Homework5

Houhao Liang

December 11, 2018

1 Question1

1.1 1a

1.1.1 1a(i)

Pr(Popularity = ’P’) = $\frac{7}{10} = 0.7$

1.1.2 1a(ii)

Pr(Popularity = ’NP’) = $\frac{3}{10} = 0.3$

1.1.3 1a(iii)

Set event x as Price = ‘\$’, Delivery = ‘Yes’ , Cuisine = ‘Korean’
Pr(Popularity = ‘P’ | Price = ‘\$’) = $\frac{3}{4}$
Pr(Popularity = ‘P’ | Delivery = ‘Yes’) = $\frac{5}{6}$
Pr(Popularity = ‘P’ | Cuisine = ‘Korean’) = $\frac{2}{3}$
Pr(x) = 0
Pr(x | Popularity = ‘P’) = Pr(Popularity = ‘P’ | Price = ‘\$’) Pr(Popularity = ‘P’ | Delivery = ‘Yes’) Pr(Popularity = ‘P’ | Cuisine = ‘Korean’) Pr(x) = 0

1.1.4 1a(iv)

Set event x as Price = ‘\$’, Delivery = ‘Yes’ , Cuisine = ‘Korean’
Pr(Popularity = ‘NP’ | Price = ‘\$’) = $\frac{1}{4}$
Pr(Popularity = ‘NP’ | Delivery = ‘Yes’) = $\frac{1}{6}$
Pr(Popularity = ‘NP’ | Cuisine = ‘Korean’) = $\frac{1}{3}$
Pr(x) = 0
Pr(x | Popularity = ‘NP’) = Pr(Popularity = ‘NP’ | Price = ‘\$’) Pr(Popularity = ‘NP’ | Delivery = ‘Yes’) Pr(Popularity = ‘NP’ | Cuisine = ‘Korean’) Pr(x) = 0

1.2 1b

Pr(Popularity = ‘P’ | x) = Pr(Price = ‘\$’ | Popularity = ‘P’) Pr(Delivery = ‘Yes’ | Popularity = ‘P’) Pr(Cuisine = ‘Korean’ | Popularity = ‘P’) Pr(Popularity = ‘P’) = $\frac{3}{7}\frac{5}{7}\frac{2}{7}\frac{7}{10} = \frac{3}{49}$

1.3 1c

1. Various Navie Bayes classifier with different features will be created.
2. Ensemble different classifier.
3. Boosting method can be used to improve the accuracy. It assigns weights for each training tuple, and use ensemble naive Bayes classifier to iteratively learn. The weights will be adjusted and finally find the best one.

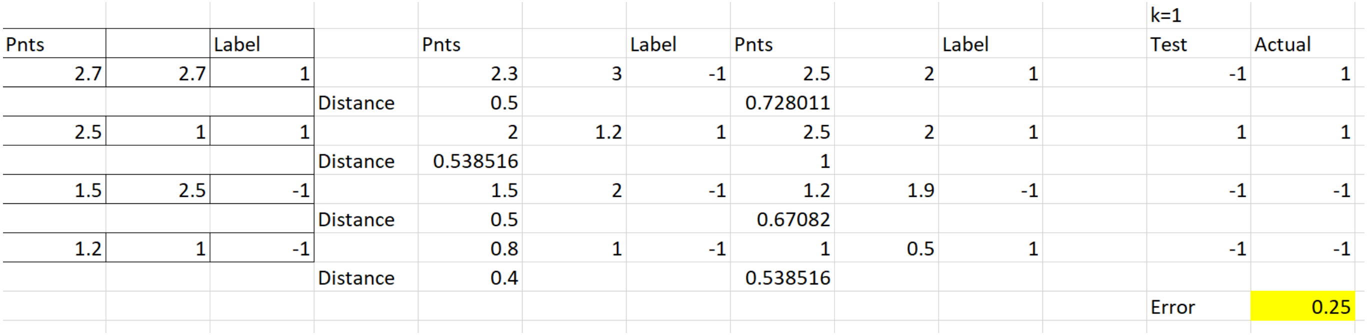
1.4 1d

Precision and sensitive. Precision: $\frac{TP}{TP+FP}$, and Sensitive: $\frac{TP}{P}$. These two metrics only use positive examples, which is good since we have rare positive examples.

2 Question 2

2.1 2a

K = 1, as you can see in the Figure 1, the testing error is 25%

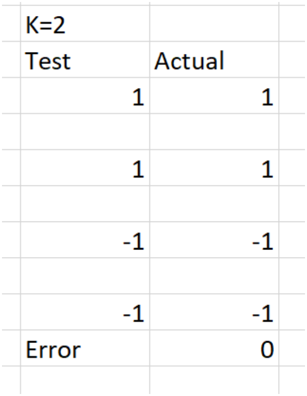


											k=1	
Pnts		Label		Pnts		Label	Pnts		Label		Test	Actual
2.7	2.7	1		2.3	3	-1	2.5	2	1		-1	1
			Distance	0.5			0.728011					
2.5	1	1		2	1.2	1	2.5	2	1		1	1
			Distance	0.538516			1					
1.5	2.5	-1		1.5	2	-1	1.2	1.9	-1		-1	-1
			Distance	0.5			0.67082					
1.2	1	-1		0.8	1	-1	1	0.5	1		-1	-1
			Distance	0.4			0.538516					
											Error	0.25

Figure 1: K=1

2.2 2b

For K = 2, we can arbitrarily select the label when two points have different labels. So in this case, testing error is 0



K=2	
Test	Actual
1	1
1	1
-1	-1
-1	-1
Error	0

Figure 2: K=2

2.3 2c

Choose a,b,c = 1, -1, 0.1, respectively. The training error is is 0, and the testing error is 50%. $f(x) = x_1 - x_2 - 0.1$, it satisfies all training dataset, and for the test dataset, the predictions for the first and last one are wrong, and the left are correct. Thus , the test error is 50%.

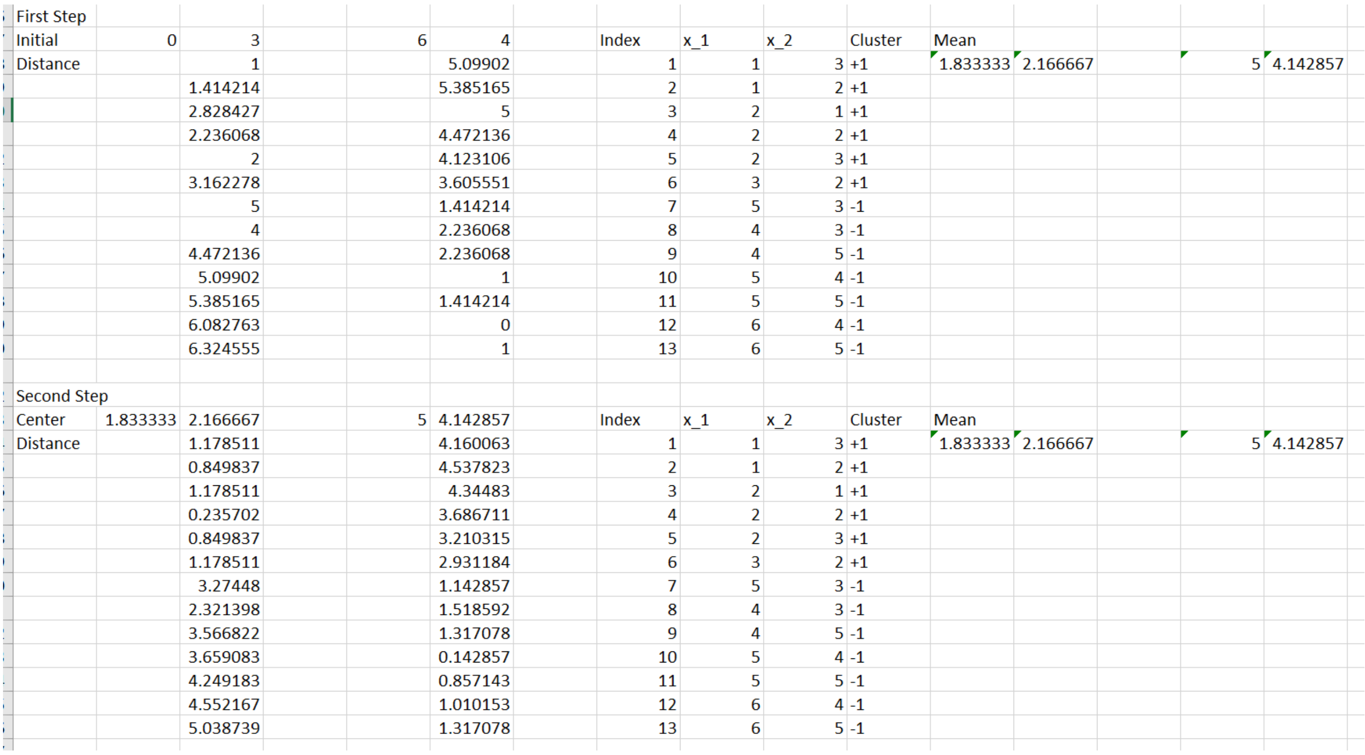
2.4 2d

Based on 2a-2c, we can find
KNN:
1. Do not assume an explicit form for f(X), providing a more flexible approach.
2. The number of cluster is pre-defined which sometimes can not ensure to get the most optimized clusters.
3. Easy to understand and implement.
Linear Classifier:
1. They make strong assumptions about the form of f(X).
2. Suppose we assume a linear relationship between X and Y but the true relationship is far from linear, then the resulting model will provide a poor fit to the data.

3 Question 3

3.1 3a

Based on the calculation, we can find the the mean points of two cluster are (1,8333, 2.1667), (5, 4.14) respectively. and index from 1 to 6 belong to cluster 1, the rest belong to cluster 2.



First Step														
Initial	0	3		6	4	Index	x_1	x_2	Cluster	Mean				
Distance		1			5.09902	1	1	3	+1	1.833333	2.166667		5	4.142857
		1.414214			5.385165	2	1	2	+1					
		2.828427			5	3	2	1	+1					
		2.236068			4.472136	4	2	2	+1					
		2			4.123106	5	2	3	+1					
		3.162278			3.605551	6	3	2	+1					
		5			1.414214	7	5	3	-1					
		4			2.236068	8	4	3	-1					
		4.472136			2.236068	9	4	5	-1					
		5.09902			1	10	5	4	-1					
		5.385165			1.414214	11	5	5	-1					
		6.082763			0	12	6	4	-1					
		6.324555			1	13	6	5	-1					
Second Step														
Center	1.833333	2.166667		5	4.142857	Index	x_1	x_2	Cluster	Mean				
Distance		1.178511			4.160063	1	1	3	+1	1.833333	2.166667		5	4.142857
		0.849837			4.537823	2	1	2	+1					
		1.178511			4.34483	3	2	1	+1					
		0.235702			3.686711	4	2	2	+1					
		0.849837			3.210315	5	2	3	+1					
		1.178511			2.931184	6	3	2	+1					
		3.27448			1.142857	7	5	3	-1					
		2.321398			1.518592	8	4	3	-1					
		3.566822			1.317078	9	4	5	-1					
		3.659083			0.142857	10	5	4	-1					
		4.249183			0.857143	11	5	5	-1					
		4.552167			1.010153	12	6	4	-1					
		5.038739			1.317078	13	6	5	-1					

Figure 3: K=2

3.2 3b

First step, arbitrarily select a point. We can find it satisfy MinPts = 2 and Eps = 1.5. So we say this point is a core point and a cluster is formed.
Then we select another points. There will be three cases.
1. It's a another core point, which means it can form a new cluster.
2. It's a density reachable point from last core point.
3. It's a directly density-reachable from last core point.
After iteration, we can find all points are either density-reachable or density-connected. Therefore, clusters are integrated, and a big cluster is formed for this entire dataset.

3.3 3c

The order is

$$(1,2),3,4,5,6,7,8,9,10,11,12,13$$

$$(1,2),(3,4),5,6,7,8,9,10,11,12,13$$

$$(1,2),(3,4),5,6,(7,8),9,10,11,12,13$$

$$(1,2),(3,4),5,6,(7,8),9,(10,11),12,13$$

$$(1,2),(3,4),5,6,(7,8),9,(10,11),(12,13)$$

$$((1,2),(3,4)),5,6,(7,8),9,(10,11),(12,13)$$

$$(((1,2),(3,4)),5),6,(7,8),9,(10,11),(12,13)$$

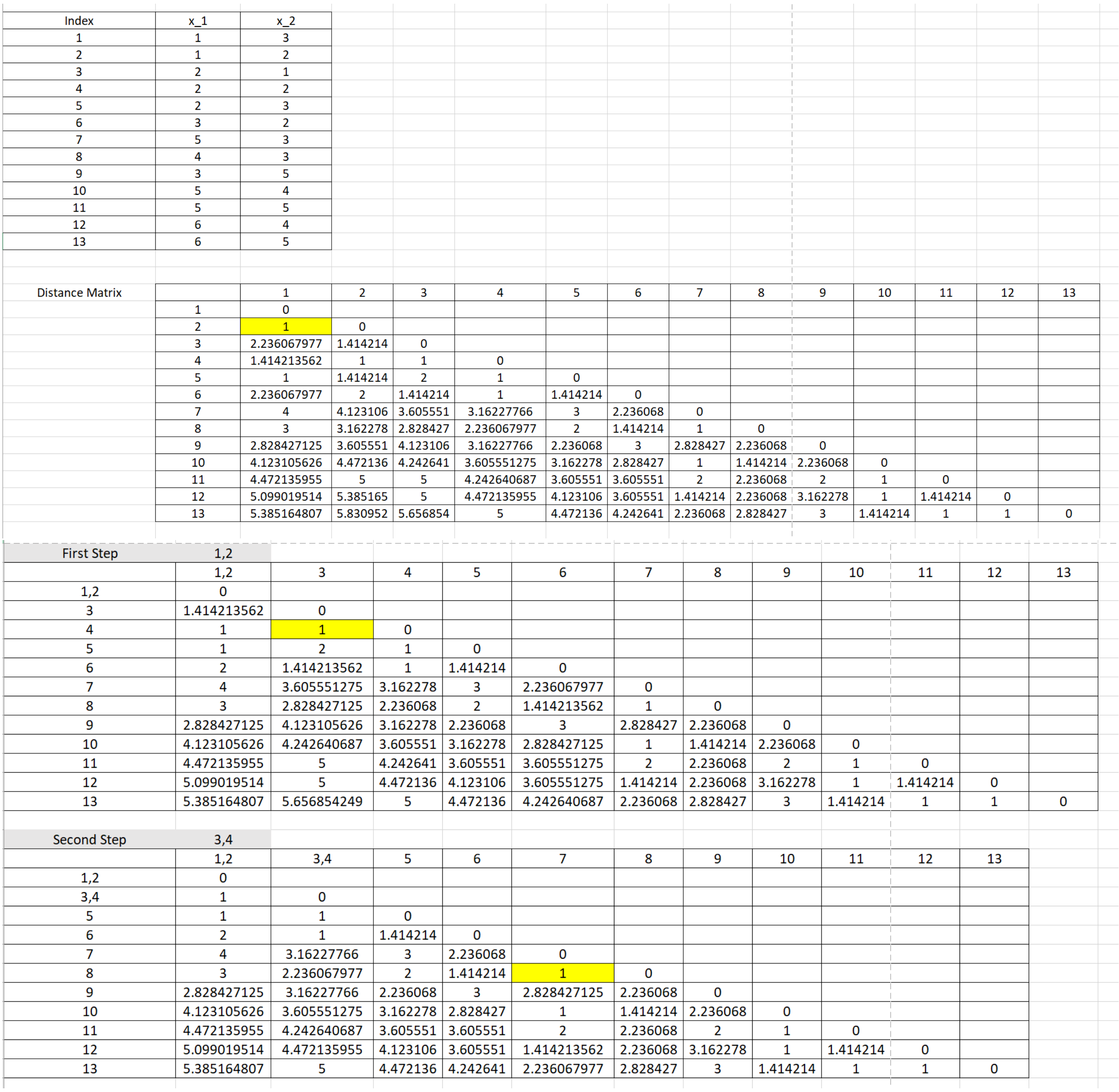
$$((((1,2),(3,4)),5),6),(7,8),9,(10,11),(12,13)$$

$$((((1,2),(3,4)),5),6),(7,8),9,((10,11),(12,13))$$

$$((((1,2),(3,4)),5),6),((7,8),((10,11),(12,13))),9$$

$$(((((1,2),(3,4)),5),6),((7,8),((10,11),(12,13)))),9$$

$$((((((1,2),(3,4)),5),6),((7,8),((10,11),(12,13)))),9)$$



Index	x_1	x_2
1	1	3
2	1	2
3	2	1
4	2	2
5	2	3
6	3	2
7	5	3
8	4	3
9	3	5
10	5	4
11	5	5
12	6	4
13	6	5

Distance Matrix

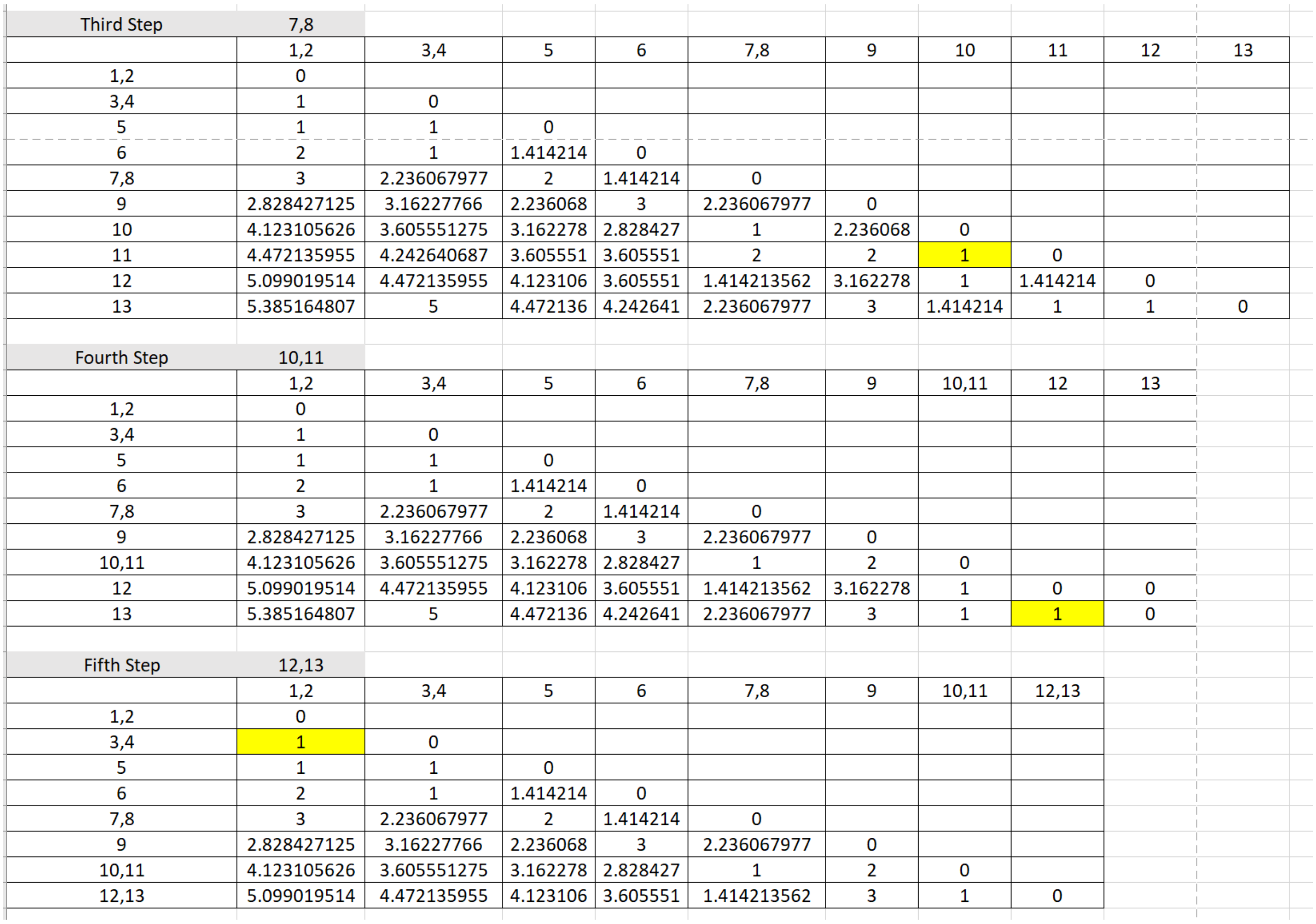
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0												
2	1	0											
3	2.236067977	1.414214	0										
4	1.414213562	1	1	0									
5	1	1.414214	2	1	0								
6	2.236067977	2	1.414214	1	1.414214	0							
7	4	4.123106	3.605551	3.16227766	3	2.236068	0						
8	3	3.162278	2.828427	2.236067977	2	1.414214	1	0					
9	2.828427125	3.605551	4.123106	3.16227766	2.236068	3	2.828427	2.236068	0				
10	4.123105626	4.472136	4.242641	3.605551275	3.162278	2.828427	1	1.414214	2.236068	0			
11	4.472135955	5	5	4.242640687	3.605551	3.605551	2	2.236068	2	1	0		
12	5.099019514	5.385165	5	4.472135955	4.123106	3.605551	1.414214	2.236068	3.162278	1	1.414214	0	
13	5.385164807	5.830952	5.656854	5	4.472136	4.242641	2.236068	2.828427	3	1.414214	1	1	0

First Step 1,2

	1,2	3	4	5	6	7	8	9	10	11	12	13
1,2	0											
3	1.414213562	0										
4	1	1	0									
5	1	2	1	0								
6	2	1.414213562	1	1.414214	0							
7	4	3.605551275	3.162278	3	2.236067977	0						
8	3	2.828427125	2.236068	2	1.414213562	1	0					
9	2.828427125	4.123105626	3.162278	2.236068	3	2.828427	2.236068	0				
10	4.123105626	4.242640687	3.605551	3.162278	2.828427125	1	1.414214	2.236068	0			
11	4.472135955	5	4.242641	3.605551	3.605551275	2	2.236068	2	1	0		
12	5.099019514	5	4.472136	4.123106	3.605551275	1.414214	2.236068	3.162278	1	1.414214	0	
13	5.385164807	5.656854249	5	4.472136	4.242640687	2.236068	2.828427	3	1.414214	1	1	0

Second Step 3,4

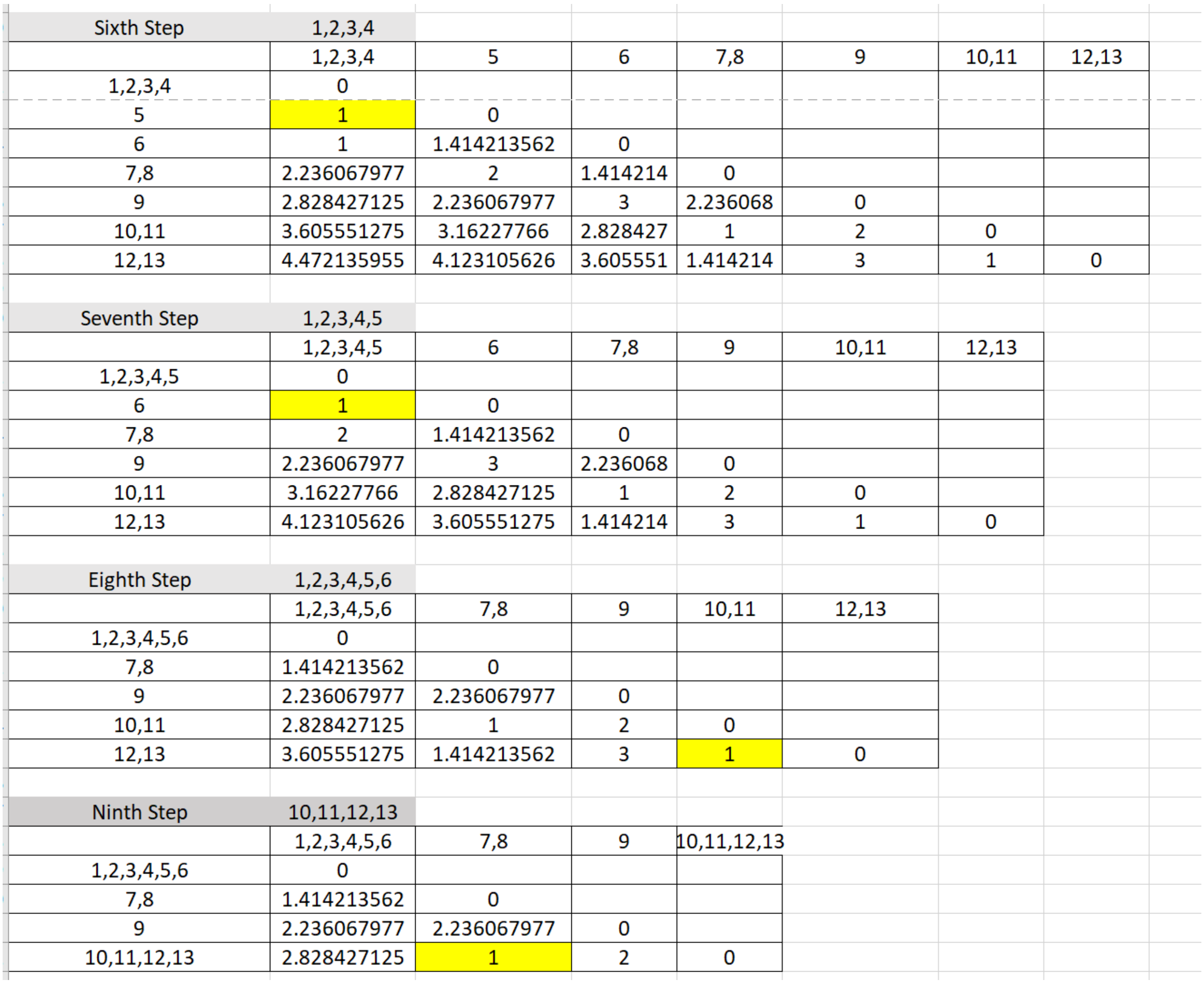
	1,2	3,4	5	6	7	8	9	10	11	12	13
1,2	0										
3,4	1	0									
5	1	1	0								
6	2	1	1.414214	0							
7	4	3.16227766	3	2.236068	0						
8	3	2.236067977	2	1.414214	1	0					
9	2.828427125	3.16227766	2.236068	3	2.828427125	2.236068	0				
10	4.123105626	3.605551275	3.162278	2.828427	1	1.414214	2.236068	0			
11	4.472135955	4.242640687	3.605551	3.605551	2	2.236068	2	1	0		
12	5.099019514	4.472135955	4.123106	3.605551	1.414213562	2.236068	3.162278	1	1.414214	0	
13	5.385164807	5	4.472136	4.242641	2.236067977	2.828427	3	1.414214	1	1	0



Third Step	7,8									
	1,2	3,4	5	6	7,8	9	10	11	12	13
1,2	0									
3,4	1	0								
5	1	1	0							
6	2	1	1.414214	0						
7,8	3	2.236067977	2	1.414214	0					
9	2.828427125	3.16227766	2.236068	3	2.236067977	0				
10	4.123105626	3.605551275	3.162278	2.828427	1	2.236068	0			
11	4.472135955	4.242640687	3.605551	3.605551	2	2	1	0		
12	5.099019514	4.472135955	4.123106	3.605551	1.414213562	3.162278	1	1.414214	0	
13	5.385164807	5	4.472136	4.242641	2.236067977	3	1.414214	1	1	0

Fourth Step	10,11								
	1,2	3,4	5	6	7,8	9	10,11	12	13
1,2	0								
3,4	1	0							
5	1	1	0						
6	2	1	1.414214	0					
7,8	3	2.236067977	2	1.414214	0				
9	2.828427125	3.16227766	2.236068	3	2.236067977	0			
10,11	4.123105626	3.605551275	3.162278	2.828427	1	2	0		
12	5.099019514	4.472135955	4.123106	3.605551	1.414213562	3.162278	1	0	0
13	5.385164807	5	4.472136	4.242641	2.236067977	3	1	1	0

Fifth Step	12,13							
	1,2	3,4	5	6	7,8	9	10,11	12,13
1,2	0							
3,4	1	0						
5	1	1	0					
6	2	1	1.414214	0				
7,8	3	2.236067977	2	1.414214	0			
9	2.828427125	3.16227766	2.236068	3	2.236067977	0		
10,11	4.123105626	3.605551275	3.162278	2.828427	1	2	0	
12,13	5.099019514	4.472135955	4.123106	3.605551	1.414213562	3	1	0

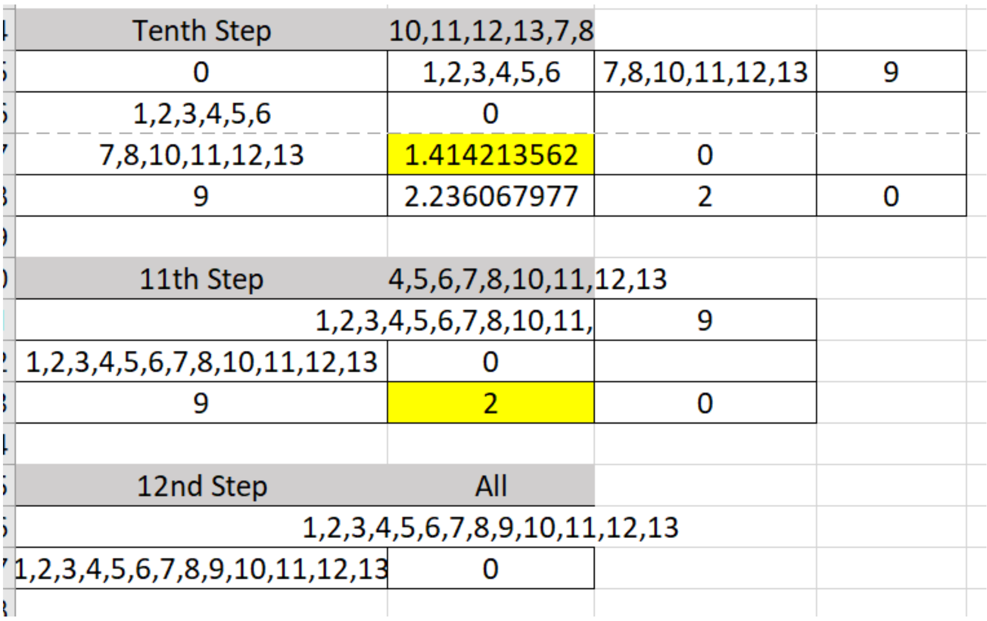


Sixth Step	1,2,3,4						
	1,2,3,4	5	6	7,8	9	10,11	12,13
1,2,3,4	0						
5	1	0					
6	1	1.414213562	0				
7,8	2.236067977	2	1.414214	0			
9	2.828427125	2.236067977	3	2.236068	0		
10,11	3.605551275	3.16227766	2.828427	1	2	0	
12,13	4.472135955	4.123105626	3.605551	1.414214	3	1	0

Seventh Step	1,2,3,4,5					
	1,2,3,4,5	6	7,8	9	10,11	12,13
1,2,3,4,5	0					
6	1	0				
7,8	2	1.414213562	0			
9	2.236067977	3	2.236068	0		
10,11	3.16227766	2.828427125	1	2	0	
12,13	4.123105626	3.605551275	1.414214	3	1	0

Eighth Step	1,2,3,4,5,6				
	1,2,3,4,5,6	7,8	9	10,11	12,13
1,2,3,4,5,6	0				
7,8	1.414213562	0			
9	2.236067977	2.236067977	0		
10,11	2.828427125	1	2	0	
12,13	3.605551275	1.414213562	3	1	0

Ninth Step	10,11,12,13			
	1,2,3,4,5,6	7,8	9	10,11,12,13
1,2,3,4,5,6	0			
7,8	1.414213562	0		
9	2.236067977	2.236067977	0	
10,11,12,13	2.828427125	1	2	0



Tenth Step	10,11,12,13,7,8		
0	1,2,3,4,5,6	7,8,10,11,12,13	9
1,2,3,4,5,6	0		
7,8,10,11,12,13	1.414213562	0	
9	2.236067977	2	0

11th Step	4,5,6,7,8,10,11,12,13	
	1,2,3,4,5,6,7,8,10,11,	9
1,2,3,4,5,6,7,8,10,11,12,13	0	
9	2	0

12nd Step	All
	1,2,3,4,5,6,7,8,9,10,11,12,13
1,2,3,4,5,6,7,8,9,10,11,12,13	0