Devon James
CSI 431
Prof. Petkov

December 3rd, 2018

Homework #4 Solutions

1.

a.

Part A

Cut 1

$W(S_1, T_1) = 1$ (Highest)

$\frac{W(S_1,T_1)}{|S_1|} + \frac{W(S_1,T_1)}{|T_1|} = \frac{1}{2} \cdot (\frac{1}{3} + \frac{1}{3}) = \frac{1}{2} \cdot \frac{2}{3}$

$\frac{W(S_1,T_1)}{vol(S_1)} + \frac{W(S_1,T_1)}{vol(T_1)} = \frac{1}{2} \cdot (\frac{1}{7} + \frac{1}{7}) = \frac{1}{2} \cdot \frac{2}{7}$ (Smallest)

Cut 2

$W(S_2, T_2) = 2$ (Highest)

$\frac{W(S_2,T_2)}{|S_2|} + \frac{W(S_2,T_2)}{|T_2|} = \frac{1}{2} \cdot (\frac{2}{1} + \frac{2}{5})$

$\frac{W(S_2,T_2)}{vol(S_2)} + \frac{W(S_2,T_2)}{vol(T_2)} = \frac{1}{2} \cdot (\frac{2}{2} + \frac{2}{12})$ (Smallest)

Cut 3

$W(S_3, T_3) = 20$ (Highest)

$\frac{W(S_3,T_3)}{|S_3|} + \frac{W(S_3,T_3)}{|T_3|} = \frac{1}{2} \cdot (\frac{20}{2} + \frac{20}{4})$

$\frac{W(S_3,T_3)}{vol(S_3)} + \frac{W(S_3,T_3)}{vol(T_3)} = \frac{1}{2} \cdot (\frac{20}{10} + \frac{20}{4})$ (Smallest)

b.

Part B

Adjacency Matrix

	1	2	3	4	5	6
1	0	1	1	0	0	0
2	1	0	1	0	0	0
3	1	1	0	1	0	0
4	0	0	1	0	1	1
5	0	0	0	1	0	1
6	0	0	0	1	1	0

Degree Matrix

	1	2	3	4	5	6
1	2	0	0	0	0	0
2	0	2	0	0	0	0
3	0	0	3	0	0	0
4	0	0	0	3	0	0
5	0	0	0	0	2	0
6	0	0	0	0	0	2

Laplacian Matrix (L)

	1	2	3	4	5	6
1	2	-1	-1	0	0	0
2	-1	2	-1	0	0	0
3	-1	-1	3	-1	0	0
4	0	0	-1	3	-1	-1
5	0	0	0	-1	2	-1
6	0	0	0	-1	-1	2

Symmetric Laplacian Matrix (L_s)

	1	2	3	4	5	6
1	4	-2	$-\sqrt{6}$	0	0	0
2	-2	4	$-\sqrt{6}$	0	0	0
3	$-\sqrt{6}$	$-\sqrt{6}$	9	-3	0	0
4	0	0	-3	9	$-\sqrt{6}$	$-\sqrt{6}$
5	0	0	0	$-\sqrt{6}$	4	-2
6	0	0	0	$-\sqrt{6}$	-2	4

Computation

$$\underbrace{\begin{bmatrix} -\sqrt{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\sqrt{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\sqrt{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\sqrt{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & -\sqrt{2} \end{bmatrix}}_{-\sqrt{D}} * \underbrace{\begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ -1 & -1 & 3 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & -1 & 2 \end{bmatrix}}_{L}$$

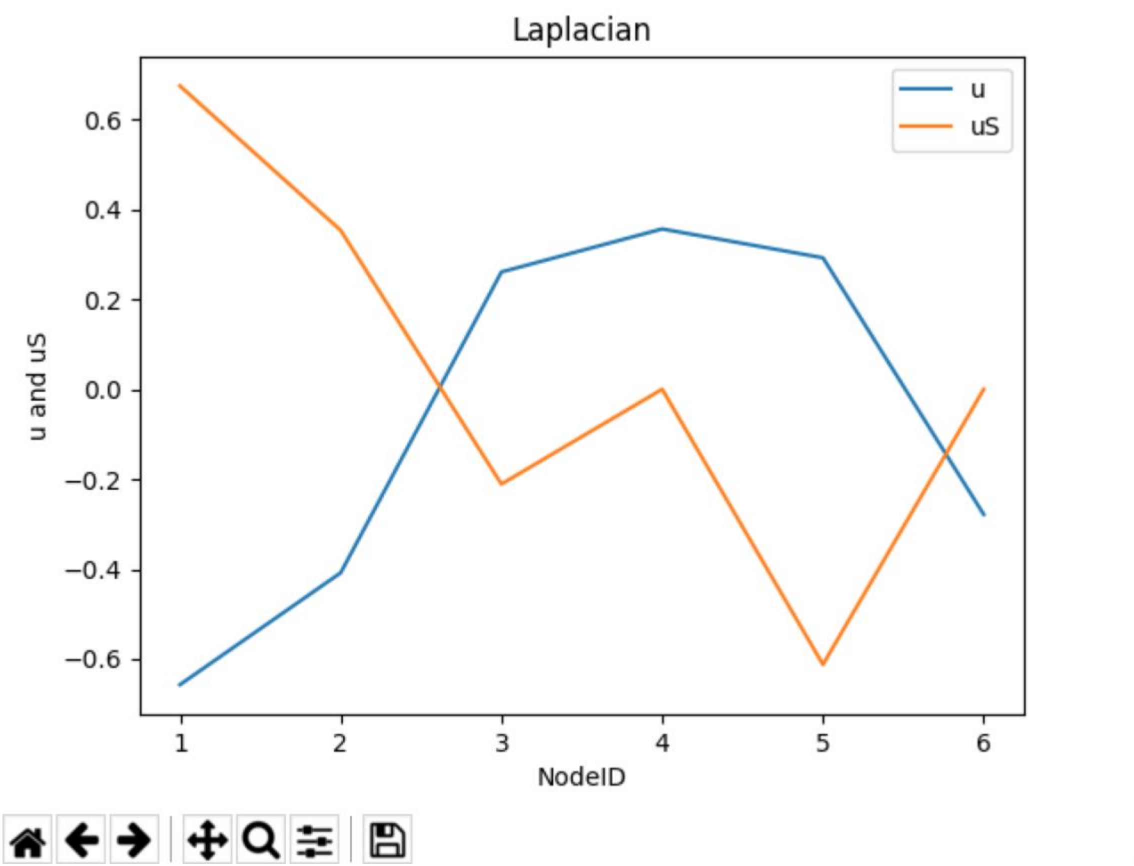
Cont. →

$$
= \underbrace{\begin{bmatrix} -2\sqrt{2} & \sqrt{2} & \sqrt{2} & 0 & 0 & 0 \\ \sqrt{2} & -2\sqrt{2} & \sqrt{2} & 0 & 0 & 0 \\ \sqrt{3} & \sqrt{3} & -3\sqrt{3} & \sqrt{3} & 0 & 0 \\ 0 & 0 & \sqrt{3} & -3\sqrt{3} & \sqrt{3} & \sqrt{3} \\ 0 & 0 & 0 & \sqrt{2} & -\sqrt{2}\cdot 2 & \sqrt{2} \\ 0 & 0 & 0 & \sqrt{2} & \sqrt{2} & -2\sqrt{2} \end{bmatrix}}_{(-\sqrt{D} * L)} * \underbrace{\begin{bmatrix} -\sqrt{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\sqrt{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\sqrt{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\sqrt{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & -\sqrt{2} \end{bmatrix}}_{-\sqrt{D}}
$$

$$
= \underbrace{\begin{bmatrix} 4 & -2 & -\sqrt{6} & 0 & 0 & 0 \\ -2 & 4 & -\sqrt{6} & 0 & 0 & 0 \\ -\sqrt{6} & -\sqrt{6} & 9 & -3 & 0 & 0 \\ 0 & 0 & -3 & 9 & -\sqrt{6} & -\sqrt{6} \\ 0 & 0 & 0 & -\sqrt{6} & 4 & -2 \\ 0 & 0 & 0 & -\sqrt{6} & -2 & 4 \end{bmatrix}}_{L_s}
$$

c. Found in Code.zip

Plot:



d.

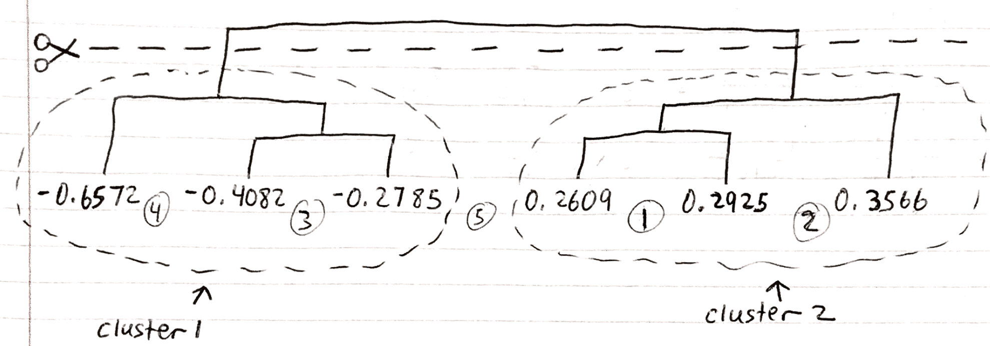
Parts D & E

$u = [-0.6572, -0.4082, 0.2609, 0.3566, 0.2925, -0.2785]$

$uS = [0.6749, 0.3536, -0.2110, -5.3557e-16, -0.6124, -2.6251e-16]$

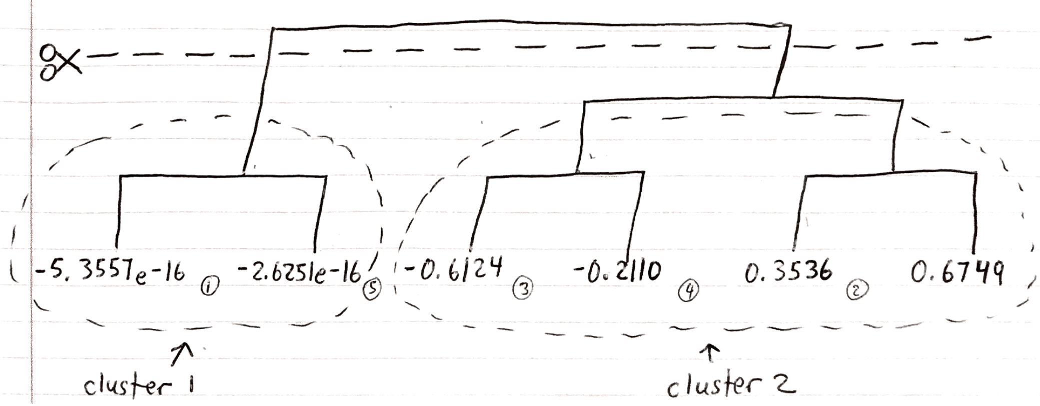
u

cut



uS

cut



e.

$\underline{u}$

Cut Weight $= 2$

Ratio Cut - $\frac{w(S_1, T_1)}{|S_1|} + \frac{w(S_1, T_1)}{|T_1|} = 1/2 \cdot (2/3 + 2/3)$

Normalized Cut - $\frac{w(S_1, T_1)}{vol(S_1)} + \frac{w(S_1, T_1)}{vol(T_1)} = 1/2 \cdot (2/5 + 2/5)$

$\underline{uS}$

Cut Weight $= 2$

Ratio Cut - $\frac{w(S_2, T_2)}{|S_2|} + \frac{w(S_2, T_2)}{|T_2|} = 1/2 \cdot (2/2 + 2/4)$

Normalized Cut - $\frac{w(S_2, T_2)}{vol(S_2)} + \frac{w(S_2, T_2)}{vol(T_2)} = 1/2 \cdot (2/3 + 2/7)$

2.

a.

2) a) $\{0, 1, 2, 2, 10\}$

$$\mu = \frac{0+1+2+2+10}{5} = \frac{15}{5} = 3$$

(Median) $m = 2$

Sum of Squared Distances – Mean (μ)

$$\Sigma_i (x_i - \mu)^2$$

$$= (0-3)^2 + (1-3)^2 + (2-3)^2 + (2-3)^2 + (10-3)^2$$
$$= (-3)^2 + (-2)^2 + (-1)^2 + (-1)^2 + (7)^2$$
$$= 9 + 4 + 1 + 1 + 49$$
$$\boxed{= 64}$$

Sum of Squared Distances – Median (m)

$$\Sigma_i (x_i - m)^2$$

$$= (0-2)^2 + (1-2)^2 + (2-2)^2 + (2-2)^2 + (10-2)^2$$
$$= (-2)^2 + (-1)^2 + (0)^2 + (0)^2 + (8)^2$$
$$= 4 + 1 + 0 + 0 + 64$$
$$\boxed{= 69}$$

$\therefore$ The sum of squared distances from the median is bigger.

b.

b) Prove that $\sum_i (x_i - \mu)^2 \le \sum_i (x_i - m)^2$

$$\sum_{i=1}^{N} (x_i - z)^2 = \sum_{i=1}^{N} (x_i^2 - 2z x_i - z^2)$$

$$= Nz^2 - 2Nz \sum_{i=1}^{N} x_i + \sum_{i=1}^{N} x_i^2$$

$$\frac{d}{dz}\left(Nz^2 - 2Nz \sum_{i=1}^{N} x_i + \sum_{i=1}^{N} x_i^2\right)$$

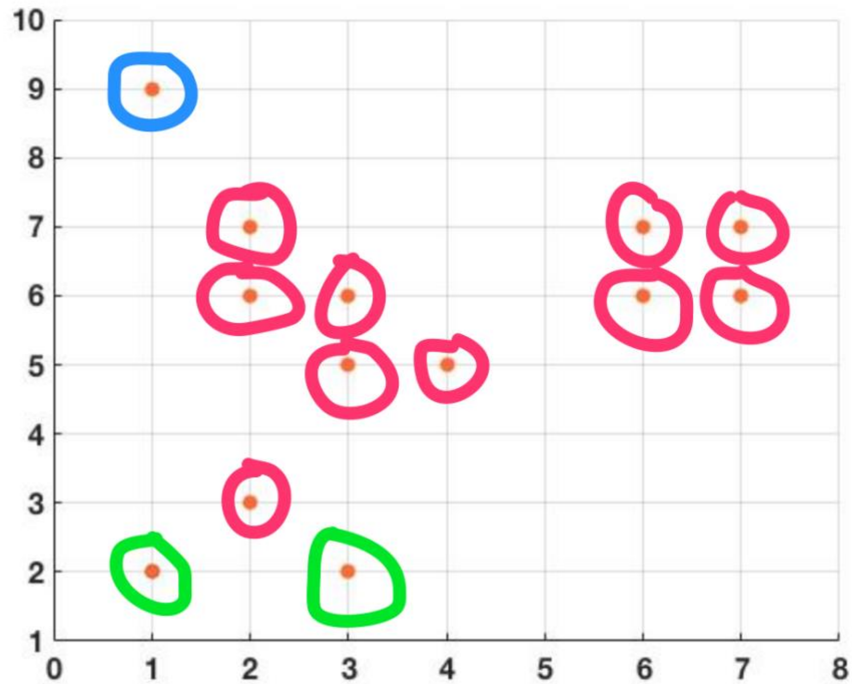
$$= 2Nz - 2N \sum_{i=1}^{N} x_i = 0$$

$$z = \sum_{i=1}^{N} x_i / N = \mu$$

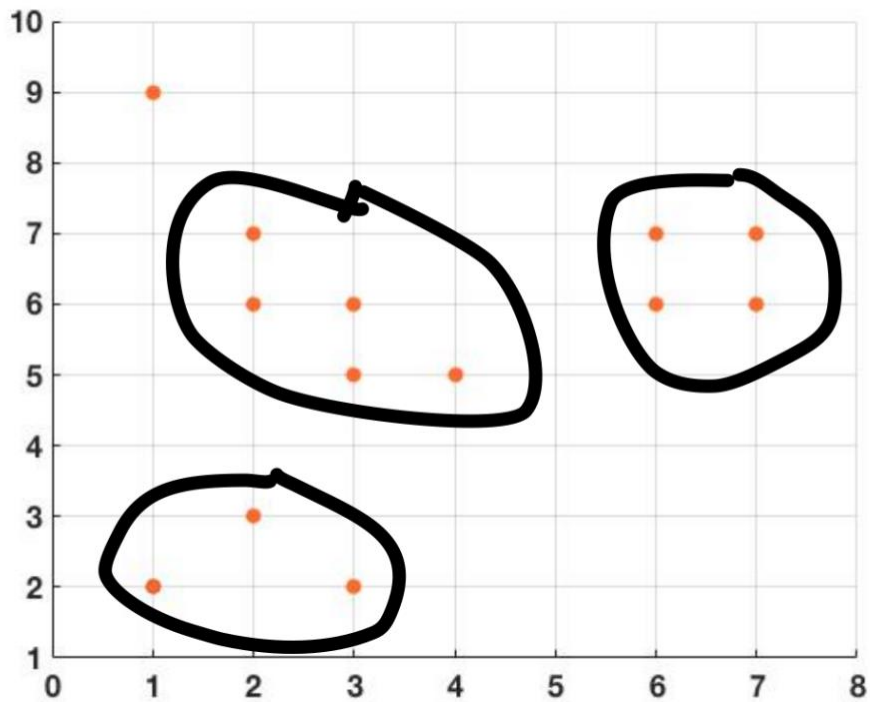
$\therefore$ the mean minimizes the sum of squared distances

3. Red – Core Points , Green – Border Points and Blue – Outlier Points
   Black – DBSCAN clusters

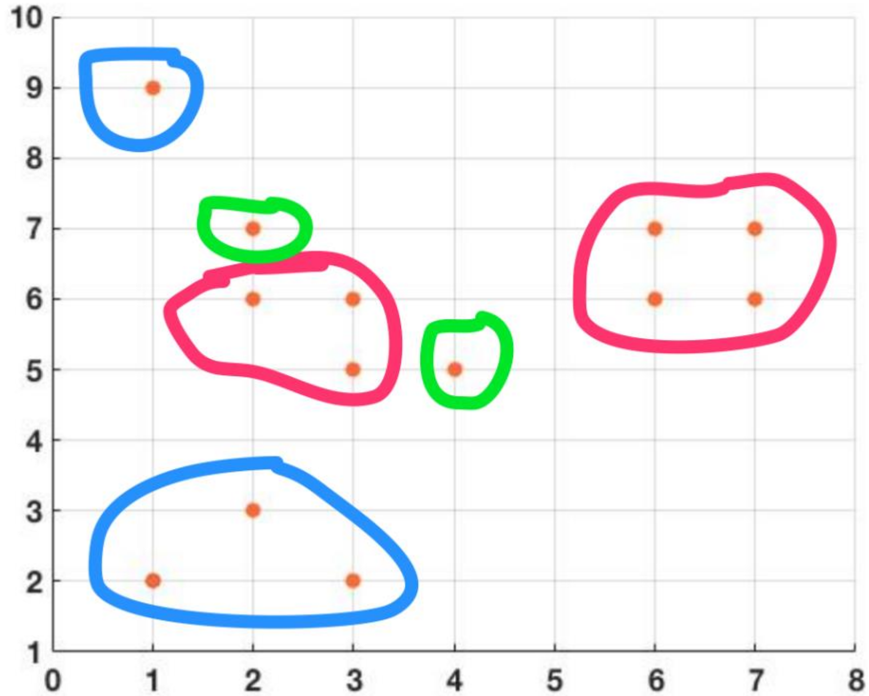
a. ______________________________



b.



c.



When compared with Part A, you can see that Points (1,2), (2,3), and (3,2) become outliers. You also notice that Points (2,7) and (4,5) change from being core points to border points. Because we are using L2 distance, Point (2,3) is not a core point anymore since when you compute the L2 distance for, example, (2,3) and (3,2), you get $\sqrt{(1+1)}$, which is greater than 1 (our epsilon value). With that being said, these points cannot match up with each other and so they all become outliers. (2,7) cannot match up with (3,6) and (4,5) cannot match up with (3,6) for the same reason. So, they become border points.

d.

When compared with Part B, you can see that DBSCAN would find 2 clusters instead of 3. This is because Point (2,3) is no longer a core point surrounded by border points; they are all outliers. Since DBSCAN only recognizes clusters of core and border points, the cluster around Point (2,3) is no longer recognized. The other two clusters remain the same because although Points (2,7) and (4,5) are now border points, they can still be included in the cluster found by DBSCAN.