**Assignment3**

**Q1:**

1. The code is in the file *kmeans.py*
2. The code is in the file *singlelinkage.py*
3. Kmeans:

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster Number | size | Most common label | Percentage of points in cluster with that label |
| 0 | 159 | 1 | 64.1% |
| 1 | 81 | 2 | 96.2% |
| 2 | 141 | 3 | 46.8% |
| 3 | 70 | 6 | 77.1% |
| 4 | 74 | 0 | 83.7% |
| 5 | 104 | 4 | 25.0% |
| 6 | 61 | 6 | 85.2% |
| 7 | 117 | 4 | 34.1% |
| 8 | 130 | 9 | 37.6% |
| 9 | 63 | 0 | 38.0% |

Single linkage:

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster Number | size | Most common label | Percentage of points in cluster with that label |
| 0 | 159 | 1 | 64.1% |
| 1 | 81 | 2 | 96.2% |
| 2 | 141 | 3 | 46.8% |
| 3 | 70 | 6 | 77.1% |
| 4 | 74 | 0 | 83.7% |
| 5 | 104 | 4 | 25.0% |
| 6 | 61 | 6 | 85.2% |
| 7 | 117 | 4 | 34.1% |
| 8 | 130 | 9 | 37.6% |
| 9 | 63 | 0 | 38.0% |

d.

e. kmeans k=6:

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster Number | size | Most common label | Percentage of points in cluster with that label |
| 0 | 192 | 1 | 61.9% |
| 1 | 87 | 0 | 93.1% |
| 2 | 170 | 6 | 51.7% |
| 3 | 161 | 3 | 42.8% |
| 4 | 243 | 4 | 32.0% |
| 5 | 147 | 8 | 36.0% |