Northeastern University
CS6220 – Data Mining Techniques
October 2017, Emily Dutile

Homework 2, Part 2

1 K-Means

Algorithm[1]:
Step 1: Choose the number of clusters k (in our case, 3)
Step 2: Make an initial selection of k centroids (this is already done for us)
Step 3: Assign each data element in S to its nearest
centroid (in this way k clusters are formed one for each centroid, where each cluster consists of all the data elements assigned to that centroid)
Step 4: For each cluster make a new selection of its centroid
Step 5: Go back to step 3, repeating the process until the centroids dont change or some other convergence criterion is met

Original Data and First iteration calculations:

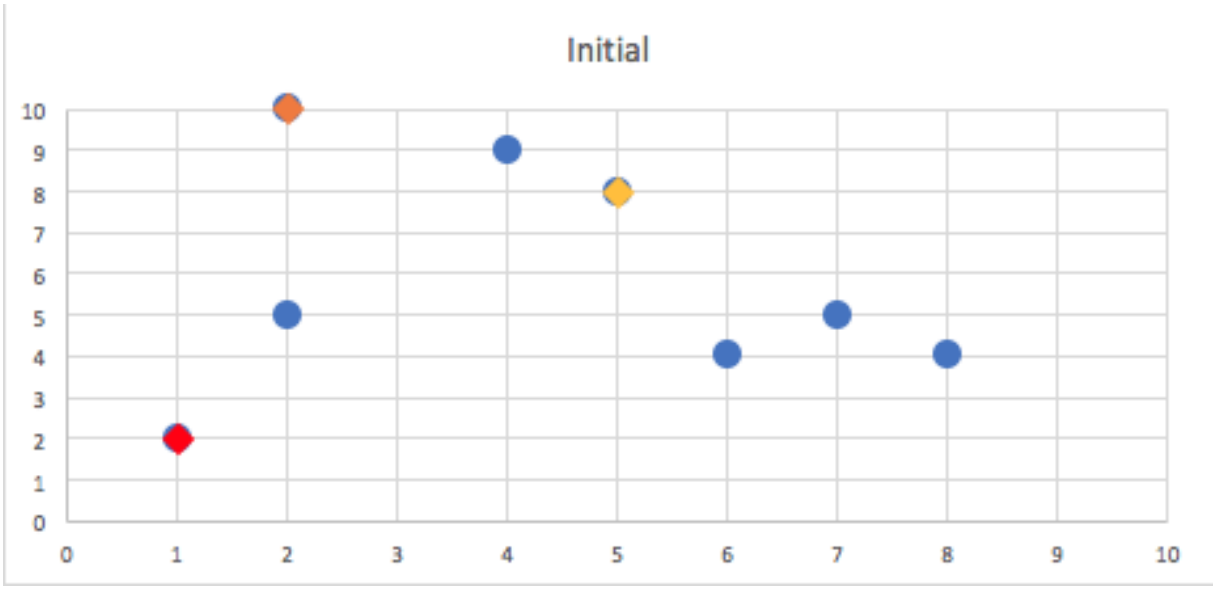
Point	X	Y	Euclidean 1	Euclidean 2	Euclidean 3	Centroid
A	4	9	2.236067977	7.615773106	1.414213562	3
B	2	10	0	8.062257748	3.605551275	1
C	1	2	8.062257748	0	7.211102551	2
D	2	5	5	3.16227766	.242640687	2
E	6	4	7.211102551	5.385164807	4.123105626	3
F	8	4	8.485281374	7.280109889	5	3
G	7	5	7.071067812	6.708203932	3.605551275	3
H	5	8	3.605551275	7.211102551	0	3

Centriod	X	Y
1	2	10
2	1	2
3	5	8

Legend:

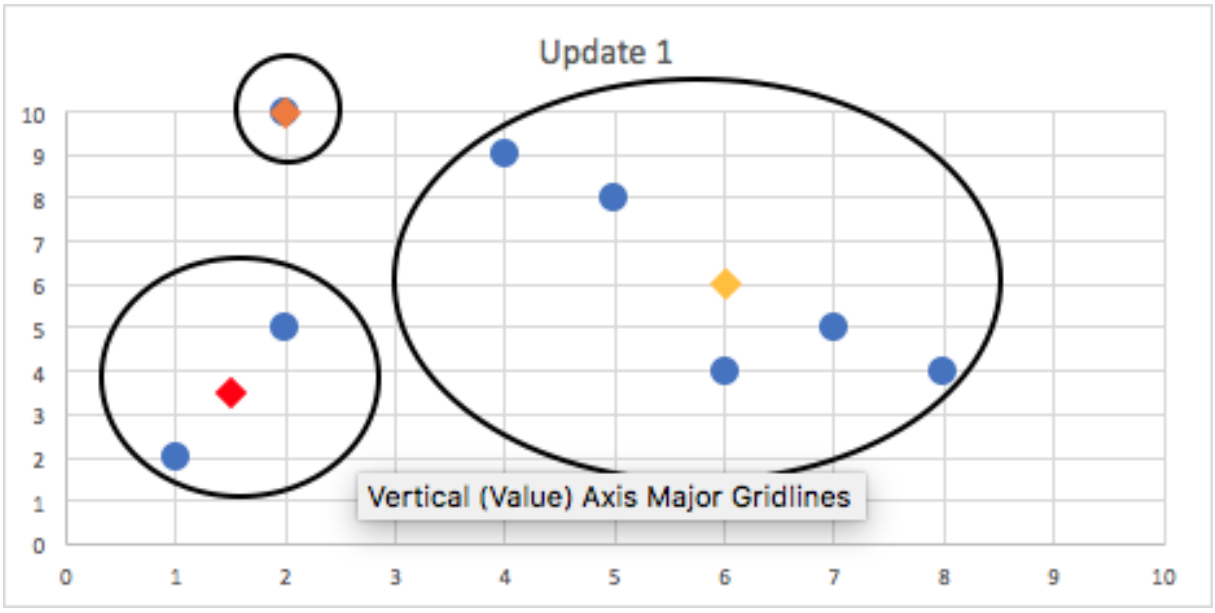
Centroid	Number	Color
Alpha	1	Orange
Beta	2	Red
Gamma	3	Yellow
Data	NA	Blue

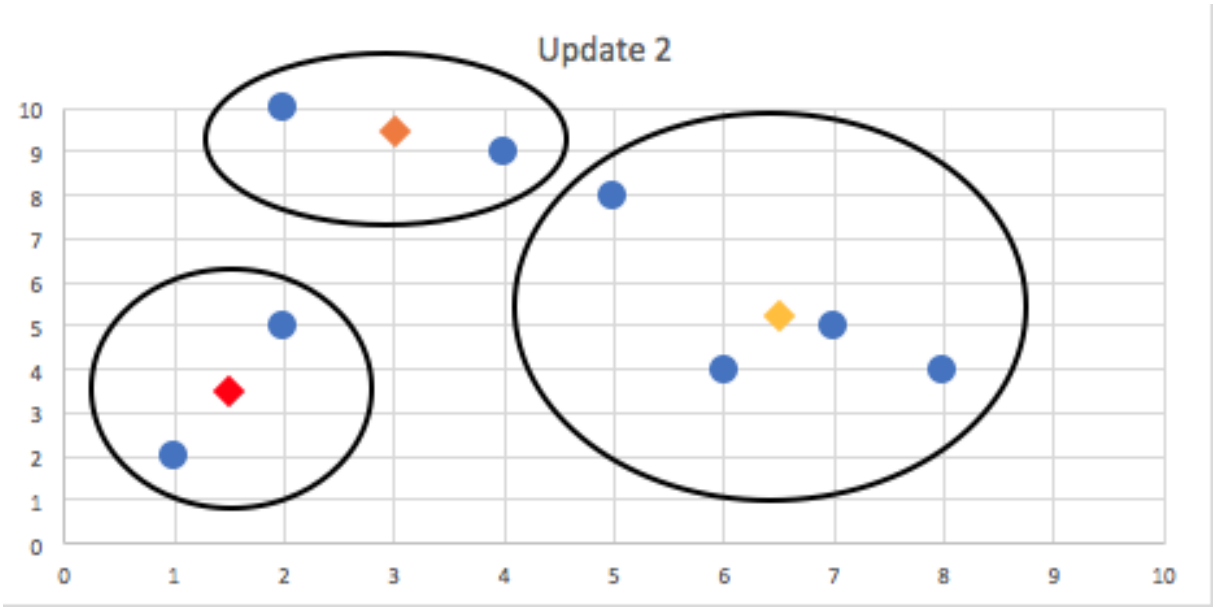
[1] http://www.real-statistics.com/multivariate-statistics/cluster-analysis/k-means-cluster-analysis/

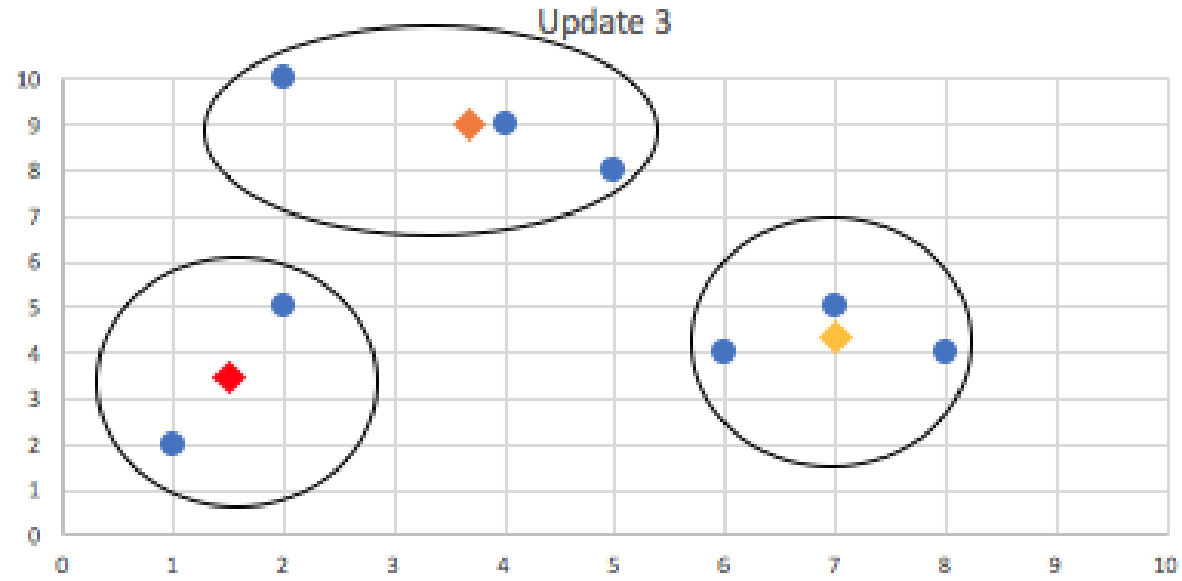


Update 1:

Centriod	X	Y
1	2	10
2	1.5	3.5
3	6	6







2 Agglomerative Hierarchical

MIN

	x	y
p1	0.4	0.53
p2	0.21	0.38
p3	0.35	0.32
p4	0.26	0.19
p5	0.08	0.41
p6	0.45	0.3

	p1	p2	p3	p4	p5	p6
p1						
p2	0.24207437					
p3	0.21587033	0.152315462				
p4	0.36769553	0.196468827	0.15811388			
p5	0.34176015	0.133416641	0.28460499	0.28425341		
p6	0.23537205	0.252982213	0.10198039	0.21954498	0.38600518	

Cluster {3,6}
d({1},{3,6}) = min(d(1,3), d(1,6)) = d(1,3)
d({2},{3,6}) = min(d(2,3), d(2,6)) = d(2,3)
d({4},{3,6}) = min(d(4,3), d(4,6)) = d(4,3)
d({5},{3,6}) = min(d(5,3), d(5,6)) = d(5,3)

	p1	p2	p3,p6	p4	p5
p1					
p2	0.24207437				
p3,p6	0.21587033	0.152315462			
p4	0.36769553	0.196468827	0.15811388		
p5	0.34176015	0.133416641	0.28460499	0.28425341	

Cluster {2,5} d({1},{2,5}) = min(d(1,2), d(1,5)) = d(1,2)
d({3},{2,5}) = min(d(3,2), d(3,5)) = d(3,2)
d({4},{2,5}) = min(d(4,2), d(4,5)) = d(4,2)

	p1	p2,p5	p3,p6	p4
p1				
p2,p5	0.24207437			
p3,p6	0.21587033	0.152315462		
p4	0.36769553	0.196468827	0.15811388	

Cluster {2,3,5,6}
d({1},{2,3,5,6}) = min(d(p1,p5), d(p1,p2), d(p1,p6), d(p1,p3) = d(1,3)
d({4},{2,3,5,6}) = min((p4,p5), d(p4,p2),d(p4,p6), d(p4,p3) = d(3,4)

	p1	p2,p3,p5,p6	p4
p1			
p2,p3,p5,p6	0.21587033		
p4	0.36769553	0.158113883	

Clusters: {1}, {2,3,4,5,6}

MAX

	x	y
p1	0.4	0.53
p2	0.21	0.38
p3	0.35	0.32
p4	0.26	0.19
p5	0.08	0.41
p6	0.45	0.3

	p1	p2	p3	p4	p5	p6
p1						
p2	0.24207437					
p3	0.21587033	0.15231546				
p4	0.36769553	0.19646883	0.15811388			
p5	0.34176015	0.13341664	0.28460499	0.28425341		
p6	0.23537205	0.25298221	0.10198039	0.21954498	0.38600518	

Cluster {3,6}
d({1},{3,6}) = max(d(1,3), d(1,6)) = d(1,6)
d({2},{3,6}) = max(d(2,3), d(2,6)) = d(2,6)
d({4},{3,6}) = max(d(4,3), d(4,6)) = d(4,6)
d({5},{3,6}) = max(d(5,3), d(5,6)) = d(5,6)

	p1	p2	p3,p6	p4	p5
p1					
p2	0.24207437				
p3,6	0.23537205	0.25298221			
p4	0.36769553	0.19646883	0.21954498		
p5	0.34176015	0.13341664	0.38600518	0.28425341	

Cluster {2,5}
d({1},{2,5}) = max(d(1,2), d(1,5)) = d(1,5)
d({4},{2,5}) = max(d(2,4), d(4,5)) = d(4,5)
d({6},{2,5}) = max(d(2,6), d(5,6)) = d(2,6)

	p1	p2,p5	p3,p6	p4
p1				
p2,p5	0.34176015			
p3,6	0.23537205	0.38600518		
p4	0.36769553	0.28425341	0.21954498	

Cluster {3,4,6}
d({1}, {3,4,6}) = max(d(1,4), d(1,6), d(1, 3)) = d(1,4])
d({2,5},{3,4,6}) = max(d(4,2), d(6,2), d({3},{2}), d(4,5), d(6,5), d(p3,p5)) = d(5,6])

	p1	p2,p5	p3,p4,p6
p1			
p2,p5	0.34176015		
p3,6,p4	0.36769553	0.38600518	

Clusters: {1,2,5}, {2,5,6}

AVG

	x	y
p1	0.4	0.53
p2	0.21	0.38
p3	0.35	0.32
p4	0.26	0.19
p5	0.08	0.41
p6	0.45	0.3

	p1	p2	p3	p4	p5	p6
p1						
p2	0.24207437					
p3	0.21587033	0.15231546				
p4	0.36769553	0.19646883	0.15811388			
p5	0.34176015	0.13341664	0.28460499	0.28425341		
p6	0.23537205	0.25298221	0.10198039	0.21954498	0.38600518	

Cluster {3,6}

	p1	p2	p4	p5	p3,p6
p1					
p2	0.24207437				
p4	0.36769553	0.19646883			
p5	0.34176015	0.13341664	0.28425341		
p3,p6	0.22	0.20248457	0.18439089	0.33526109	

Cluster {2,5}

	p1	p4	p2,p5	p3,p6
p1				
p4	0.36769553			
p2,p5	0.28853076	0.23505319		
p3,p6	0.22	0.18439089	0.2687936	

Cluster {4,6}

	p1	p2,p5	p3,p4,p6
p1			
p2,p5	0.28853076		
p3,p6,p4	0.2729793	0.24295633	

Clusters: {1}, {2,3,4,5,6}

3 DBSCAN

DBSCAN (= 7.5, MinPts = 3)

pt 0: 2 < MinPts, so cluster=-1

pt 1: 3 ≥ MinPts, so cluster=0 to_visit=[40, 75], visited=1

- pt 40: cluster=0, 3 ≥ MinPts, so adding neighbors to_visit=[75, 28], visited=1, 40
- pt 75: cluster=0, 3 ≥ MinPts, so adding neighbors to_visit=[28, 4], visited=1, 40, 75
- pt 28: cluster=0, 3 ≥ MinPts, so adding neighbors to_visit=[4, 12], visited=1, 28, 40, 75
- pt 4: cluster=0, 3 ≥ MinPts, so adding neighbors to_visit=[12, 56], visited=1, 4, 28, 40, 75
- pt 12: cluster=0, 2 < MinPts, to_visit=[56], visited=1, 4, 12, 28, 40, 75
- pt 56: cluster=0, 3 ≥ MinPts, so adding neighbors to_visit=[66], visited=1, 4, 12, 28, 40, 56, 75
- pt 66: cluster=0, 2 < MinPts to_visit=[], visited=1, 4, 12, 28, 40, 56, 66, 75

pt 2: 1 < MinPts, so cluster=-1

pt 3: 2 < MinPts, so cluster=-1

pt 4: cluster=0, so skip

pt 5: 3 ≥ MinPts, so cluster=1
to_visit=[70, 74], visited=5:

- pt 70: cluster=1, 5 ≥ MinPts, so adding neighbors to_visit=[74, 32, 69, 72], visited=5, 70
- pt 74: cluster=1, 4 ≥ MinPts, so adding neighbors to_visit=[32, 69, 72, 19, 54], visited=5, 70, 74
- pt 32: cluster=1, 5 ≥ MinPts, so adding neighbors to_visit=[69, 72, 19, 54, 63, 69], visited=5, 32, 70, 74
- pt 69: cluster=1, 4 ≥ MinPts, so adding neighbors to_visit=[72, 19, 54, 63], visited=5, 32, 69, 70, 74
- pt 72: cluster=1, 7 ≥ MinPts, so adding neighbors to_visit=[19, 54, 63, 8, 60], visited=5, 32, 69, 70, 72, 74
- pt 19: cluster=1, 3 ≥ MinPts, so adding neighbors to_visit=[54, 63, 8, 60], visited=5, 19, 32, 69, 70, 72, 74
- pt 54: cluster=1, 4 ≥ MinPts, so adding neighbors to_visit=[63, 8, 60, 25], visited=5, 19, 32, 54, 69, 70, 72, 74
- pt 63: cluster=1, 7 ≥ MinPts, so adding neighbors to_visit=[8, 60, 25], visited=5, 19, 32, 54, 63, 69, 70, 72, 74
- pt 8: cluster=1, 5 ≥ MinPts, so adding neighbors to_visit=[60, 25, 11], visited=5, 8, 19, 32, 54, 63, 69, 70, 72, 74

- pt 60: cluster=1, 6 $\geq$ MinPts, so adding neighbors to_visit=[25, 11, 50, 68], visited=5, 8, 19, 32, 54, 60, 63, 69, 70, 72, 74
- pt 25: cluster=1, 4 $\geq$ MinPts, so adding neighbors to_visit=[11, 50, 68, 26, 67], visited=5, 8, 19, 25, 32, 54, 60, 63, 69, 70, 72, 74
- pt 11: cluster=1, 3 $\geq$ MinPts, so adding neighbors to_visit=[50, 68, 26, 67, 14], visited=5, 8, 11, 19, 25, 32, 54, 60, 63, 69, 70, 72, 74
- pt 50: cluster=1, 5 $\geq$ MinPts, so adding neighbors to_visit=[68, 26, 67, 14, 39], visited=5, 8, 11, 19, 25, 32, 50, 54, 60, 63, 69, 70, 72, 74
- pt 68: cluster=1, 5 $\geq$ MinPts, so adding neighbors to_visit=[26, 67, 14, 39], visited=5, 8, 11, 19, 25, 32, 50, 54, 60, 63, 68, 69, 70, 72, 74
- pt 26: cluster=1, 3 $\geq$ MinPts, so adding neighbors to_visit=[67, 14, 39, 34], visited=5, 8, 11, 19, 25, 26, 32, 50, 54, 60, 63, 68, 69, 70, 72, 74
- pt 67: cluster=1, 2 $<$ MinPts, to_visit=[14, 39, 34], visited=5, 8, 11, 19, 25, 26, 32, 50, 54, 60, 63, 67, 68, 69, 70, 72, 74
- pt 14: cluster=1, 3 $\geq$ MinPts, so adding neighbors to_visit=[39, 34, 6], visited=5, 8, 11, 14, 19, 25, 26, 32, 50, 54, 60, 63, 68, 69, 70, 72, 74
- pt 39: cluster=1, 3 $\geq$ MinPts, so adding neighbors to_visit=[34, 6, 10, 71], visited=5, 8, 11, 14, 19, 25, 26, 32, 39, 50, 54, 60, 63, 68, 69, 70, 72, 74
- pt 34: cluster=1, 4 $\geq$ MinPts, so adding neighbors to_visit=[6, 10, 71, 29, 46], visited=5, 8, 11, 14, 19, 25, 26, 32, 34, 39, 50, 54, 60, 63, 68, 69, 70, 72, 74
- pt 6: cluster=1, 3 $\geq$ MinPts, so adding neighbors to_visit=[10, 71, 29, 46, 42], visited=5, 6, 8, 11, 14, 19, 25, 26, 32, 34, 39, 50, 54, 60, 63, 68, 69, 70, 72, 74
- pt 10: cluster=1, 3 $\geq$ MinPts, so adding neighbors to_visit=[71, 29, 46, 42, 22], visited=5, 6, 8, 10, 11, 14, 19, 25, 26, 32, 34, 39, 50, 54, 60, 63, 68, 69, 70, 72, 74
- pt 71: cluster=1, 4 $\geq$ MinPts, so adding neighbors to_visit=[29, 46, 42, 22], visited=5, 6, 8, 10, 11, 14, 19, 25, 26, 32, 34, 39, 50, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 29: cluster=1, 4 $\geq$ MinPts, so adding neighbors to_visit=[46, 42, 22, 16], visited=5, 6, 8, 10, 11, 14, 19, 25, 26, 29, 32, 34, 39, 50, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 46: cluster=1, 4 $\geq$ MinPts, so adding neighbors to_visit=[42, 22, 16], visited=5, 6, 8, 10, 11, 14, 19, 25, 26, 29, 32, 34, 39, 46, 50, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 42: cluster=1, 4 $\geq$ MinPts, so adding neighbors to_visit=[22, 16, 17, 20], visited=5, 6, 8, 10, 11, 14, 19, 25, 26, 29, 32, 34, 39, 42, 46, 50, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 22: cluster=1, 3 $\geq$ MinPts, so adding neighbors to_visit=[16, 17, 20], visited=5, 6, 8, 10, 11, 14, 19, 22, 25, 26, 29, 32, 34, 39, 42, 46, 50, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 16: cluster=1, 4 $\geq$ MinPts, so adding neighbors to_visit=[17, 20, 48], visited=5, 6, 8, 10, 11, 14, 16, 19, 22, 25, 26, 29, 32, 34, 39, 42, 46, 50, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 17: cluster=1, 3 $\geq$ MinPts, so adding neighbors to_visit=[20, 48], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 22, 25, 26, 29, 32, 34, 39, 42, 46, 50, 54, 60, 63, 68, 69, 70, 71, 72, 74

- pt 20: cluster=1, 4 ≥ MinPts, so adding neighbors to_visit=[48, 38], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 20, 22, 25, 26, 29, 32, 34, 39, 42, 46, 50, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 48: cluster=1, 2 < MinPts, to_visit=[38], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 20, 22, 25, 26, 29, 32, 34, 39, 42, 46, 48, 50, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 38: cluster=1, 5 ≥ MinPts, so adding neighbors to_visit=[30, 37, 45], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 20, 22, 25, 26, 29, 32, 34, 38, 39, 42, 46, 48, 50, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 30: cluster=1, 4 ≥ MinPts, so adding neighbors to_visit=[37, 45, 52], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 20, 22, 25, 26, 29, 30, 32, 34, 38, 39, 42, 46, 48, 50, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 37: cluster=1, 4 ≥ MinPts, so adding neighbors to_visit=[45, 52, 53], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 20, 22, 25, 26, 29, 30, 32, 34, 37, 38, 39, 42, 46, 48, 50, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 45: cluster=1, 4 ≥ MinPts, so adding neighbors to_visit=[52, 53], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 20, 22, 25, 26, 29, 30, 32, 34, 37, 38, 39, 42, 45, 46, 48, 50, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 52: cluster=1, 4 ≥ MinPts, so adding neighbors to_visit=[53, 49, 64], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 20, 22, 25, 26, 29, 30, 32, 34, 37, 38, 39, 42, 45, 46, 48, 50, 52, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 53: cluster=1, 3 ≥ MinPts, so adding neighbors to_visit=[49, 64, 47], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 20, 22, 25, 26, 29, 30, 32, 34, 37, 38, 39, 42, 45, 46, 48, 50, 52, 53, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 49: cluster=1, 4 ≥ MinPts, so adding neighbors to_visit=[64, 47, 31, 76], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 20, 22, 25, 26, 29, 30, 32, 34, 37, 38, 39, 42, 45, 46, 48, 49, 50, 52, 53, 54, 60, 63, 68, 69, 70, 71, 72, 74
- pt 64: cluster=1, 2 < MinPts, to_visit=[47, 31, 76], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 20, 22, 25, 26, 29, 30, 32, 34, 37, 38, 39, 42, 45, 46, 48, 49, 50, 52, 53, 54, 60, 63, 64, 68, 69, 70, 71, 72, 74
- pt 47: cluster=1, 2 < MinPts, to_visit=[31, 76], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 20, 22, 25, 26, 29, 30, 32, 34, 37, 38, 39, 42, 45, 46, 47, 48, 49, 50, 52, 53, 54, 60, 63, 64, 68, 69, 70, 71, 72, 74
- pt 31: cluster=1, 2 < MinPts, to_visit=[76], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 20, 22, 25, 26, 29, 30, 31, 32, 34, 37, 38, 39, 42, 45, 46, 47, 48, 49, 50, 52, 53, 54, 60, 63, 64, 68, 69, 70, 71, 72, 74
- pt 76: cluster=1, 3 ≥ MinPts, so adding neighbors to_visit=[21], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 20, 22, 25, 26, 29, 30, 31, 32, 34, 37, 38, 39, 42, 45, 46, 47, 48, 49, 50, 52, 53, 54, 60, 63, 64, 68, 69, 70, 71, 72, 74
- pt 21: cluster=1, 3 ≥ MinPts, so adding neighbors to_visit=[], visited=5, 6, 8, 10, 11, 14, 16, 17, 19, 20, 21, 22, 25, 26, 29, 30, 31, 32, 34, 37, 38, 39, 42, 45, 46, 47, 48, 49, 50, 52, 53, 54, 60, 63, 64, 68, 69, 70, 71, 72, 74

pt 6: cluster=1, so skip

pt 7: 1 < MinPts, so cluster=-1

pt 8: cluster=1, so skip

pt 9: 3 ≥ MinPts, so cluster=2, to_visit=[33, 78], visited=9

- pt 33: cluster=2, 3 $\geq$ MinPts, so adding neighbors to_visit=[78], visited=9, 33
- pt 78: cluster=2, 3 $\geq$ MinPts, so adding neighbors to_visit=[], visited=9, 33, 78

pt 10: cluster=1, so skip

pt 11: cluster=1, so skip

pt 12: cluster=0, so skip

pt 13: 2 $<$ MinPts, so cluster=-1

pt 14: cluster=1, so skip

pt 15: 1 $<$ MinPts, so cluster=-1

pt 16: cluster=1, so skip

pt 17: cluster=1, so skip

pt 18: 1 $<$ MinPts, so cluster=-1

pt 19: cluster=1, so skip

pt 20: cluster=1, so skip

pt 21: cluster=1, so skip

pt 22: cluster=1, so skip

pt 23: 1 $<$ MinPts, so cluster=-1

pt 24: 1 $<$ MinPts, so cluster=-1

pt 25: cluster=1, so skip

pt 26: cluster=1, so skip

pt 27: 2 $<$ MinPts, so cluster=-1

pt 28: cluster=0, so skip

pt 29: cluster=1, so skip

pt 30: cluster=1, so skip

pt 31: cluster=1, so skip

pt 32: cluster=1, so skip

pt 33: cluster=2, so skip

pt 34: cluster=1, so skip

pt 35: 2 < MinPts, so cluster=-1

pt 36: 1 < MinPts, so cluster=-1

pt 37: cluster=1, so skip

pt 38: cluster=1, so skip

pt 39: cluster=1, so skip

pt 40: cluster=0, so skip

pt 41: 1 < MinPts, so cluster=-1

pt 42: cluster=1, so skip

pt 43: 2 < MinPts, so cluster=-1

pt 44: 1 < MinPts, so cluster=-1

pt 45: cluster=1, so skip

pt 46: cluster=1, so skip

pt 47: cluster=1, so skip

pt 48: cluster=1, so skip

pt 49: cluster=1, so skip

pt 50: cluster=1, so skip

pt 51: 2 < MinPts, so cluster=-1

pt 52: cluster=1, so skip

pt 53: cluster=1, so skip

pt 54: cluster=1, so skip

pt 55: 2 < MinPts, so cluster=-1

pt 56: cluster=0, so skip

pt 57: 1 < MinPts, so cluster=-1

pt 58: 1 < MinPts, so cluster=-1

pt 59: 2 < MinPts, so cluster=-1

pt 60: cluster=1, so skip

pt 61: 1 < MinPts, so cluster=-1

pt 62: 2 < MinPts, so cluster=-1

pt 63: cluster=1, so skip

pt 64: cluster=1, so skip

pt 65: 1 < MinPts, so cluster=-1

pt 66: cluster=0, so skip

pt 67: cluster=1, so skip

pt 68: cluster=1, so skip

pt 69: cluster=1, so skip

pt 70: cluster=1, so skip

pt 71: cluster=1, so skip

pt 72: cluster=1, so skip

pt 73: 1 < MinPts, so cluster=-1

pt 74: cluster=1, so skip

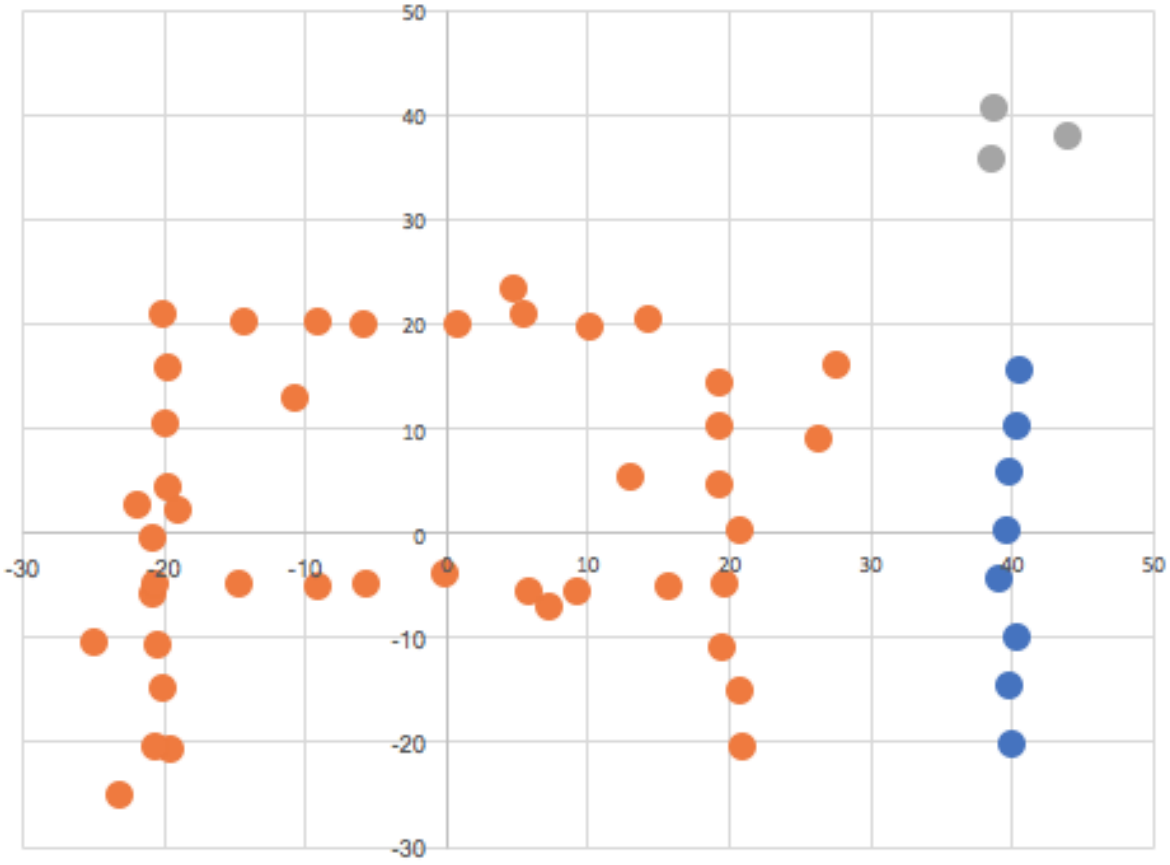
pt 75: cluster=0, so skip

pt 76: cluster=1, so skip

pt 77: 2 < MinPts, so cluster=-1

pt 78: cluster=2, so skip

pt 79: 1 < MinPts, so cluster=-1



4 Extra Credit

Magazine: AI Magazine, Vol 38, No 3: Fall 2017 (current)
Article: Steps Toward Robust Artificial Intelligence
Author: Thomas G. Dietterich
Employment: Oregon State University
Facts: One of the founders of ML, Born in MA, Has obtained over 30 million in research grants