

## Question 2:

Code attached to show working

Input of Node 4	0.04	Error at Node 6	0.052
Output of Node 4	0.51	Error at Node 4	0.005
Input of Node 5	0.15	Error at Node 5	0.003
Output of Node 5	0.5374	Updated $w_{46}$	0.42
Input of Node 6	0.3384	Updated $w_{56}$	0.27
Output of Node 6	0.5838	Updated $w_{14}$	0.352
Updated $w_{25}$	-0.099	Updated $w_{24}$	0.1512
Updated $w_{35}$	0.2022	Updated $w_{34}$	-0.1965

## Question 3(d):

One of the algorithms which does not require us to mention  $K$  (i.e. the number of clusters) before clustering is hierarchical clustering. As the name suggests, the clusters obtained after this algorithm are in the form of hierarchy, and thus looking at the hierarchy, we decide how many clusters to obtain.

The algorithm starts with considering each node as a separate cluster, and then,

1. Compute dissimilarity matrix, i.e. how much each cluster is away from another cluster.
2. Merge the two clusters how are the least dissimilar (closest together, speaking in terms of distance)
3. Repeat until all clusters have been merged into one. In this algorithm we have to keep record at what iteration data points are merged. This information is displayed in the form of dendrogram, which shows the level of merging of each cluster.

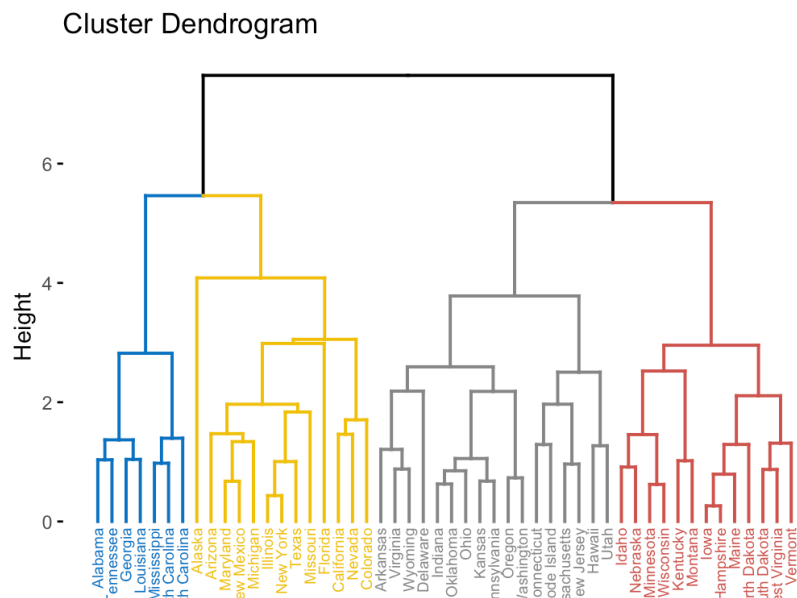


Figure 1, displays an example dendrogram. The height at which two clusters are merged illustrates the dissimilarity of each cluster. The lower the height, more similar they are to each other. After such a dendrogram is obtained, we decide at what height to cut the dendrogram so as to obtain clusters.

This algorithm requires us to specify,

- a. The dissimilarity measure (i.e. Euclidean Distance, or Manhattan Distance)
- b. The measure to decide the closest clusters (since the cluster has more than one point).  
Commonly, single linkage or complete linkage is used.

The advantage this method provides over kmeans is that it provides a visual representation of the data before deciding how many clusters to obtain, hence we decide after looking at the representation. Obvious disadvantage is, it takes much more time to compute, and similarly, the advantage is highly sensitive to the methods used to compute the closest cluster.