Homework 3

Artificial Intelligence

CS 540 Section 3

Ritvik Upadhyaya

Contents

1 Hierarchical Clustering

1.1 HAC

1.1.1 Single Linkage

{6}, {8}, {10}, {20}, {26}, {30}, {32}, {33}
{6}, {8}, {10}, {20}, {26}, {30}, {32, 33}
{6, 8}, {10}, {20}, {26}, {30}, {32, 33}
{6, 8, 10}, {20}, {26}, {30}, {32, 33}
{6, 8, 10}, {20}, {26}, {30, 32, 33}
{6, 8, 10}, {20}, {26, 30, 32, 33}
{6, 8, 10}, {20, 26, 30, 32, 33}
{6, 8, 10, 20, 26, 30, 32, 33}

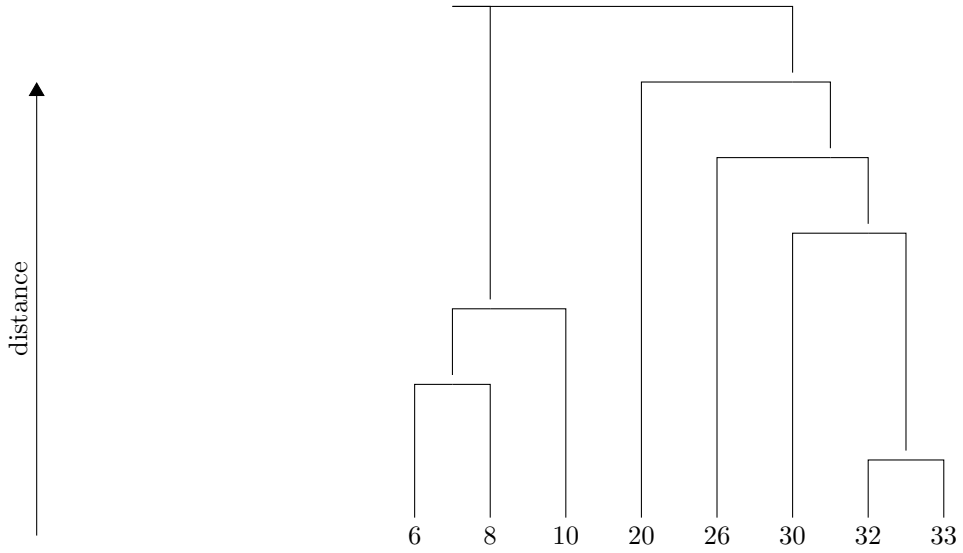
1.1.2 Complete Linkage

{6}, {8}, {10}, {20}, {26}, {30}, {32}, {33}
{6}, {8}, {10}, {20}, {26}, {30}, {32, 33}
{6, 8}, {10}, {20}, {26}, {30}, {32, 33}
{6, 8}, {10}, {20}, {26}, {30, 32, 33}
{6, 8, 10}, {20}, {26}, {30, 32, 33}
{6, 8, 10}, {20, 26}, {30, 32, 33}
{6, 8, 10}, {20, 26, 30, 32, 33}
{6, 8, 10, 20, 26, 30, 32, 33}

1.1.3 Average Linkage

{6}, {8}, {10}, {20}, {26}, {30}, {32}, {33}
{6}, {8}, {10}, {20}, {26}, {30}, {32, 33}
{6, 8}, {10}, {20}, {26}, {30}, {32, 33}
{6, 8, 10}, {20}, {26}, {30}, {32, 33}
{6, 8, 10}, {20}, {26}, {30, 32, 33}
{6, 8, 10}, {20}, {26, 30, 32, 33}
{6, 8, 10}, {20, 26, 30, 32, 33}
{6, 8, 10, 20, 26, 30, 32, 33}

1.2 Single Linkage Dendogram



Note: Ignore the extending line at the top

1.3 Good Clustering

There is no real "good" clustering method. A good cluster depends on the data being clustered and the characteristics of that data. We see the characteristics of the cluster after each step and decide when do we have a data that can be used and has a high number of homogeneous clusters. In our question we see that single and average result in the same dendogram so it can be debated that those are the best way even though they are completely different. Hence there is no really good way to cluster data and it is a matter of instance.

2 k Nearest Neighbor

2.1 Classification

Table 1: My caption

Weight (lbs)	Height (in)	Shoe Size	Age	Square Root Distance From First Data	Distance from Second Data
90	52	7	10	104	2348
130	69	9.5	20	967.25	541.25
50	45	6	10	2526	6934
63	51.5	6.5	10	1394.25	4737.5
145	70	11	20	2441	1029
160	69.5	10	20	3989.25	2021

From the first instance we get that 1,2,4 (row numbers) are the closest match. Thus the predicted age is 10 from the majority of the vote (1 and 4). From the second instance we get closest match at 2,5,6. The predicted age from this instance is 20 (all 3).

2.2 Regression

First data gives the predicted age as (10+20+10)/3 = 13.33 Second data gives the predicted age as (20+20+20)/3 = 20

3 Implementation of k Means

File is attached under the name "KMeans.java"