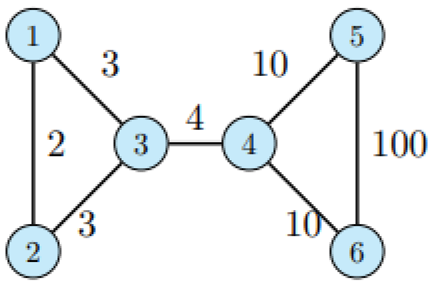
Assignment 4

Name: Kalpita Dapkekar

1)



	1	2	3	4	5	6
1	0	2	3	0	0	0
2	2	0	3	0	0	0
3	3	3	0	4	0	0
4	0	0	4	0	10	10
5	0	0	0	10	0	100
6	0	0	0	10	100	0

A) Cuts: Consider the following cuts:

cut 1: S1 = {1, 2, 3}, T1 = {4, 5, 6}
cut 2: S2 = {1}, T2 = {2, 3, 4, 5, 6}
cut 3: S3 = {1, 2, 3, 4}, T3 = {5, 6}

cut 1: S_1 = {1, 2, 3}, T_1 = {4, 5, 6}

1) $W(S_1, T_1) = 0 + 0 + 4 + 0 + 0 + 0$

$= 4$

2) Ratio Cut $= \frac{W(S_1,T_1)}{|S_1|} + \frac{W(S_1,T_1)}{|T_1|}$

$= \frac{4}{3} + \frac{4}{3} = \frac{8}{3} = 2.6$

3) Normalized Cut $= \frac{W(S_1,T_1)}{Vol(S_1)} + \frac{W(S_1,T_1)}{Vol(T_1)}$

$= \frac{4}{20} + \frac{4}{244} = 0.216$

cut 2: S_2 = {1}, T_2 = {2, 3, 4, 5, 6}

1) $W(S_2, T_2) = 2 + 3$

$= 5$

2) Ratio Cut $= \frac{W(S_2,T_2)}{|S_2|} + \frac{W(S_2,T_2)}{|T_2|}$

$= \frac{5}{1} + \frac{5}{5} = 6$

3) Normalized Cut $= \frac{W(S_2,T_2)}{Vol(S_2)} + \frac{W(S_2,T_2)}{Vol(T_2)}$

$= \frac{5}{5} + \frac{5}{259} = 1.019$

cut 3: S_3 = {1, 2, 3, 4}, T_3 = {5, 6}

1) W (S_3, T_3) = 10
2) Ratio Cut = $\frac{W(S_3,T_3)}{|S_3|} + \frac{W(S_3,T_3)}{|T_3|}$

$= \frac{10}{4} + \frac{10}{2}$ = 2.5 + 5 = 7.5

3) Normalized Cut = $\frac{W(S_3,T_3)}{Vol(S_3)} + \frac{W(S_3,T_3)}{Vol(T_3)}$

$= \frac{10}{34} + \frac{10}{210}$ = 0.23 + 0.045 = 0.275

From Smallest to highest cut criteria for **Cut Weight:** *Cut1* $\leq$ *Cut2* $\leq$ *Cut3*

From Smallest to highest cut criteria for **Ratio Cut:** *Cut1* $\leq$ *Cut2* $\leq$ *Cut3*

From Smallest to highest cut criteria for **Normalized Cut:** *Cut1* $\leq$ *Cut3* $\leq$ *Cut2*

B) Laplacians: Compute the adjacency matrix A, degree matrix D, Laplacian matrix L = D−A and Symmetric Laplacian matrix Ls = D−1/2LD−1/2 of the graph above (do this by hand and show them in your Solutions file).

Adjacency Matrix (A)

	1	**2**	**3**	**4**	**5**	**6**
1	0	2	3	0	0	0
2	2	0	3	0	0	0
3	3	3	0	4	0	0
4	0	0	4	0	10	10
5	0	0	0	10	0	100
6	0	0	0	10	100	0

Degree Matrix (D)

	1	**2**	**3**	**4**	**5**	**6**
1	5	0	0	0	0	0
2	0	5	0	0	0	0
3	0	0	10	0	0	0
4	0	0	0	24	0	0
5	0	0	0	0	110	0
6	0	0	0	0	0	110

L = D – A

	1	**2**	**3**	**4**	**5**	**6**
1	5	-2	-3	0	0	0
2	-2	5	-3	0	0	0
3	-3	-3	10	-4	0	0
4	0	0	-4	24	-10	-10
5	0	0	0	-10	110	-100
6	0	0	0	-10	-100	110

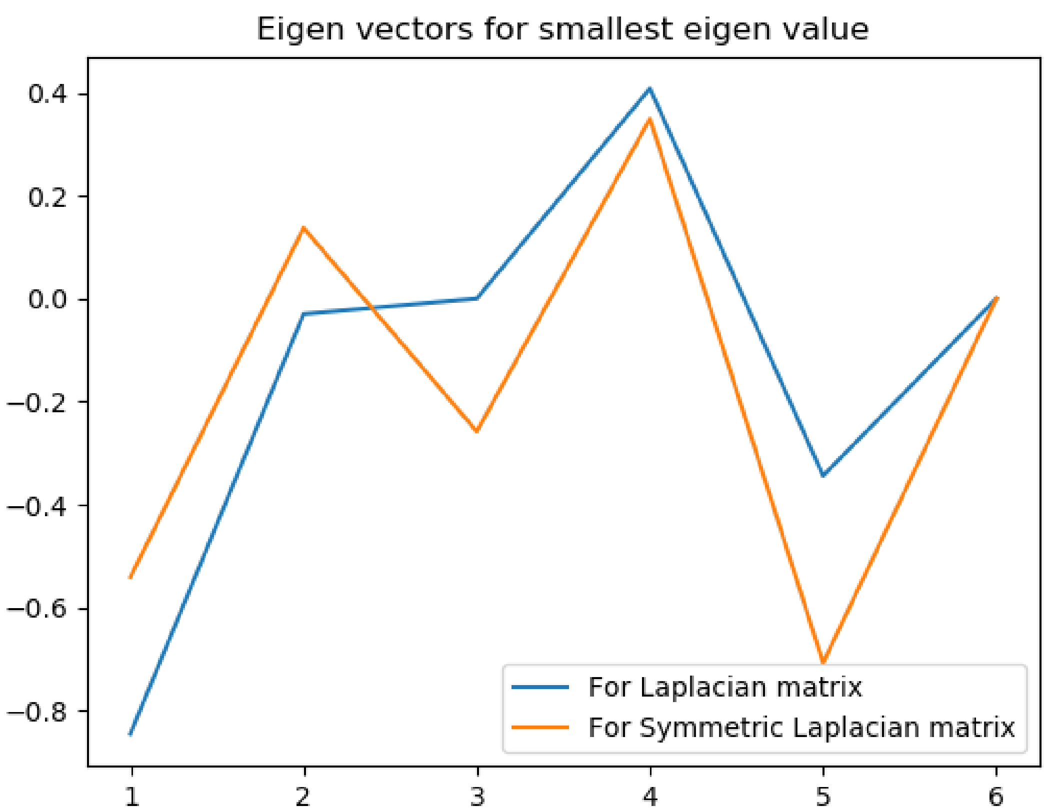
Symmetric Laplacian Matrix

$L_S = D^{-1/2} L D^{-1/2}$

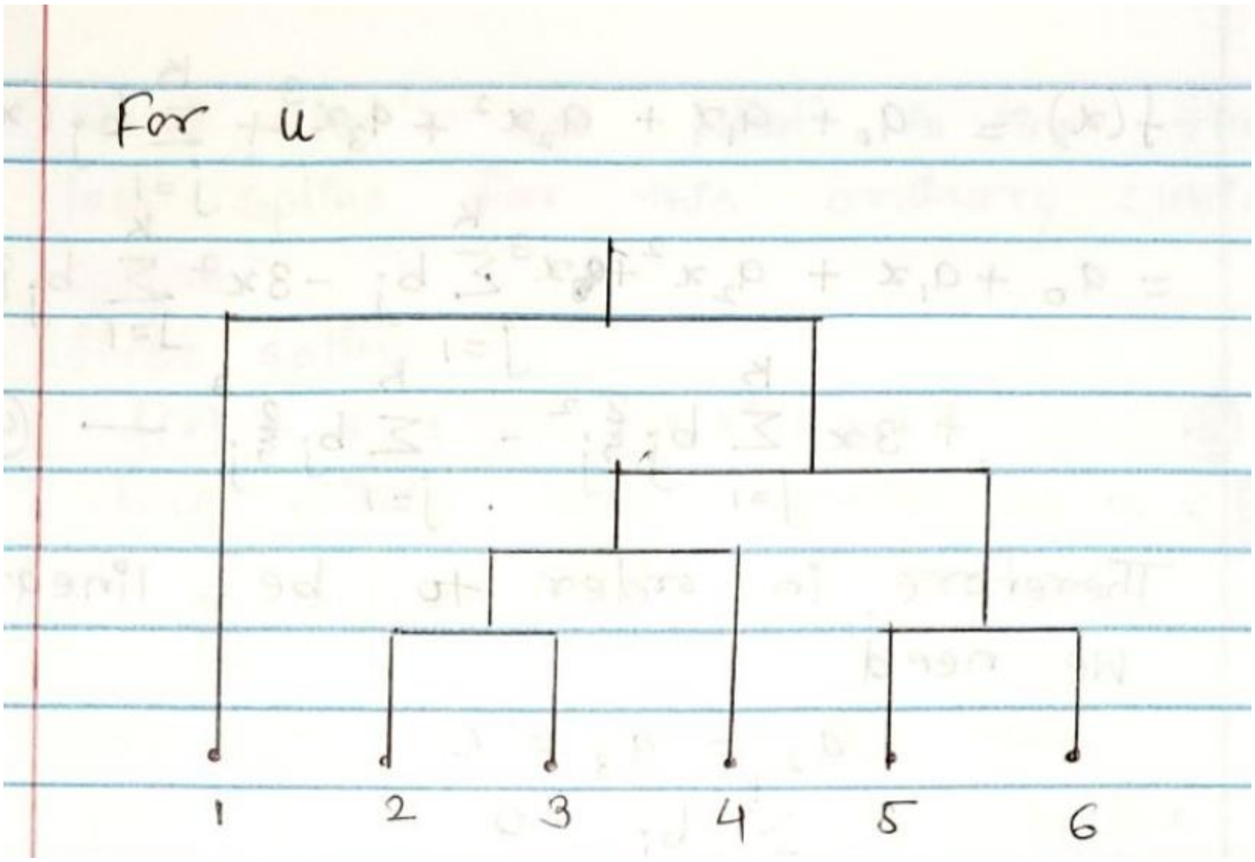
$L_S =$

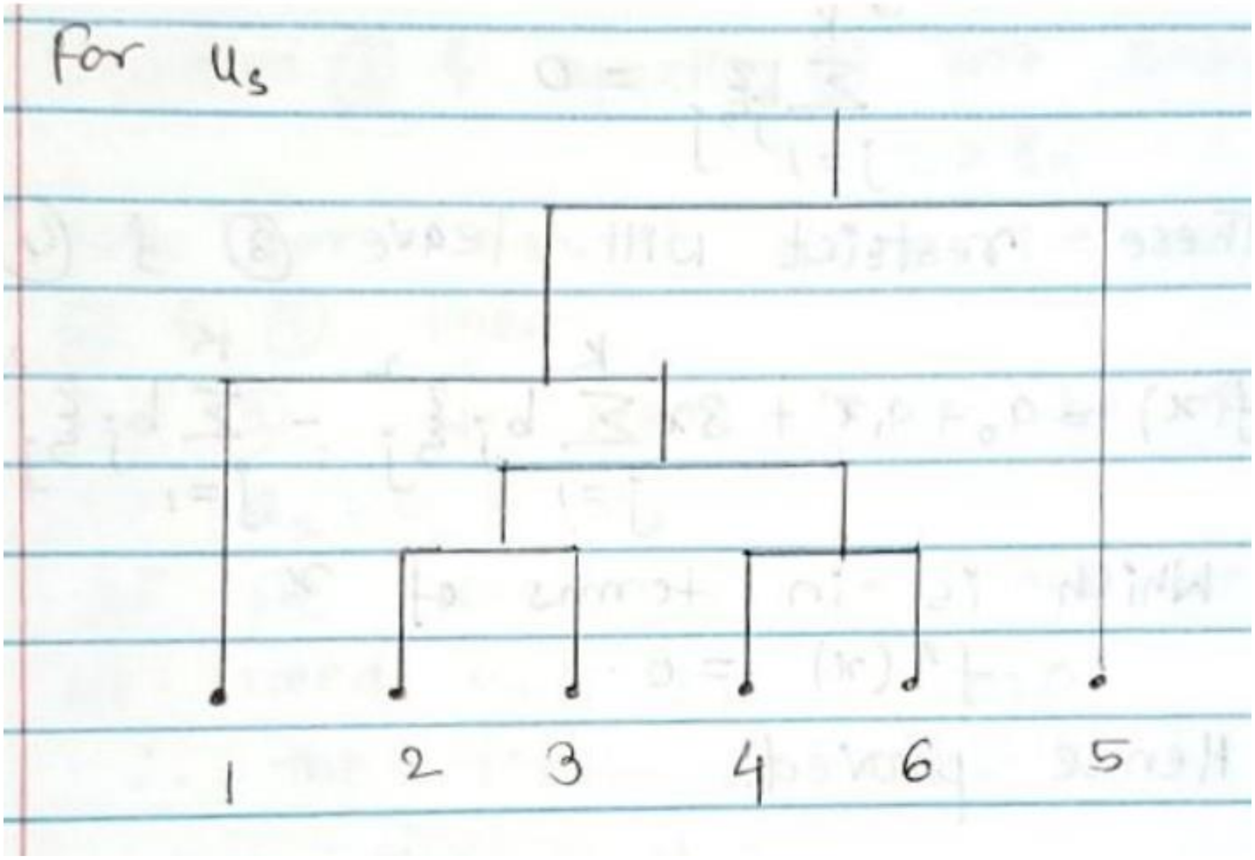
1	-0.4	-0.42	0	0	0
-0.42	1	-0.42	0	0	0
-0.42	-0.42	1	-0.25	0	0
0	0	-0.25	1	-0.19	-0.19
0	0	0	-0.19	1	-0.90
0	0	0	-0.19	-0.90	1

C) Eigen values and eigen vector:



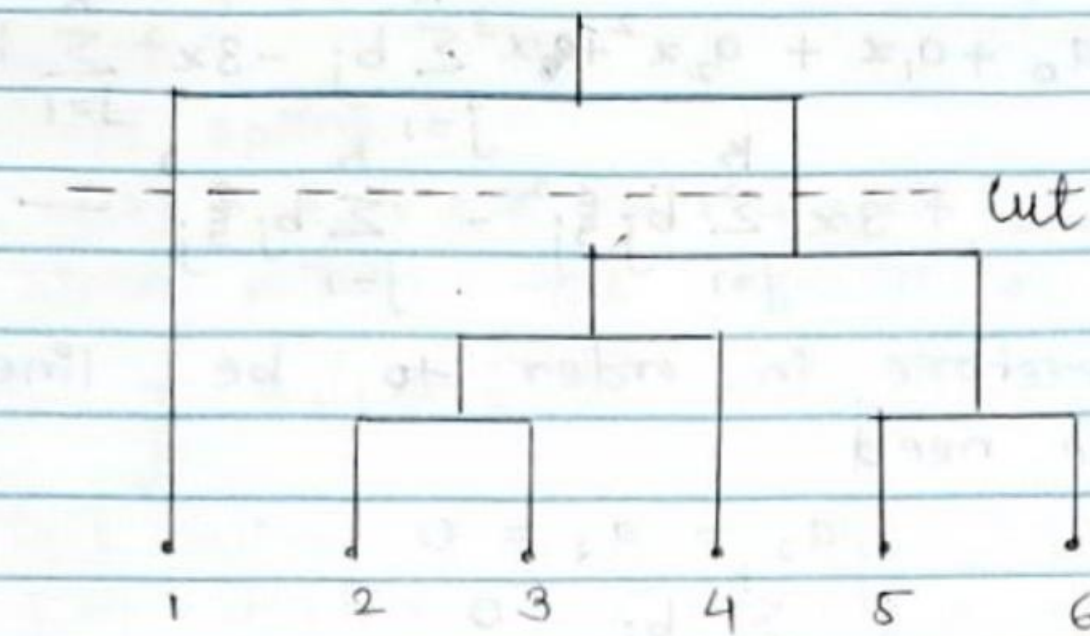
D) Hierarchical clusters from eigen vectors:



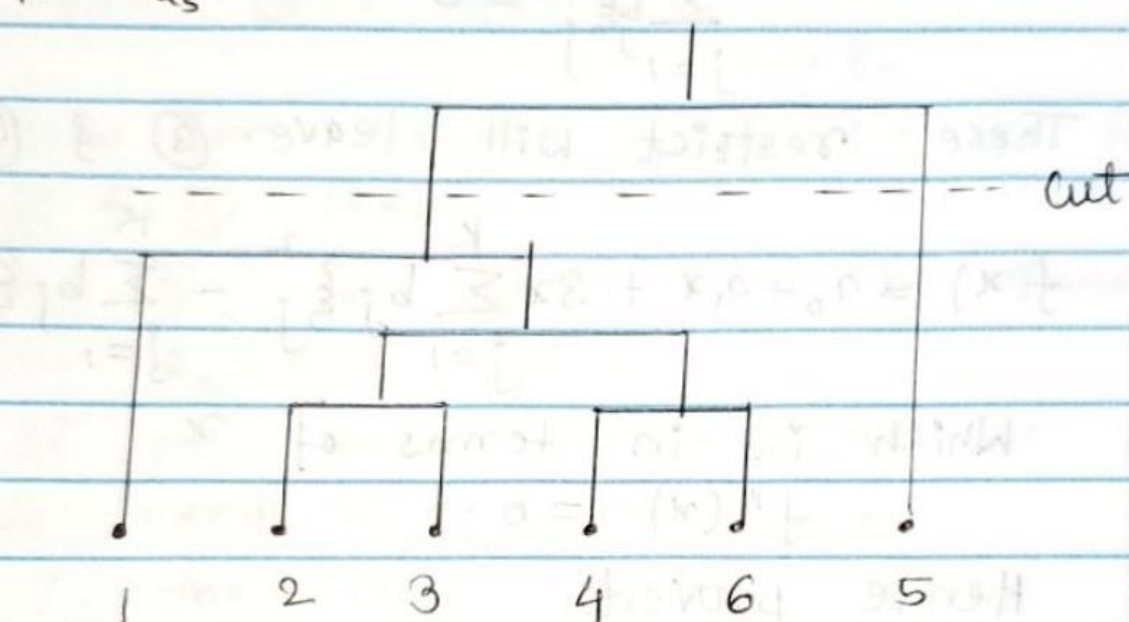


E) Hierarchical clusters from eigen vectors:

For u



For u_s



2

A) Given points: {0, 1, 2, 2, 10}

$\mu = 3$

$m = 2$

1) $\sum_i (x_i - \mu)^2$ = $(0-3)^2 + (1-3)^2 + (2-3)^2 + (2-3)^2 + (10-3)^2$
= 9 + 4 + 1 + 1 + 49 = 64

$\sum_i (x_i - m)^2$ = $(0-2)^2 + (1-2)^2 + (2-2)^2 + (2-2)^2 + (10-2)^2$
= 4 + 1 + 0 + 0 + 64 = 69

64 < 69

$$\sum_i (x_i - \mu)^2 \leq \sum_i (x_i - m)^2$$

2) $\sum_i |x_i - m|$ = |0 – 2| + |1 – 2| + |2 -2| + |2 – 2| + |10 – 2|

= 2 + 1 + 0 + 0 + 8 = 11

$\sum_i |x_i - \mu|$ = |0 – 3| + |1 – 3| + |2 -3| + |2 – 3| + |10 – 3|
= 3 + 2 + 1 + 1 + 7 = 14

11 < 14

$$\sum_i |x_i - m| \leq \sum_i |x_i - \mu|$$

B) b) $\mu = \arg\min_a \sum_i (x_i - a)^2$

The above expression can be written in the form of L_2 norm

$\mu = \arg\min_a \sum_{x \in X} ||x_i - a||^2 = \arg\min_a \sum_{x \in X} < x - a, x - a >$

$= \arg\min_a \sum_{x \in X} (< x, x > - 2 < x, a > + < a, a >)$ ……………… by taking the inner product

$= \arg\min_a < a, a > - 2n < \frac{1}{n} \sum_{x \in X} x, \mu >$

$= \arg\min_a ||a - \bar{x}||^2$ Where, $\bar{x} = \frac{1}{n} \sum_{x \in X} x$

L_2 norm can never be smaller than 0, "a" should be equal to $\bar{x}$ so that $\arg\min_a ||a - \bar{x}||^2$ is minimum i.e., zero

That is, $a = \bar{x}$

C) m = arg min$_a \sum_{i=1}^{n} |x_i - a|$ .... $x_i \in X$

So, for calculating median, we need to arrange all numbers x_i in X set in ascending order.

$X_1 \leq X_2 \leq X_3 .... \leq X_n$

Case 1: The odd case, where there are 2n – 1 points, choose "*a*" such that the sum of absolute distance is minimized.
Let us consider a point on number line (not between X_1 & X_{2n-1})
The sum of absolute distances of 2n-1 points from that point would be greater than a point which is chosen between X_1 & X_{2n-1}

If $X_1 < a < X_{2n-1}$ then the absolute distance $| X_1 - a| + | X_{2n-1} - a|$ will be less choose a such that $| X_4 - a|$ is least *This is least when* $X_4 - a = 0$

Thus $X_4 = a$

m = argmin$_{a=X4} \sum |X_1 - X_4|$

$= | X_4 - X_1 | + | X_4 - X_2 | + | X_4 - X_3 | + | X_4 - X_4 | + | X_5 - X_4 | + | X_6 - X_4 | + | X_7 - X_4 |$

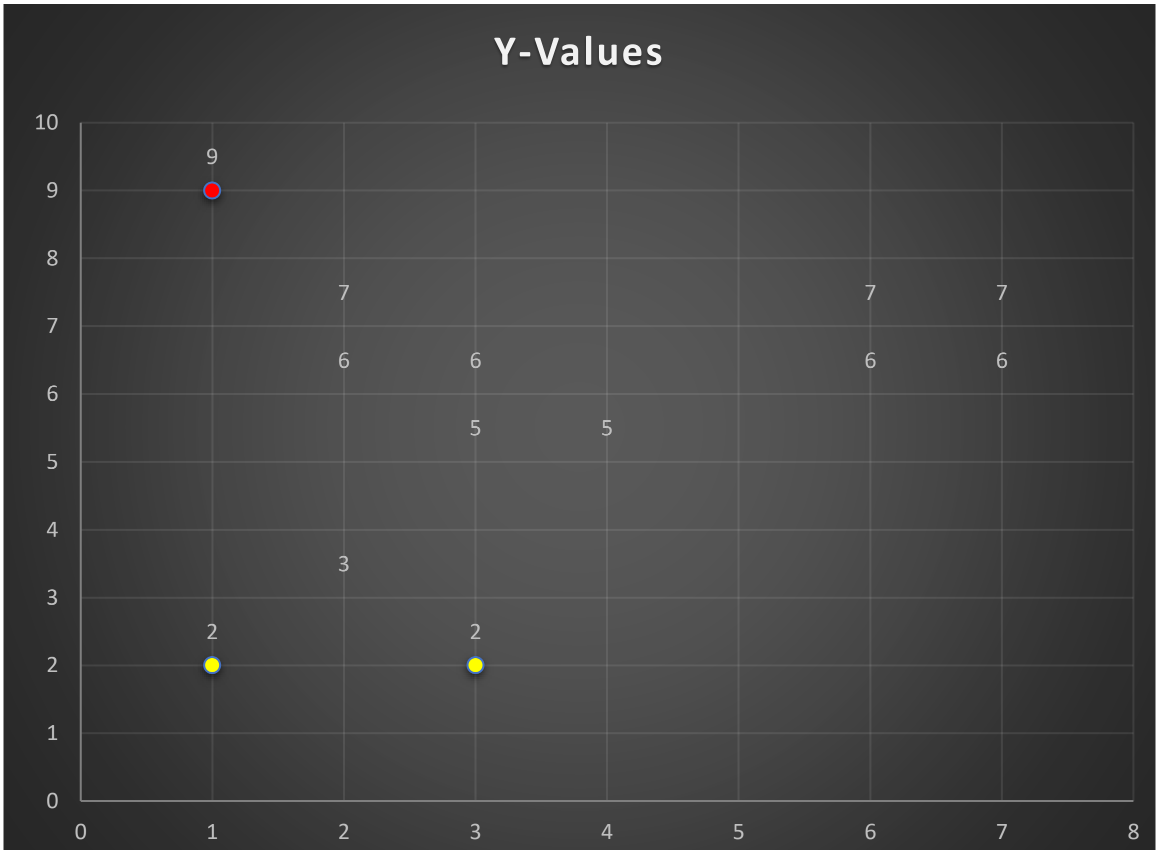
$= (X_5 + X_6 + X_7) - (X_1 + X_2 + X_3)$

Case 2: Even number, where there are 2n points, choose "*a*" such that the sum of absolute distance is minimized that is $X_n \leq a \leq X_{n+1}$

"*a*" will be between X_n & X_{n+1} to minimize $\sum_{i=1}^{n} |x_i - a|$

3

A)



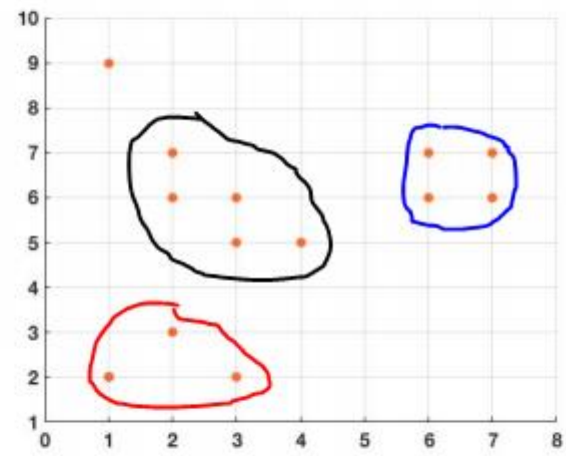
Blue point are **core**
Yellow points are **Border**
Red points are **noise**

B) Clusters

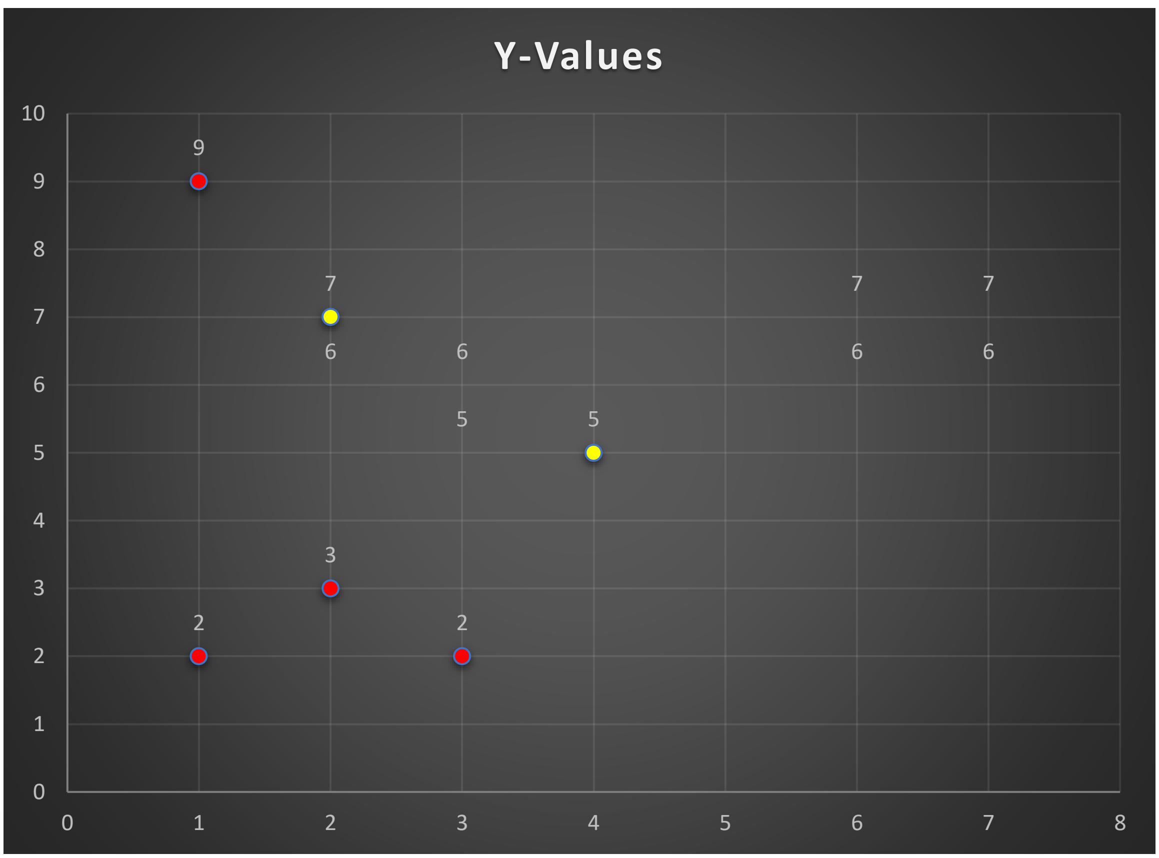
A Clusters is a subset of samples that satisfy following two criteria:

1. **Connected**
2. **Maximal**

a. Points in bottom of the graph labeled as 2,2 and 3 satisfies connected and maximal criteria, so they form a clusters, C = {(1,2),(2,3),(3,2)}.
b. Points in the middle of the graph satisfies both criteria
c. Points to the right of the graph forms another cluster.



C)



Blue point are **core**
Yellow points are **Border**
Red points are **noise**

D) Clusters

a. Points in the middle of the graph Satisfies both the criteria.

b. Points to the right of the graph Satisfies both the criteria.