IDM Exercise 6
Selina Klees
Submission date: 8.12.17

1. **Assignment**
    a. **Hierarchical clustering**

       1, 4, 9, 16, 25, 36, 49, 64, 81

       {1, 4}, 9, 16, 25, 36, 49, 64, 81 → 2.5, 9, 16, 25, 36, 49, 64, 81

       {1, 4, 9}, 16, 25, 36, 49, 64, 81 → 4.6, 16, 25, 36, 49, 64, 81

       {1, 4, 9}, {16, 25}, 36, 49, 64, 81 → 4.6, 20.5, 36, 49, 64, 81

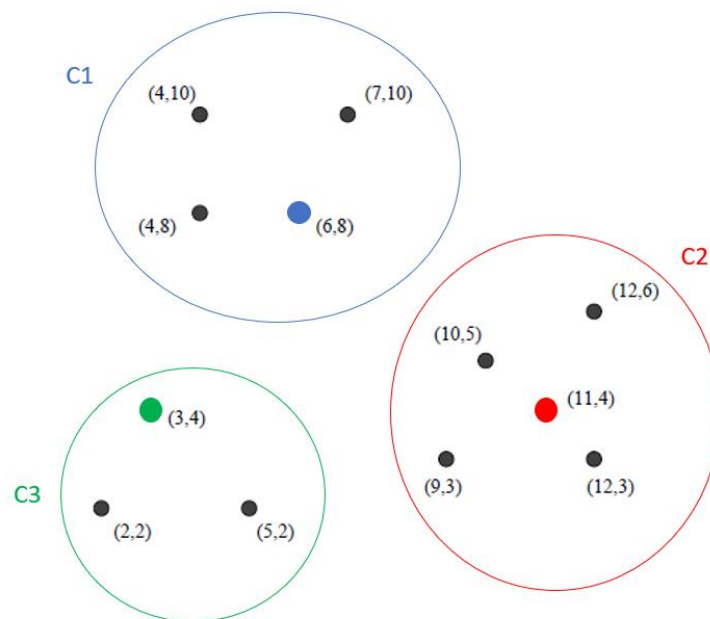   {1, 4, 9}, {16, 25}, {36, 49}, 64, 81 → 4.6, 20.5, 42.5, 64, 81

       {1, 4, 9, 16, 25}, {36, 49}, 64, 81 → 11, 42.5, 64, 81

       {1, 4, 9, 16, 25}, {36, 49}, {64, 81} → 11, 42.5, 72.5

       {1, 4, 9, 16, 25}, {36, 49, 64, 81} → 11, 57.5

       {1, 4, 9, 16, 25, 36, 49, 64, 81} → 31.7

    b. **Clustroids (minimum sum of distances)**

C1	4,10	7,10	4,8	6,8	Sum
4,10	-	3	2	$\sqrt{8}$	7.83
7,10	3	-	$\sqrt{13}$	$\sqrt{5}$	8.84
4,8	2	$\sqrt{13}$	-	2	7.61
6,8	$\sqrt{8}$	$\sqrt{5}$	2	-	7.06

C2	10,5	9,3	11,4	12,6	12,3	Sum
10,5	-	$\sqrt{5}$	$\sqrt{2}$	$\sqrt{5}$	$\sqrt{8}$	8.71
9,3	$\sqrt{5}$	-	$\sqrt{5}$	$\sqrt{18}$	3	11.71

11,4	$\sqrt{2}$	$\sqrt{5}$	-	$\sqrt{5}$	$\sqrt{2}$	7.3
12,6	$\sqrt{5}$	$\sqrt{18}$	$\sqrt{5}$	-	3	11.71
12,3	$\sqrt{8}$	3	$\sqrt{2}$	3	-	10.24

C3	3,4	2,2	5,2	Sum
3,4	-	$\sqrt{5}$	$\sqrt{8}$	5.06
2,2	$\sqrt{5}$	-	9	11.24
5,2	$\sqrt{8}$	9	-	11.83

2. **Assignment**

   a. **BFR – cluster representation**

	C1	C2	C3
N	4	5	3
SUM	(21,36)	(54,21)	(10,8)
SUMSQ	(117,328)	(590,95)	(38,24)

   b. **Variance and standard deviation**

C1	Var_i	Sd_i
Dimension = 1	$\dfrac{117}{4} - \left(\dfrac{21}{4}\right)^2 = 1.69$	$\sqrt{1.69} = 1.3$
Dimension = 2	$\dfrac{328}{4} - \left(\dfrac{36}{4}\right)^2 = 1$	1

C2	Var_i	Sd_i
Dimension = 1	$\dfrac{590}{5} - \left(\dfrac{54}{5}\right)^2 = 1.36$	$\sqrt{1.36} = 1.17$
Dimension = 2	$\dfrac{95}{5} - \left(\dfrac{21}{5}\right)^2 = 1.36$	$\sqrt{1.36} = 1.17$

C3	Var_i	Sd_i
Dimension = 1	$\dfrac{38}{3} - \left(\dfrac{10}{3}\right)^2 = 1.56$	$\sqrt{1.56} = 1.25$
Dimension = 2	$\dfrac{24}{3} - \left(\dfrac{8}{3}\right)^2 = 0.89$	$\sqrt{0.89} = 0.94$

   c. **Mahalanobis distance**

$$dis_{mahalanobis} = \sqrt{\sum_{i=1}^{d} \left(\frac{p_i - c_i}{\sigma_i}\right)^2} = \left(\frac{1-0}{2}\right)^2 + \left(\frac{-3-0}{3}\right)^2 + \left(\frac{4-0}{5}\right)^2$$

$$= 0.25 + 1 + 0.64 = 1.89$$