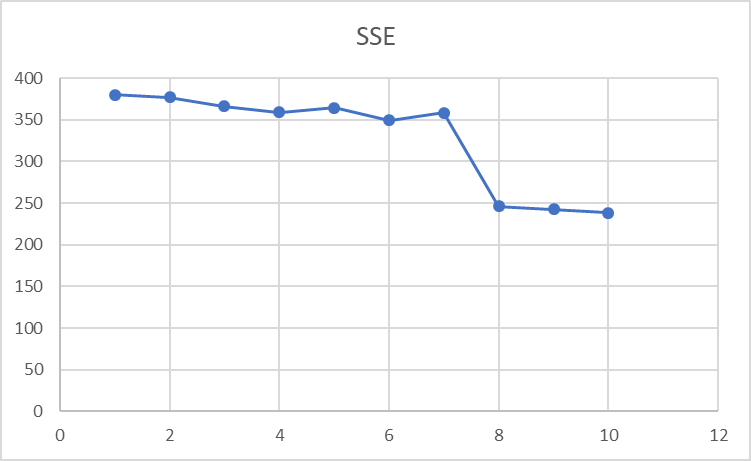
Ideal k:

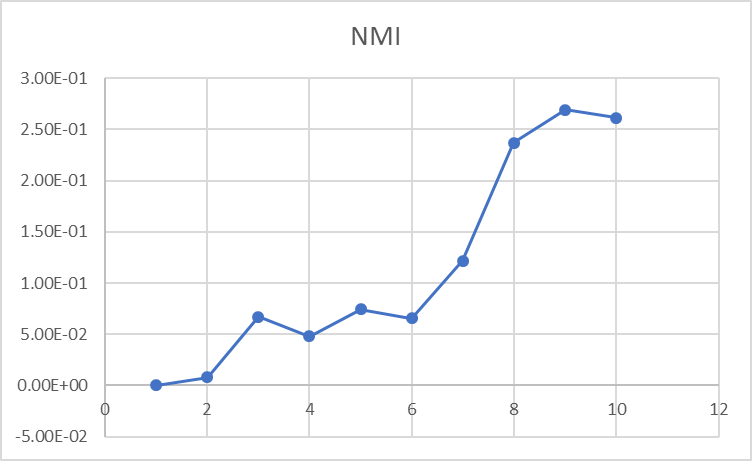
Actual: 11

GMM: 12

Yeast:

|  |  |  |
| --- | --- | --- |
| K | SSE | NMI |
| 1 | 380.2507 | -5.41E-06 |
| 2 | 377.2615 | 0.00806847 |
| 3 | 366.3949 | 0.06717392 |
| 4 | 358.7768 | 0.04798059 |
| 5 | 364.3685 | 0.07440909 |
| 6 | 349.5884 | 0.06580907 |
| 7 | 358.2323 | 0.12188477 |
| 8 | 246.1029 | 2.37E-01 |
| 9 | 242.6619 | 0.269087 |
| 10 | 238.1064 | 0.261163 |
| 11 | 227.7592 | 0.2698 |
| 12 | 225.7414 | 0.270948 |
| 13 | 230.6102 | 0.25985 |
| 14 | 217.8066 | 0.262554 |





Ideal k:

Actual: 9

K Means: 8

4.4)

4)

NMI is better. NMI uses actual class labels while SSE uses inter cluster distance.

5.1)

1. Dermatology: Kmeans

2. Vowels: GMM

3. Glass:. Kmeans

4. Ecoli: Kmeans

5. Yeast: GMM

6. Soybean: GMM

5.2) Yes clustering does provide insight about the separability of classes.