Name:___Toghrul Tahirov____ Student Number:_ G47609664

Exercise 9

Hierarchical Clustering

Given the following document/document matrix along with four clusters (CL1-CL4):

C1 = D1, D2, D3; C2 = D4,D5,D6; C3=D7, D8, D9; CL4 = D10, D11, D12

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
D1		5	15	5	6	12	8	2	12	11	5	3
D2	5		8	18	23	7	17	10	20	10	5	8
D3	15	8		20	5	12	18	7	18	12	19	7
D4	5	18	20		17	8	16	9	19	14	13	10
D5	6	23	5	17		20	13	6	16	8	12	6
D6	12	7	12	8	20		14	17	4	16	18	22
D7	8	17	18	16	13	14		10	5	5	6	17
D8	2	10	7	9	6	17	10		8	2	17	2
D9	12	20	18	19	16	4	5	8		12	10	18
D10	11	10	12	14	8	16	5	2	12		19	12
D11	5	5	19	13	12	18	6	17	10	19		13
D12	3	8	7	10	6	22	17	2	18	12	13	

1. Show the order of combining clusters using single link. For each combination tell which cells/values caused the clusters to be combined.
   Cluster combination will be based on the highest pairwise similarities between the documents across the existing classes. Using this strategy, we have:
   - For C1, C2: D2xD5 (25) -> C1, C2 combined -> CombC1C2
   - For CombC1C2, C3: D2xD9 (20) -> CombC1C2C3
   - CombC1C2C3, C4: D3xD11 (19) -> CombC1C2C3C4
2. Show the order of combining the clusters using complete link. For each combination tell which cells/values caused the clusters to be combined. NOTE: was not clear in last class but find minimum similarity between all combinations of classes (CL1:CL2; CL1:CL3; CL1:CL4; CL2:CL3; CL2:CL4; CL3:CL4) then choose **MAXIMUM** of them to decide on clusters to combine. Then do same with three clusters from that action.
   Using the proposed method, we get the following:
   - CL1:CL2=5, CL1:CL3=2, CL1:CL4=3, CL2:CL3=4, CL2:CL4=6, CL3:CL4=2. The highest score amongst these is between CL2, CL4, so they are combined to form a new cluster called CombCL2CL4

- CL1 x CombCL2CL4 = 3; CL1 x CL3 = 2; CL3 x CombCL2CL4 = 2. he highest similarity score is between CL1, CombCL2CL4, so they are combined to form a new cluster called CombCL2CL4CL1
- CL3 x CombCL2CL4CL1 = 2. Final cluster: CombCL2CL4CL1CL3

3. Show the order of combining the clusters using Average Link.
   - AVG(C1, C2) = 5+6+12+18+23+7+20+5+12=108/9
   - AVG(C1, C3) = 8+2+12+17+10+20+18+7+18=112/9
   - AVG(C1, C4) = 11+5+3+10+5+8+12+19+7=80/9
   - AVG(C2, C3) = 16+9+19+13+6+16+14+17+4=114/9
   - AVG(C2, C4) = 14+13+10+8+12+6+16+18+22=119/9
   - AVG(C3, C4) = 5+6+17+2+17+2+12+10+18=89/9

   Max: C2, C4 = 119/9 -> CombC2C4

   - AVG(CombC2C4, C1) = (107+80) / 18
   - AVG(CombC2C4, C3) = (114+89) / 18
   - AVG(C1, C3) = 8+2+12+17+10+20+18+7+18=112/9

   Max: C1, C3 = 112/9 - > CombC1C3. Finally, the full cluster becomes Comb(CombC2C4, CombC1C3)

4. Show the order of combining the clusters using Group Average Link.
   - AVG(C1, C2) = (5+6+12+18+23+7+20+5+12+ 28+45)/15=181/15
   - AVG(C1, C3) = 8+2+12+17+10+20+18+7+18+28+23=163/15
   - AVG(C1, C4) = 11+5+3+10+5+8+12+19+7+28+44=152/15
   - AVG(C2, C3) = 16+9+19+13+6+16+14+17+4+45+23=182/15
   - AVG(C2, C4) = 14+13+10+8+12+6+16+18+22+45+44=208/15
   - AVG(C3, C4) = 5+6+17+2+17+2+12+10+18+23+44=156/15

   Max: C2, C4 = 208/15 -> CombC2C4

   - AVG(C1, CombC2C4) = 28+45+44+188=305/36
   - AVG(C3, CombC2C4) = 45+23+44+203=315/36
   - AVG(C1, C3) = 8+2+12+17+10+20+18+7+18+28+23=163/15

   Max: C3, CombC2C4= 315/36 -> Comb(C3, CombC2C4)
   Final: Comb(C4, Comb(C3, CombC2C4))