

## CSE 6363 - *Machine Learning*

Homework/Project 3- Spring 2019

Due Date: Apr. 30 2019, 11:59 pm

### Hierarchical Clustering

1. Consider an unlabeled version of our height/weight/age data set used in the previous assignments (and shown below).

$$D = \{ \begin{array}{ll} ((170, 57, 32), & W), \\ ((192, 95, 28), & M), \\ ((150, 45, 30), & W), \\ ((170, 65, 29), & M), \\ ((175, 78, 35), & M), \\ ((185, 90, 32), & M), \\ ((170, 65, 28), & W), \\ ((155, 48, 31), & W), \\ ((160, 55, 30), & W), \\ ((182, 80, 30), & M), \\ ((175, 69, 28), & W), \\ ((180, 80, 27), & M), \\ ((160, 50, 31), & W), \\ ((175, 72, 30), & M), \end{array} \}$$

- a) Apply hierarchical clustering with single (minimum) linkage to this data and show the resulting cluster hierarchy. Indicate the order of the merge operations and the distance (linkage) value between the merged sets at each merge. You can do this on the cluster hierarchy tree if you want.

To perform hierarchical clustering with single linkage we first have every point form its own cluster and compute the distance function between all the clusters. For simplicity we will index each data item and use it as a cluster number and when merging use the smallest index in the cluster as the index of the resulting cluster. Also, when merging only the distances from the new cluster have to be recomputed as the minimum of the distance to any of the merged clusters.

- i. Initial clusters and distance function:

## CSE 6363 - Machine Learning Homework/Project 3: Clustering and Semi-Supervised Learning

- Initial Clusters:

1 : {(170, 57, 32)}  
 2 : {(192, 95, 28)}  
 3 : {(150, 45, 30)}  
 4 : {(170, 65, 29)}  
 5 : {(175, 78, 35)}  
 6 : {(185, 90, 32)}  
 7 : {(170, 65, 28)}  
 8 : {(155, 48, 31)}  
 9 : {(160, 55, 30)}  
 10 : {(182, 80, 30)}  
 11 : {(175, 69, 28)}  
 12 : {(180, 80, 27)}  
 13 : {(160, 50, 31)}  
 14 : {(175, 72, 30)}

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 23.4094 | 8.5440  | 21.7945 | 36.2491 | 8.9443  | 17.5214 | 10.3923 | 26.0192 | 13.6015 | 25.5734 | 12.2474 | 15.9374 |
| 44.0908 |         | 65.3299 | 37.2156 | 25.0400 | 9.4868  | 37.2022 | 59.8916 | 51.2640 | 18.1384 | 31.0644 | 19.2354 | 55.2992 | 28.6705 |
| 23.4094 | 65.3299 |         | 28.3019 | 41.7013 | 57.0438 | 28.3549 | 5.9161  | 14.1421 | 47.4236 | 34.7131 | 46.1952 | 11.2250 | 36.7967 |
| 8.5440  | 37.2156 | 28.3019 |         | 15.1658 | 29.3087 | 1.0000  | 22.7596 | 14.1774 | 19.2354 | 6.4807  | 18.1384 | 18.1384 | 8.6603  |
| 21.7945 | 25.0400 | 41.7013 | 15.1658 |         | 15.9060 | 15.5885 | 36.2767 | 27.9106 | 8.8318  | 11.4018 | 9.6437  | 32.0156 | 7.8102  |
| 36.2491 | 9.4868  | 57.0438 | 29.3087 | 15.9060 |         | 29.4279 | 51.6236 | 43.0581 | 10.6301 | 23.6008 | 12.2474 | 47.1805 | 20.6882 |
| 8.9443  | 37.2022 | 28.3549 | 1.0000  | 15.5885 | 29.4279 |         | 22.8692 | 14.2829 | 19.3132 | 6.4031  | 18.0555 | 18.2757 | 8.8318  |
| 17.5214 | 59.8916 | 5.9161  | 22.7596 | 36.2767 | 51.6236 | 22.8692 |         | 8.6603  | 41.8808 | 29.1548 | 40.8044 | 5.3852  | 31.2570 |
| 10.3923 | 51.2640 | 14.1421 | 14.1774 | 27.9106 | 43.0581 | 14.2829 | 8.6603  |         | 33.3017 | 20.6155 | 32.1559 | 5.0990  | 22.6716 |
| 26.0192 | 18.1384 | 47.4236 | 19.2354 | 8.8318  | 10.6301 | 19.3132 | 41.8808 | 33.3017 |         | 13.1909 | 3.6056  | 37.2156 | 10.6301 |
| 13.6015 | 31.0644 | 34.7131 | 6.4807  | 11.4018 | 23.6008 | 6.4031  | 29.1548 | 20.6155 | 13.1909 |         | 12.1244 | 24.3926 | 3.6056  |
| 25.5734 | 19.2354 | 46.1952 | 18.1384 | 9.6437  | 12.2474 | 18.0555 | 40.8044 | 32.1559 | 3.6056  | 12.1244 |         | 36.2767 | 9.8995  |
| 12.2474 | 55.2992 | 11.2250 | 18.1384 | 32.0156 | 47.1805 | 18.2757 | 5.3852  | 5.0990  | 37.2156 | 24.3926 | 36.2767 |         | 26.6458 |
| 15.9374 | 28.6705 | 36.7967 | 8.6603  | 7.8102  | 20.6882 | 8.8318  | 31.2570 | 22.6716 | 10.6301 | 3.6056  | 9.8995  | 26.6458 |         |

### ii. Merge minimum distance clusters: 4, 7 (dist = 1.0)

- Initial Clusters:

4 : {(4, 7)}

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 23.4094 | 8.5440  | 21.7945 | 36.2491 | 17.5214 | 10.3923 | 26.0192 | 13.6015 | 25.5734 | 12.2474 | 15.9374 |
| 44.0908 |         | 65.3299 | 37.2022 | 25.0400 | 9.4868  | 59.8916 | 51.2640 | 18.1384 | 31.0644 | 19.2354 | 55.2992 | 28.6705 |
| 23.4094 | 65.3299 |         | 28.3019 | 41.7013 | 57.0438 | 5.9161  | 14.1421 | 47.4236 | 34.7131 | 46.1952 | 11.2250 | 36.7967 |
| 8.5440  | 37.2022 | 28.3019 |         | 15.1658 | 29.3087 | 22.7596 | 14.1774 | 19.2354 | 6.4031  | 18.0555 | 18.1384 | 8.6603  |
| 21.7945 | 25.0400 | 41.7013 | 15.1658 |         | 15.9060 | 36.2767 | 27.9106 | 8.8318  | 11.4018 | 9.6437  | 32.0156 | 7.8102  |
| 36.2491 | 9.4868  | 57.0438 | 29.3087 | 15.9060 |         | 51.6236 | 43.0581 | 10.6301 | 23.6008 | 12.2474 | 47.1805 | 20.6882 |
| 17.5214 | 59.8916 | 5.9161  | 22.7596 | 36.2767 | 51.6236 |         | 8.6603  | 41.8808 | 29.1548 | 40.8044 | 5.3852  | 31.2570 |
| 10.3923 | 51.2640 | 14.1421 | 14.1774 | 27.9106 | 43.0581 | 8.6603  |         | 33.3017 | 20.6155 | 32.1559 | 5.0990  | 22.6716 |
| 26.0192 | 18.1384 | 47.4236 | 19.2354 | 8.8318  | 10.6301 | 41.8808 | 33.3017 |         | 13.1909 | 3.6056  | 37.2156 | 10.6301 |
| 13.6015 | 31.0644 | 34.7131 | 6.4031  | 11.4018 | 23.6008 | 29.1548 | 20.6155 | 13.1909 |         | 12.1244 | 24.3926 | 3.6056  |
| 25.5734 | 19.2354 | 46.1952 | 18.0555 | 9.6437  | 12.2474 | 40.8044 | 32.1559 | 3.6056  | 12.1244 |         | 36.2767 | 9.8995  |
| 12.2474 | 55.2992 | 11.2250 | 18.1384 | 32.0156 | 47.1805 | 5.3852  | 5.0990  | 37.2156 | 24.3926 | 36.2767 |         | 26.6458 |
| 15.9374 | 28.6705 | 36.7967 | 8.6603  | 7.8102  | 20.6882 | 31.2570 | 22.6716 | 10.6301 | 3.6056  | 9.8995  | 26.6458 |         |

### iii. Merge minimum distance clusters: 10, 12 (dist = 3.6056)

- Initial Clusters:

10 : {10, 12}

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 23.4094 | 8.5440  | 21.7945 | 36.2491 | 17.5214 | 10.3923 | 25.5734 | 13.6015 | 12.2474 | 15.9374 |
| 44.0908 |         | 65.3299 | 37.2022 | 25.0400 | 9.4868  | 59.8916 | 51.2640 | 18.1384 | 31.0644 | 55.2992 | 28.6705 |
| 23.4094 | 65.3299 |         | 28.3019 | 41.7013 | 57.0438 | 5.9161  | 14.1421 | 46.1952 | 34.7131 | 11.2250 | 36.7967 |
| 8.5440  | 37.2022 | 28.3019 |         | 15.1658 | 29.3087 | 22.7596 | 14.1774 | 18.0555 | 6.4031  | 18.1384 | 8.6603  |
| 21.7945 | 25.0400 | 41.7013 | 15.1658 |         | 15.9060 | 36.2767 | 27.9106 | 8.8318  | 11.4018 | 32.0156 | 7.8102  |
| 36.2491 | 9.4868  | 57.0438 | 29.3087 | 15.9060 |         | 51.6236 | 43.0581 | 10.6301 | 23.6008 | 47.1805 | 20.6882 |
| 17.5214 | 59.8916 | 5.9161  | 22.7596 | 36.2767 | 51.6236 |         | 8.6603  | 40.8044 | 29.1548 | 5.3852  | 31.2570 |
| 10.3923 | 51.2640 | 14.1421 | 14.1774 | 27.9106 | 43.0581 | 8.6603  |         | 32.1559 | 20.6155 | 5.0990  | 22.6716 |
| 25.5734 | 18.1384 | 46.1952 | 18.0555 | 8.8318  | 10.6301 | 40.8044 | 32.1559 |         | 12.1244 | 36.2767 | 9.8995  |
| 13.6015 | 31.0644 | 34.7131 | 6.4031  | 11.4018 | 23.6008 | 29.1548 | 20.6155 | 12.1244 |         | 24.3926 | 3.6056  |
| 12.2474 | 55.2992 | 11.2250 | 18.1384 | 32.0156 | 47.1805 | 5.3852  | 5.0990  | 36.2767 | 24.3926 |         | 26.6458 |
| 15.9374 | 28.6705 | 36.7967 | 8.6603  | 7.8102  | 20.6882 | 31.2570 | 22.6716 | 9.8995  | 3.6056  | 26.6458 |         |

## iv. Merge minimum distance clusters: 11, 14 (dist = 3.6056)

- Initial Clusters:

11 : {11, 14}

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 23.4094 | 8.5440  | 21.7945 | 36.2491 | 17.5214 | 10.3923 | 25.5734 | 13.6015 | 12.2474 |
| 44.0908 |         | 65.3299 | 37.2022 | 25.0400 | 9.4868  | 59.8916 | 51.2640 | 18.1384 | 28.6705 | 55.2992 |
| 23.4094 | 65.3299 |         | 28.3019 | 41.7013 | 57.0438 | 5.9161  | 14.1421 | 46.1952 | 34.7131 | 11.2250 |
| 8.5440  | 37.2022 | 28.3019 |         | 15.1658 | 29.3087 | 22.7596 | 14.1774 | 18.0555 | 6.4031  | 18.1384 |
| 21.7945 | 25.0400 | 41.7013 | 15.1658 |         | 15.9060 | 36.2767 | 27.9106 | 8.8318  | 7.8102  | 32.0156 |
| 36.2491 | 9.4868  | 57.0438 | 29.3087 | 15.9060 |         | 51.6236 | 43.0581 | 10.6301 | 20.6882 | 47.1805 |
| 17.5214 | 59.8916 | 5.9161  | 22.7596 | 36.2767 | 51.6236 |         | 8.6603  | 40.8044 | 29.1548 | 5.3852  |
| 10.3923 | 51.2640 | 14.1421 | 14.1774 | 27.9106 | 43.0581 | 8.6603  |         | 32.1559 | 20.6155 | 5.0990  |
| 25.5734 | 18.1384 | 46.1952 | 18.0555 | 8.8318  | 10.6301 | 40.8044 | 32.1559 |         | 9.8995  | 36.2767 |
| 13.6015 | 28.6705 | 34.7131 | 6.4031  | 7.8102  | 20.6882 | 29.1548 | 20.6155 | 9.8995  |         | 24.3926 |
| 12.2474 | 55.2992 | 11.2250 | 18.1384 | 32.0156 | 47.1805 | 5.3852  | 5.0990  | 36.2767 | 24.3926 |         |

## v. Merge minimum distance clusters: 9, 13 (dist = 5.099)

- Initial Clusters:

9 : {9, 13}

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 23.4094 | 8.5440  | 21.7945 | 36.2491 | 17.5214 | 10.3923 | 25.5734 | 13.6015 |
| 44.0908 |         | 65.3299 | 37.2022 | 25.0400 | 9.4868  | 59.8916 | 51.2640 | 18.1384 | 28.6705 |
| 23.4094 | 65.3299 |         | 28.3019 | 41.7013 | 57.0438 | 5.9161  | 11.2250 | 46.1952 | 34.7131 |
| 8.5440  | 37.2022 | 28.3019 |         | 15.1658 | 29.3087 | 22.7596 | 14.1774 | 18.0555 | 6.4031  |
| 21.7945 | 25.0400 | 41.7013 | 15.1658 |         | 15.9060 | 36.2767 | 27.9106 | 8.8318  | 7.8102  |
| 36.2491 | 9.4868  | 57.0438 | 29.3087 | 15.9060 |         | 51.6236 | 43.0581 | 10.6301 | 20.6882 |
| 17.5214 | 59.8916 | 5.9161  | 22.7596 | 36.2767 | 51.6236 |         | 5.3852  | 40.8044 | 29.1548 |
| 10.3923 | 51.2640 | 11.2250 | 14.1774 | 27.9106 | 43.0581 | 5.3852  |         | 32.1559 | 20.6155 |
| 25.5734 | 18.1384 | 46.1952 | 18.0555 | 8.8318  | 10.6301 | 40.8044 | 32.1559 |         | 9.8995  |
| 13.6015 | 28.6705 | 34.7131 | 6.4031  | 7.8102  | 20.6882 | 29.1548 | 20.6155 | 9.8995  |         |

## vi. Merge minimum distance clusters: 8, 9 (dist = 5.3852)

- Initial Clusters:

8 : {8, 9}

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 23.4094 | 8.5440  | 21.7945 | 36.2491 | 10.3923 | 25.5734 | 13.6015 |
| 44.0908 |         | 65.3299 | 37.2022 | 25.0400 | 9.4868  | 51.2640 | 18.1384 | 28.6705 |
| 23.4094 | 65.3299 |         | 28.3019 | 41.7013 | 57.0438 | 5.9161  | 46.1952 | 34.7131 |
| 8.5440  | 37.2022 | 28.3019 |         | 15.1658 | 29.3087 | 14.1774 | 18.0555 | 6.4031  |
| 21.7945 | 25.0400 | 41.7013 | 15.1658 |         | 15.9060 | 27.9106 | 8.8318  | 7.8102  |
| 36.2491 | 9.4868  | 57.0438 | 29.3087 | 15.9060 |         | 43.0581 | 10.6301 | 20.6882 |
| 10.3923 | 51.2640 | 5.9161  | 14.1774 | 27.9106 | 43.0581 |         | 32.1559 | 20.6155 |
| 25.5734 | 18.1384 | 46.1952 | 18.0555 | 8.8318  | 10.6301 | 32.1559 |         | 9.8995  |
| 13.6015 | 28.6705 | 34.7131 | 6.4031  | 7.8102  | 20.6882 | 20.6155 | 9.8995  |         |

## vii. Merge minimum distance clusters: 3, 8 (dist = 5.9161)

- Initial Clusters:

3 : {3, 8}

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 10.3923 | 8.5440  | 21.7945 | 36.2491 | 25.5734 | 13.6015 |
| 44.0908 |         | 51.2640 | 37.2022 | 25.0400 | 9.4868  | 18.1384 | 28.6705 |
| 10.3923 | 51.2640 |         | 14.1774 | 27.9106 | 43.0581 | 32.1559 | 20.6155 |
| 8.5440  | 37.2022 | 14.1774 |         | 15.1658 | 29.3087 | 18.0555 | 6.4031  |
| 21.7945 | 25.0400 | 27.9106 | 15.1658 |         | 15.9060 | 8.8318  | 7.8102  |
| 36.2491 | 9.4868  | 43.0581 | 29.3087 | 15.9060 |         | 10.6301 | 20.6882 |
| 25.5734 | 18.1384 | 32.1559 | 18.0555 | 8.8318  | 10.6301 |         | 9.8995  |
| 13.6015 | 28.6705 | 20.6155 | 6.4031  | 7.8102  | 20.6882 | 9.8995  |         |

## viii. Merge minimum distance clusters: 4, 11 (dist = 6.4031)

- Initial Clusters:

4 : {4, 11}

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 10.3923 | 8.5440  | 21.7945 | 36.2491 | 25.5734 |
| 44.0908 |         | 51.2640 | 28.6705 | 25.0400 | 9.4868  | 18.1384 |
| 10.3923 | 51.2640 |         | 14.1774 | 27.9106 | 43.0581 | 32.1559 |
| 8.5440  | 28.6705 | 14.1774 |         | 7.8102  | 20.6882 | 9.8995  |
| 21.7945 | 25.0400 | 27.9106 | 7.8102  |         | 15.9060 | 8.8318  |
| 36.2491 | 9.4868  | 43.0581 | 20.6882 | 15.9060 |         | 10.6301 |
| 25.5734 | 18.1384 | 32.1559 | 9.8995  | 8.8318  | 10.6301 |         |

## ix. Merge minimum distance clusters: 4, 5 (dist = 7.8102)

- Initial Clusters:

4 : {4, 5}

- Cluster Distance Matrix:

|         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 10.3923 | 8.5440  | 36.2491 | 25.5734 |
| 44.0908 |         | 51.2640 | 25.0400 | 9.4868  | 18.1384 |
| 10.3923 | 51.2640 |         | 14.1774 | 43.0581 | 32.1559 |
| 8.5440  | 25.0400 | 14.1774 |         | 15.9060 | 8.8318  |
| 36.2491 | 9.4868  | 43.0581 | 15.9060 |         | 10.6301 |
| 25.5734 | 18.1384 | 32.1559 | 8.8318  | 10.6301 |         |

## x. Merge minimum distance clusters: 1, 4 (dist = 8.5440)

- Initial Clusters:

1 : {1, 4}

- Cluster Distance Matrix:

|         |         |         |         |         |
|---------|---------|---------|---------|---------|
|         | 25.0400 | 10.3923 | 15.9060 | 8.8318  |
| 25.0400 |         | 51.2640 | 9.4868  | 18.1384 |
| 10.3923 | 51.2640 |         | 43.0581 | 32.1559 |
|         |         |         |         |         |
| 15.9060 | 9.4868  | 43.0581 |         | 10.6301 |
|         |         |         |         |         |
| 8.8318  | 18.1384 | 32.1559 | 10.6301 |         |

xi. Merge minimum distance clusters: 1, 10 (dist = 8.8318)

- Initial Clusters:

|     |         |
|-----|---------|
| 1 : | {1, 10} |
|-----|---------|
- Cluster Distance Matrix:

|         |         |         |         |
|---------|---------|---------|---------|
|         | 18.1384 | 10.3923 | 10.6301 |
| 18.1384 |         | 51.2640 | 9.4868  |
| 10.3923 | 51.2640 |         | 43.0581 |
|         |         |         |         |
| 10.6301 | 9.4868  | 43.0581 |         |

xii. Merge minimum distance clusters: 2, 6 (dist = 9.4868)

- Initial Clusters:

|     |        |
|-----|--------|
| 2 : | {2, 6} |
|-----|--------|
- Cluster Distance Matrix:

|         |         |         |
|---------|---------|---------|
|         | 10.6301 | 10.3923 |
| 10.6301 |         | 43.0581 |
| 10.3923 | 43.0581 |         |

xiii. Merge minimum distance clusters: 1, 3 (dist = 10.3923)

- Initial Clusters:

|     |        |
|-----|--------|
| 1 : | {1, 3} |
|-----|--------|

- Cluster Distance Matrix:

10.6301  
10.6301

xiv. Merge minimum distance clusters: 1, 2 (dist = 10.6301)

- Initial Clusters:

1 : {1, 2}

b) Repeat the clustering using complete (maximum) linkage. Again, make sure you indicate the linkage value for every cluster merge.

To perform hierarchical clustering with complete linkage we first have every point form its own cluster and compute the distance function between all the clusters. For simplicity we will index each data item and use it as a cluster number and when merging use the smallest index in the cluster as the index of the resulting cluster. Also, when merging only the distances from the new cluster have to be recomputed as the maximum of the distance to any of the merged clusters. .

i. Initial clusters and distance function:

- Initial Clusters:

1 : {(170, 57, 32)}  
2 : {(192, 95, 28)}  
3 : {(150, 45, 30)}  
4 : {(170, 65, 29)}  
5 : {(175, 78, 35)}  
6 : {(185, 90, 32)}  
7 : {(170, 65, 28)}  
8 : {(155, 48, 31)}  
9 : {(160, 55, 30)}  
10 : {(182, 80, 30)}  
11 : {(175, 69, 28)}  
12 : {(180, 80, 27)}  
13 : {(160, 50, 31)}  
14 : {(175, 72, 30)}

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 23.4094 | 8.5440  | 21.7945 | 36.2491 | 8.9443  | 17.5214 | 10.3923 | 26.0192 | 13.6015 | 25.5734 | 12.2474 | 15.9374 |
| 44.0908 |         | 65.3299 | 37.2156 | 25.0400 | 9.4868  | 37.2022 | 59.8916 | 51.2640 | 18.1384 | 31.0644 | 19.2354 | 55.2992 | 28.6705 |
| 23.4094 | 65.3299 |         | 28.3019 | 41.7013 | 57.0438 | 28.3549 | 5.9161  | 14.1421 | 47.4236 | 34.7131 | 46.1952 | 11.2250 | 36.7967 |
| 8.5440  | 37.2156 | 28.3019 |         | 15.1658 | 29.3087 | 1.0000  | 22.7596 | 14.1774 | 19.2354 | 6.4807  | 18.1384 | 18.1384 | 8.6603  |
| 21.7945 | 25.0400 | 41.7013 | 15.1658 |         | 15.9060 | 15.5885 | 36.2767 | 27.9106 | 8.8318  | 11.4018 | 9.6437  | 32.0156 | 7.8102  |
| 36.2491 | 9.4868  | 57.0438 | 29.3087 | 15.9060 |         | 29.4279 | 51.6236 | 43.0581 | 10.6301 | 23.6008 | 12.2474 | 47.1805 | 20.6882 |
| 8.9443  | 37.2022 | 28.3549 | 1.0000  | 15.5885 | 29.4279 |         | 22.8692 | 14.2829 | 19.3132 | 6.4031  | 18.0555 | 18.2757 | 8.8318  |
| 17.5214 | 59.8916 | 5.9161  | 22.7596 | 36.2767 | 51.6236 | 22.8692 |         | 8.6603  | 41.8808 | 29.1548 | 40.8044 | 5.3852  | 31.2570 |
| 10.3923 | 51.2640 | 14.1421 | 14.1774 | 27.9106 | 43.0581 | 14.2829 | 8.6603  |         | 33.3017 | 20.6155 | 32.1559 | 5.0990  | 22.6716 |
| 26.0192 | 18.1384 | 47.4236 | 19.2354 | 8.8318  | 10.6301 | 19.3132 | 41.8808 | 33.3017 |         | 13.1909 | 3.6056  | 37.2156 | 10.6301 |
| 13.6015 | 31.0644 | 34.7131 | 6.4807  | 11.4018 | 23.6008 | 6.4031  | 29.1548 | 20.6155 | 13.1909 |         | 12.1244 | 24.3926 | 3.6056  |
| 25.5734 | 19.2354 | 46.1952 | 18.1384 | 9.6437  | 12.2474 | 18.0555 | 40.8044 | 32.1559 | 3.6056  | 12.1244 |         | 36.2767 | 9.8995  |
| 12.2474 | 55.2992 | 11.2250 | 18.1384 | 32.0156 | 47.1805 | 18.2757 | 5.3852  | 5.0990  | 37.2156 | 24.3926 | 36.2767 |         | 26.6458 |
| 15.9374 | 28.6705 | 36.7967 | 8.6603  | 7.8102  | 20.6882 | 8.8318  | 31.2570 | 22.6716 | 10.6301 | 3.6056  | 9.8995  | 26.6458 |         |

ii. Merge minimum distance clusters: 4, 7 (dist = 1.0)

- Initial Clusters:

4 : {(4, 7)}

## CSE 6363 - Machine Learning Homework/Project 3: Clustering and Semi-Supervised Learning

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 23.4094 | 8.9443  | 21.7945 | 36.2491 | 17.5214 | 10.3923 | 26.0192 | 13.6015 | 25.5734 | 12.2474 | 15.9374 |
| 44.0908 |         | 65.3299 | 37.2156 | 25.0400 | 9.4868  | 59.8916 | 51.2640 | 18.1384 | 31.0644 | 19.2354 | 55.2992 | 28.6705 |
| 23.4094 | 65.3299 |         | 28.3549 | 41.7013 | 57.0438 | 5.9161  | 14.1421 | 47.4236 | 34.7131 | 46.1952 | 11.2250 | 36.7967 |
| 8.9443  | 37.2156 | 28.3549 |         | 15.5885 | 29.4279 | 22.8692 | 14.2829 | 19.3132 | 6.4807  | 18.1384 | 18.2757 | 8.8318  |
| 21.7945 | 25.0400 | 41.7013 | 15.5885 |         | 15.9060 | 36.2767 | 27.9106 | 8.8318  | 11.4018 | 9.6437  | 32.0156 | 7.8102  |
| 36.2491 | 9.4868  | 57.0438 | 29.4279 | 15.9060 |         | 51.6236 | 43.0581 | 10.6301 | 23.6008 | 12.2474 | 47.1805 | 20.6882 |
| 17.5214 | 59.8916 | 5.9161  | 22.8692 | 36.2767 | 51.6236 |         | 8.6603  | 41.8808 | 29.1548 | 40.8044 | 5.3852  | 31.2570 |
| 10.3923 | 51.2640 | 14.1421 | 14.2829 | 27.9106 | 43.0581 | 8.6603  |         | 33.3017 | 20.6155 | 32.1559 | 5.0990  | 22.6716 |
| 26.0192 | 18.1384 | 47.4236 | 19.3132 | 8.8318  | 10.6301 | 41.8808 | 33.3017 |         | 13.1909 | 3.6056  | 37.2156 | 10.6301 |
| 13.6015 | 31.0644 | 34.7131 | 6.4807  | 11.4018 | 23.6008 | 29.1548 | 20.6155 | 13.1909 |         | 12.1244 | 24.3926 | 3.6056  |
| 25.5734 | 19.2354 | 46.1952 | 18.1384 | 9.6437  | 12.2474 | 40.8044 | 32.1559 | 3.6056  | 12.1244 |         | 36.2767 | 9.8995  |
| 12.2474 | 55.2992 | 11.2250 | 18.2757 | 32.0156 | 47.1805 | 5.3852  | 5.0990  | 37.2156 | 24.3926 | 36.2767 |         | 26.6458 |
| 15.9374 | 28.6705 | 36.7967 | 8.8318  | 7.8102  | 20.6882 | 31.2570 | 22.6716 | 10.6301 | 3.6056  | 9.8995  | 26.6458 |         |

### iii. Merge minimum distance clusters: 10, 12 (dist = 3.6056)

- Initial Clusters:

10 : {(10, 12)}

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 23.4094 | 8.9443  | 21.7945 | 36.2491 | 17.5214 | 10.3923 | 26.0192 | 13.6015 | 12.2474 | 15.9374 |
| 44.0908 |         | 65.3299 | 37.2156 | 25.0400 | 9.4868  | 59.8916 | 51.2640 | 19.2354 | 31.0644 | 55.2992 | 28.6705 |
| 23.4094 | 65.3299 |         | 28.3549 | 41.7013 | 57.0438 | 5.9161  | 14.1421 | 47.4236 | 34.7131 | 11.2250 | 36.7967 |
| 8.9443  | 37.2156 | 28.3549 |         | 15.5885 | 29.4279 | 22.8692 | 14.2829 | 19.3132 | 6.4807  | 18.2757 | 8.8318  |
| 21.7945 | 25.0400 | 41.7013 | 15.5885 |         | 15.9060 | 36.2767 | 27.9106 | 9.6437  | 11.4018 | 32.0156 | 7.8102  |
| 36.2491 | 9.4868  | 57.0438 | 29.4279 | 15.9060 |         | 51.6236 | 43.0581 | 12.2474 | 23.6008 | 47.1805 | 20.6882 |
| 17.5214 | 59.8916 | 5.9161  | 22.8692 | 36.2767 | 51.6236 |         | 8.6603  | 41.8808 | 29.1548 | 5.3852  | 31.2570 |
| 10.3923 | 51.2640 | 14.1421 | 14.2829 | 27.9106 | 43.0581 | 8.6603  |         | 33.3017 | 20.6155 | 5.0990  | 22.6716 |
| 26.0192 | 19.2354 | 47.4236 | 19.3132 | 9.6437  | 12.2474 | 41.8808 | 33.3017 |         | 13.1909 | 37.2156 | 10.6301 |
| 13.6015 | 31.0644 | 34.7131 | 6.4807  | 11.4018 | 23.6008 | 29.1548 | 20.6155 | 13.1909 |         | 24.3926 | 3.6056  |
| 12.2474 | 55.2992 | 11.2250 | 18.2757 | 32.0156 | 47.1805 | 5.3852  | 5.0990  | 37.2156 | 24.3926 |         | 26.6458 |
| 15.9374 | 28.6705 | 36.7967 | 8.8318  | 7.8102  | 20.6882 | 31.2570 | 22.6716 | 10.6301 | 3.6056  | 26.6458 |         |

### iv. Merge minimum distance clusters: 11, 14 (dist = 3.6056)

- Initial Clusters:

11 : {(11, 14)}

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 23.4094 | 8.9443  | 21.7945 | 36.2491 | 17.5214 | 10.3923 | 26.0192 | 15.9374 | 12.2474 |
| 44.0908 |         | 65.3299 | 37.2156 | 25.0400 | 9.4868  | 59.8916 | 51.2640 | 19.2354 | 31.0644 | 55.2992 |
| 23.4094 | 65.3299 |         | 28.3549 | 41.7013 | 57.0438 | 5.9161  | 14.1421 | 47.4236 | 36.7967 | 11.2250 |
| 8.9443  | 37.2156 | 28.3549 |         | 15.5885 | 29.4279 | 22.8692 | 14.2829 | 19.3132 | 8.8318  | 18.2757 |
| 21.7945 | 25.0400 | 41.7013 | 15.5885 |         | 15.9060 | 36.2767 | 27.9106 | 9.6437  | 11.4018 | 32.0156 |
| 36.2491 | 9.4868  | 57.0438 | 29.4279 | 15.9060 |         | 51.6236 | 43.0581 | 12.2474 | 23.6008 | 47.1805 |
| 17.5214 | 59.8916 | 5.9161  | 22.8692 | 36.2767 | 51.6236 |         | 8.6603  | 41.8808 | 31.2570 | 5.3852  |
| 10.3923 | 51.2640 | 14.1421 | 14.2829 | 27.9106 | 43.0581 | 8.6603  |         | 33.3017 | 22.6716 | 5.0990  |
| 26.0192 | 19.2354 | 47.4236 | 19.3132 | 9.6437  | 12.2474 | 41.8808 | 33.3017 |         | 13.1909 | 37.2156 |
| 15.9374 | 31.0644 | 36.7967 | 8.8318  | 11.4018 | 23.6008 | 31.2570 | 22.6716 | 13.1909 |         | 26.6458 |
| 12.2474 | 55.2992 | 11.2250 | 18.2757 | 32.0156 | 47.1805 | 5.3852  | 5.0990  | 37.2156 | 26.6458 |         |

### v. Merge minimum distance clusters: 9, 13 (dist = 5.0990)

- Initial Clusters:

9 : {(9, 13)}

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 23.4094 | 8.9443  | 21.7945 | 36.2491 | 17.5214 | 12.2474 | 26.0192 | 15.9374 |
| 44.0908 |         | 65.3299 | 37.2156 | 25.0400 | 9.4868  | 59.8916 | 55.2992 | 19.2354 | 31.0644 |
| 23.4094 | 65.3299 |         | 28.3549 | 41.7013 | 57.0438 | 5.9161  | 14.1421 | 47.4236 | 36.7967 |
| 8.9443  | 37.2156 | 28.3549 |         | 15.5885 | 29.4279 | 22.8692 | 18.2757 | 19.3132 | 8.8318  |
| 21.7945 | 25.0400 | 41.7013 | 15.5885 |         | 15.9060 | 36.2767 | 32.0156 | 9.6437  | 11.4018 |
| 36.2491 | 9.4868  | 57.0438 | 29.4279 | 15.9060 |         | 51.6236 | 47.1805 | 12.2474 | 23.6008 |
| 17.5214 | 59.8916 | 5.9161  | 22.8692 | 36.2767 | 51.6236 |         | 8.6603  | 41.8808 | 31.2570 |
| 12.2474 | 55.2992 | 14.1421 | 18.2757 | 32.0156 | 47.1805 | 8.6603  |         | 37.2156 | 26.6458 |
| 26.0192 | 19.2354 | 47.4236 | 19.3132 | 9.6437  | 12.2474 | 41.8808 | 37.2156 |         | 13.1909 |
| 15.9374 | 31.0644 | 36.7967 | 8.8318  | 11.4018 | 23.6008 | 31.2570 | 26.6458 | 13.1909 |         |

## vi. Merge minimum distance clusters: 3, 8 (dist = 5.9161)

- Initial Clusters:

3 : {(3, 8)}

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 23.4094 | 8.9443  | 21.7945 | 36.2491 | 12.2474 | 26.0192 | 15.9374 |
| 44.0908 |         | 65.3299 | 37.2156 | 25.0400 | 9.4868  | 55.2992 | 19.2354 | 31.0644 |
| 23.4094 | 65.3299 |         | 28.3549 | 41.7013 | 57.0438 | 14.1421 | 47.4236 | 36.7967 |
| 8.9443  | 37.2156 | 28.3549 |         | 15.5885 | 29.4279 | 18.2757 | 19.3132 | 8.8318  |
| 21.7945 | 25.0400 | 41.7013 | 15.5885 |         | 15.9060 | 32.0156 | 9.6437  | 11.4018 |
| 36.2491 | 9.4868  | 57.0438 | 29.4279 | 15.9060 |         | 47.1805 | 12.2474 | 23.6008 |
| 12.2474 | 55.2992 | 14.1421 | 18.2757 | 32.0156 | 47.1805 |         | 37.2156 | 26.6458 |
| 26.0192 | 19.2354 | 47.4236 | 19.3132 | 9.6437  | 12.2474 | 37.2156 |         | 13.1909 |
| 15.9374 | 31.0644 | 36.7967 | 8.8318  | 11.4018 | 23.6008 | 26.6458 | 13.1909 |         |

## vii. Merge minimum distance clusters: 4, 11 (dist = 8.8318)

- Initial Clusters:

4 : {(4, 11)}

- Cluster Distance Matrix:

|         |         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 23.4094 | 15.9374 | 21.7945 | 36.2491 | 12.2474 | 26.0192 |
| 44.0908 |         | 65.3299 | 37.2156 | 25.0400 | 9.4868  | 55.2992 | 19.2354 |
| 23.4094 | 65.3299 |         | 36.7967 | 41.7013 | 57.0438 | 14.1421 | 47.4236 |
| 15.9374 | 37.2156 | 36.7967 |         | 15.5885 | 29.4279 | 26.6458 | 19.3132 |
| 21.7945 | 25.0400 | 41.7013 | 15.5885 |         | 15.9060 | 32.0156 | 9.6437  |
| 36.2491 | 9.4868  | 57.0438 | 29.4279 | 15.9060 |         | 47.1805 | 12.2474 |
| 12.2474 | 55.2992 | 14.1421 | 26.6458 | 32.0156 | 47.1805 |         | 37.2156 |
| 26.0192 | 19.2354 | 47.4236 | 19.3132 | 9.6437  | 12.2474 | 37.2156 |         |

## viii. Merge minimum distance clusters: 2, 6 (dist = 9.4868)

- Initial Clusters:

2 : {(2, 6)}



- | ● Cluster Distance Matrix: |         |         |         |         |         |         |
|----------------------------|---------|---------|---------|---------|---------|---------|
|                            | 44.0908 | 23.4094 | 15.9374 | 21.7945 | 12.2474 | 26.0192 |
| 44.0908                    |         | 65.3299 | 37.2156 | 25.0400 | 55.2992 | 19.2354 |
| 23.4094                    | 65.3299 |         | 36.7967 | 41.7013 | 14.1421 | 47.4236 |
| 15.9374                    | 37.2156 | 36.7967 |         | 15.5885 | 26.6458 | 19.3132 |
| 21.7945                    | 25.0400 | 41.7013 | 15.5885 |         | 32.0156 | 9.6437  |
|                            |         |         |         |         |         |         |
| 12.2474                    | 55.2992 | 14.1421 | 26.6458 | 32.0156 |         | 37.2156 |
| 26.0192                    | 19.2354 | 47.4236 | 19.3132 | 9.6437  | 37.2156 |         |

ix. Merge minimum distance clusters: 5, 10 (dist = 9.6437)

- Initial Clusters:
 

$5 : \{(5, 10)\}$
- Cluster Distance Matrix:
 

|         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|
|         | 44.0908 | 23.4094 | 15.9374 | 26.0192 | 12.2474 |
| 44.0908 |         | 65.3299 | 37.2156 | 25.0400 | 55.2992 |
| 23.4094 | 65.3299 |         | 36.7967 | 47.4236 | 14.1421 |
| 15.9374 | 37.2156 | 36.7967 |         | 19.3132 | 26.6458 |
| 26.0192 | 25.0400 | 47.4236 | 19.3132 |         | 37.2156 |
| 12.2474 | 55.2992 | 14.1421 | 26.6458 | 37.2156 |         |

x. Merge minimum distance clusters: 1, 9 (dist = 12.2474)

- Initial Clusters:
 

$1: \{(1, 9)\}$
- Cluster Distance Matrix:
 

|         |         |         |         |         |
|---------|---------|---------|---------|---------|
|         | 55.2992 | 23.4094 | 26.6458 | 37.2156 |
| 55.2992 |         | 65.3299 | 37.2156 | 25.0400 |
| 23.4094 | 65.3299 |         | 36.7967 | 47.4236 |
| 26.6458 | 37.2156 | 36.7967 |         | 19.3132 |
| 37.2156 | 25.0400 | 47.4236 | 19.3132 |         |

xi. Merge minimum distance clusters: 4, 5 (dist = 19.3132)

- Initial Clusters:
  - 4 :  $\{(4, 5)\}$

- Cluster Distance Matrix:
 

|         |         |         |         |
|---------|---------|---------|---------|
|         | 55.2992 | 23.4094 | 37.2156 |
| 55.2992 |         | 65.3299 | 37.2156 |
| 23.4094 | 65.3299 |         | 47.4236 |
| 37.2156 | 37.2156 | 47.4236 |         |

xii. Merge minimum distance clusters: 1, 3 (dist = 23.4094)

- Initial Clusters:
 

|     |          |
|-----|----------|
| 1 : | {(1, 3)} |
|-----|----------|
- Cluster Distance Matrix:
 

|         |         |         |
|---------|---------|---------|
|         | 65.3299 | 47.4236 |
| 65.3299 |         | 37.2156 |
| 47.4236 | 37.2156 |         |

xiii. Merge minimum distance clusters: 2, 4 (dist = 37.2156)

- Initial Clusters:
 

|     |          |
|-----|----------|
| 2 : | {(2, 4)} |
|-----|----------|
- Cluster Distance Matrix:
 

|         |         |
|---------|---------|
|         | 65.3299 |
| 65.3299 |         |

xiv. Merge minimum distance clusters: 1, 2 (dist = 65.3299)

- Initial Clusters:
 

|     |          |
|-----|----------|
| 1 : | {(1, 3)} |
|-----|----------|

## Self-Training

- Consider the following linearly separable training data set:

$$D_s = \{ \begin{array}{l} ((170, 57, 32), \text{ } W), \\ ((190, 95, 28), \text{ } M), \\ ((150, 45, 35), \text{ } W), \\ ((168, 65, 29), \text{ } M), \\ ((175, 78, 26), \text{ } M), \\ ((185, 90, 32), \text{ } M), \\ ((171, 65, 28), \text{ } W), \\ ((155, 48, 31), \text{ } W), \\ ((165, 60, 27), \text{ } W) \end{array} \}$$

$$D_u = \{ \begin{array}{lll} (182, 80, 30), & (175, 69, 28), & (178, 80, 27), \\ (160, 50, 31), & (170, 72, 30), & (152, 45, 29), \\ (177, 79, 28), & (171, 62, 27), & (185, 90, 30), \\ (181, 83, 28), & (168, 59, 24), & (158, 45, 28), \\ (178, 82, 28), & (165, 55, 30), & (162, 58, 28), \\ (180, 80, 29), & (173, 75, 28), & (172, 65, 27), \\ (160, 51, 29), & (178, 77, 28), & (182, 84, 27), \\ (175, 67, 28), & (163, 50, 27), & (177, 80, 30), \\ (170, 65, 28) \end{array} \}$$

- a) Implement a self-training system using a logistic regression classifier for this problem.
- b) Learn a classifier using the semi-supervised learning algorithm and compare it against a classifier learned only from the labeled data  $D_s$  using the following test set:

$$D_t = \{ \begin{array}{l} ((169, 58, 30), \text{ } W), \\ ((185, 90, 29), \text{ } M), \\ ((148, 40, 31), \text{ } W), \\ ((177, 80, 29), \text{ } M), \\ ((170, 62, 27), \text{ } W), \\ ((172, 72, 30), \text{ } M), \\ ((175, 68, 27), \text{ } W), \\ ((178, 80, 29), \text{ } M) \end{array} \}$$