Michael Root

Data Mining

Homework 7

Writeup

1.) I'm going to assume maybe 3 hours of actively working on it. Probably more because I can get distracted pretty easily.

2.) Remove the ID because we can't use unique ID's for clustering.

3.) I just ended up using all attributes for centering the data in clusters.

4.) Code submitted separately.

5.) For the most part it was all 1's, 2's, and then some 3's, 4's, 5's, and an 8 or so. Until the last 2 iterations before completion where it went to 17, then 33, then was finished at 100

6.) First cluster:

IDs:
92+25+26+6+8+66+71+68+12+77+58+17+56+40+46+96+78+93+19+16+51+55+49+62+9+99+31+27+94+29+47+30+97

Size: 33

Second cluster:

IDs: 70+75+59+42+11+72+28+32+53+20+43+3+13+22+83+21+44

Size: 17

Third cluster:

IDs:

54+4+5+61+80+57+91+23+63+87+14+85+64+48+10+36+82+88+1+79+7+67+38+73+74+84+24+33+89+34+52+41+65+18+86+2+37+50+35+76+98+0+60+69+90+15+39+95+45+81

Size: 50

7.) Milk, veggies, and nuts are their main types. We call those people vegetarians because they don't partake in meat foods.

8.) We would need to keep track of the radii from the center of each point, and then measure from all radii to other radii to see where the shortest distance is.

9.) About 5 hours with a lot of procrastination in the mix of it. The hardest part was the time it took to write out the merging and calculation Euclidean distances functions.

10.) What are 4 types of inter-cluster distances?
- Single link (shortest distance)
- Complete link (longest distance)
- Average link (average distance)
- Central linkage (between centers)