1 Clustering Algorithm

To better understand how a clustering algorithm works, lets manually examine some of the single linkage procedures calculation steps. We start off by looking at the initial (Euclidean) distance matrix displayed previously.

Objects	A	B	C	D	E	F	G
A	0						
B	3	0					
C	2.236	1.414	0				
D	2	3.606	2.236	0			
E	3.606	2	1.414	3	0		
F	4.123	4.472	3.162	2.236	2.828	0	
G	5.385	7.071	5.657	3.606	5.831	3.162	0

- In the very first step, the two objects exhibiting the smallest distance in the matrix are merged. Note that we always merge those objects with the smallest distance, regardless of the clustering procedure (e.g., single or complete linkage). (N.B. In the following example, ties will be broken at random.)
- As we can see, this happens to two pairs of objects, namely B and C (d(B, C) = 1.414), as well as C and E (d(C, E) = 1.414). In the next step, we will see that it does not make any difference whether we first merge the one or the other, so lets proceed by forming a new cluster, using objects B and C.

Objects	A	B, C	D	E	F	G
A	0					
B, C	2.236	0				
D	2	2.236	0			
E	3.606	1.414	3	0		
F	4.123	3.162	2.236	2.828	0	
G	5.385	5.657	3.606	5.831	3.162	0

- Having made this decision, we then form a new distance matrix by considering the single linkage decision rule as discussed above. According to this rule, the distance from, for example, object A to the newly formed cluster is the minimum of $d(A, B)$ and $d(A, C)$. As $d(A, C)$ is smaller than d(A, B), the distance from A to the newly formed cluster is equal to d(A, C); that is, 2.236.
- We also compute the distances from cluster [B,C] (clusters are indicated by means of squared brackets) to all other objects (i.e. D, E, F, G) and simply copy the remaining distances such as $d(E, F)$ that the previous clustering has not affected.
- Continuing the clustering procedure, we simply repeat the last step by merging the objects in the new distance matrix that exhibit the smallest distance (in this case, the newly

Objects	A	B, C, E	D	F	G
A	0				
B, C, E	2.236	0			
D	2	2.236	0		
F	4.123	2.828	2.236	0	
G	5.385	5.657	3.606	3.162	0

formed cluster [B, C] and object E) and calculate the distance from this cluster to all other objects.

Objects	A, D	B, C, E	F	G
A, D	0			
B, C, E	2.236	0		
F	2.236	2.828	0	
G	3.606	5.657	3.162	0

- We continue in the same fashion until one cluster is left. By following the single linkage procedure, the last steps involve the merger of cluster [A,B,C,D,E,F] and object G at a distance of 3.162.

Objects	A, B, C, D, E	F	G
A, B, C, D, E	0		
F	2.236	0	
G	3.606	3.162	0

Objects	A, B, C, D, E, F	G
A, B, C, D, E, F	0	
G	3.162	0

2 Clustering Algorithm

To better understand how a clustering algorithm works, lets manually examine some of the single linkage procedures calculation steps. We start off by looking at the initial (Euclidean) distance matrix displayed previously.

Objects	A	B	C	D	E	F	G
A	0						
B	3	0					
C	2.236	1.414	0				
D	2	3.606	2.236	0			
E	3.606	2	1.414	3	0		
F	4.123	4.472	3.162	2.236	2.828	0	
G	5.385	7.071	5.657	3.606	5.831	3.162	0

- In the very first step, the two objects exhibiting the smallest distance in the matrix are merged. Note that we always merge those objects with the smallest distance, regardless of the clustering procedure (e.g., single or complete linkage). (N.B. In the following example, ties will be broken at random.)
- As we can see, this happens to two pairs of objects, namely B and C (d(B, C) = 1.414), as well as C and E (d(C, E) = 1.414). In the next step, we will see that it does not make any difference whether we first merge the one or the other, so lets proceed by forming a new cluster, using objects B and C.

Objects	A	B, C	D	E	F	G
A	0					
B, C	2.236	0				
D	2	2.236	0			
E	3.606	1.414	3	0		
F	4.123	3.162	2.236	2.828	0	
G	5.385	5.657	3.606	5.831	3.162	0

- Having made this decision, we then form a new distance matrix by considering the single linkage decision rule as discussed above. According to this rule, the distance from, for example, object A to the newly formed cluster is the minimum of d(A, B) and d(A, C). As d(A, C) is smaller than d(A, B), the distance from A to the newly formed cluster is equal to d(A, C); that is, 2.236.
- We also compute the distances from cluster [B,C] (clusters are indicated by means of squared brackets) to all other objects (i.e. D, E, F, G) and simply copy the remaining distances such as d(E, F) that the previous clustering has not affected.
- Continuing the clustering procedure, we simply repeat the last step by merging the objects in the new distance matrix that exhibit the smallest distance (in this case, the newly

Objects	A	B, C, E	D	F	G
A	0				
B, C, E	2.236	0			
D	2	2.236	0		
F	4.123	2.828	2.236	0	
G	5.385	5.657	3.606	3.162	0

formed cluster [B, C] and object E) and calculate the distance from this cluster to all other objects.

Objects	A, D	B, C, E	F	G
A, D	0			
B, C, E	2.236	0		
F	2.236	2.828	0	
G	3.606	5.657	3.162	0

- We continue in the same fashion until one cluster is left. By following the single linkage procedure, the last steps involve the merger of cluster [A,B,C,D,E,F] and object G at a distance of 3.162.

Objects	A, B, C, D, E	F	G
A, B, C, D, E	0		
F	2.236	0	
G	3.606	3.162	0

Objects	A, B, C, D, E, F	G
A, B, C, D, E, F	0	
G	3.162	0