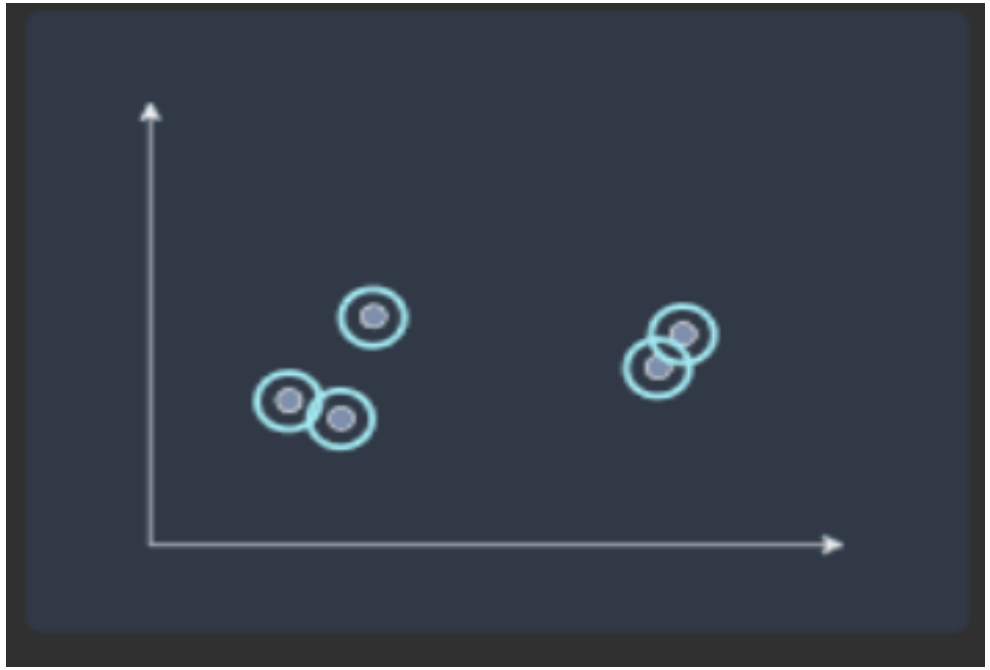
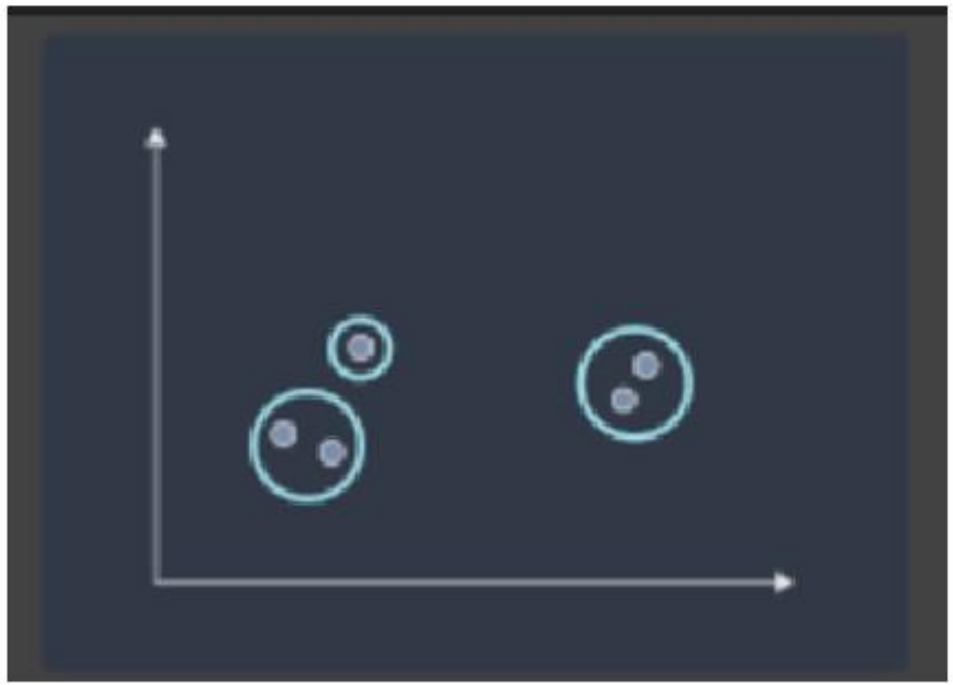
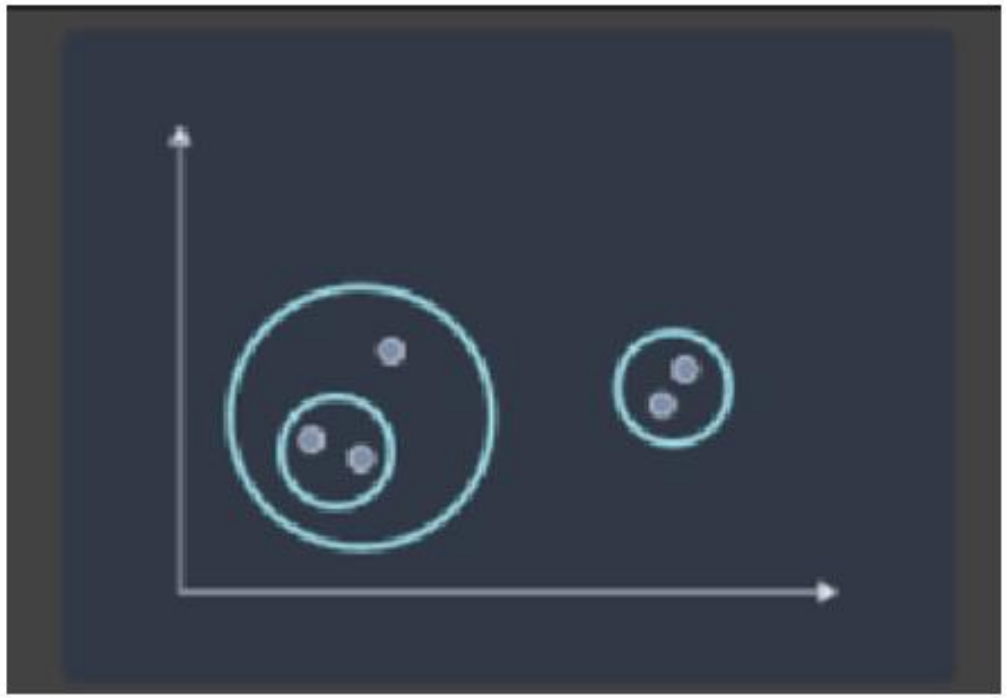
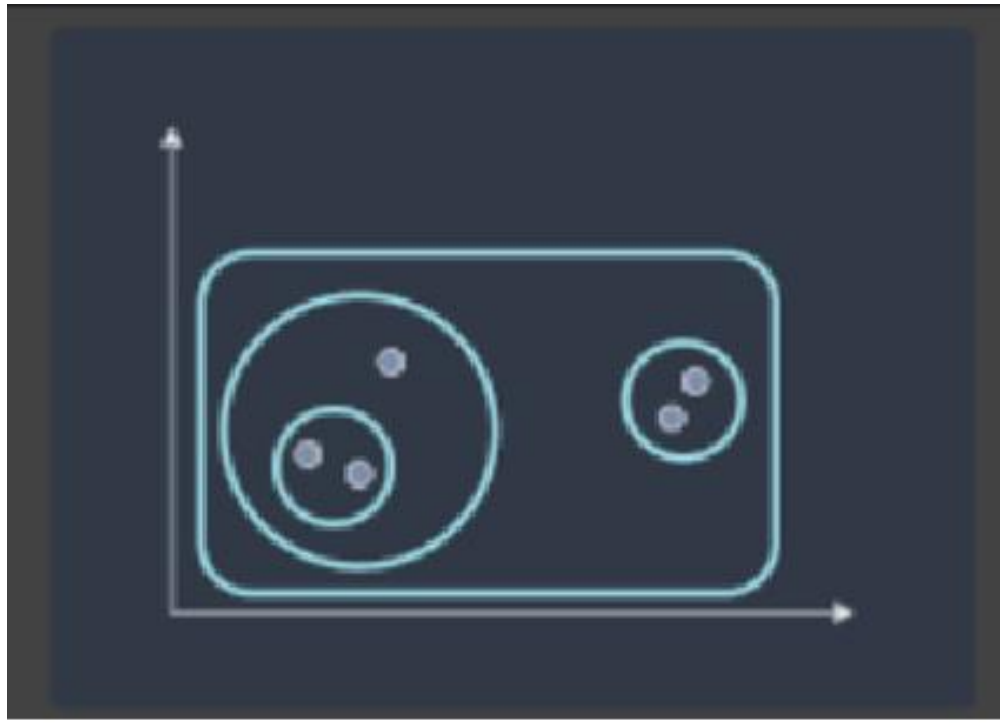
Hierarchical Clustering

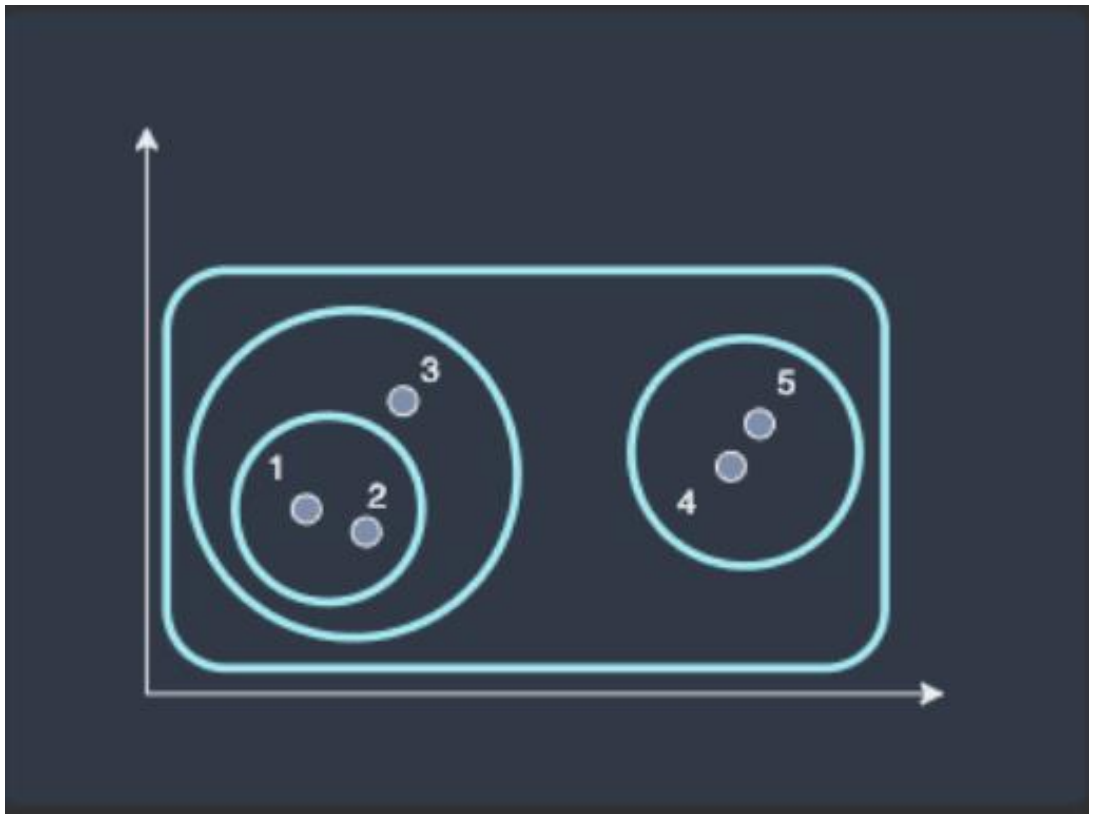
Agglomerative (bottom-up approach)

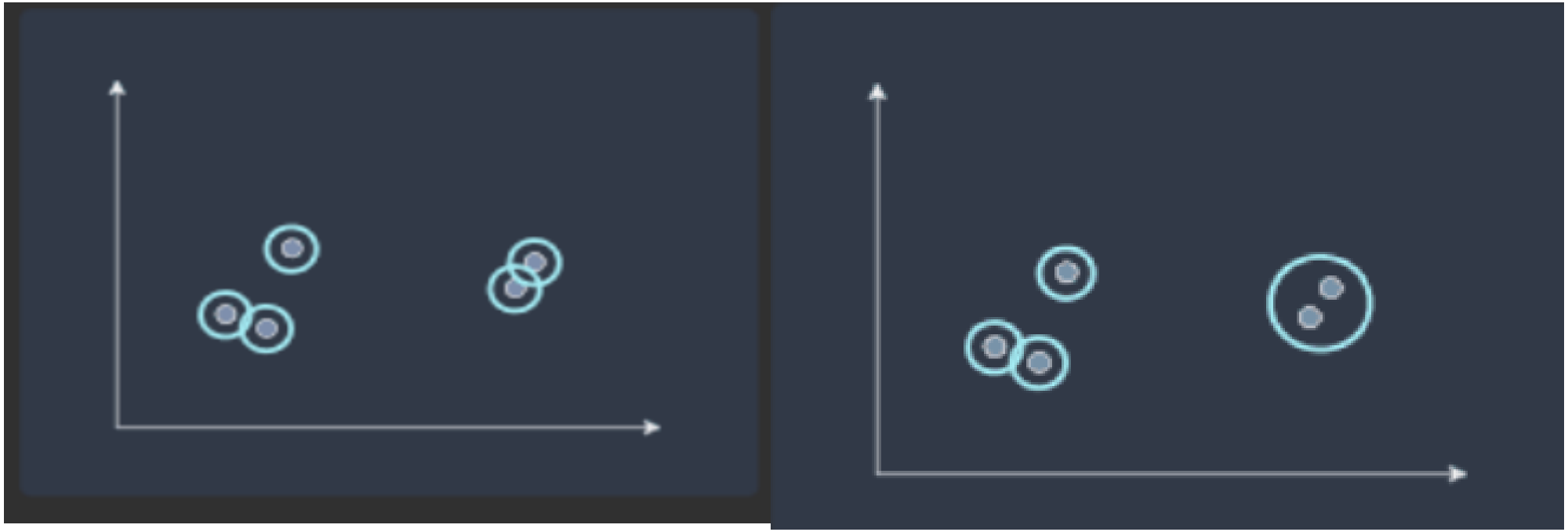


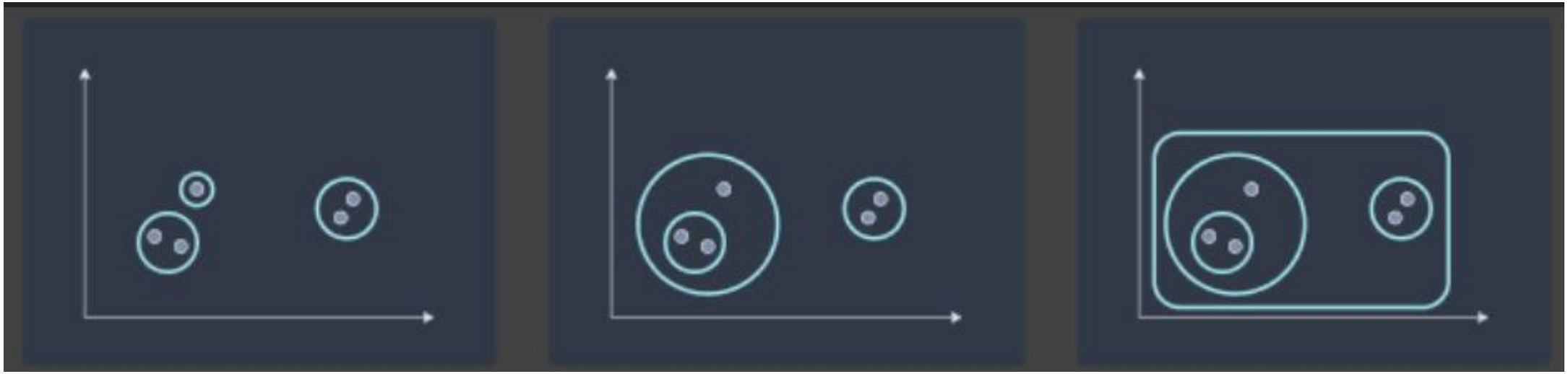







3
1
2
5
4





Distance Computation

1. Euclidean distance

2. Manhattan distance

Euclidean distance

Measures the straight-line distance between two points

For two points A(x1,y1) and B(x2,y2), the Euclidean distance *d* is calculated as:

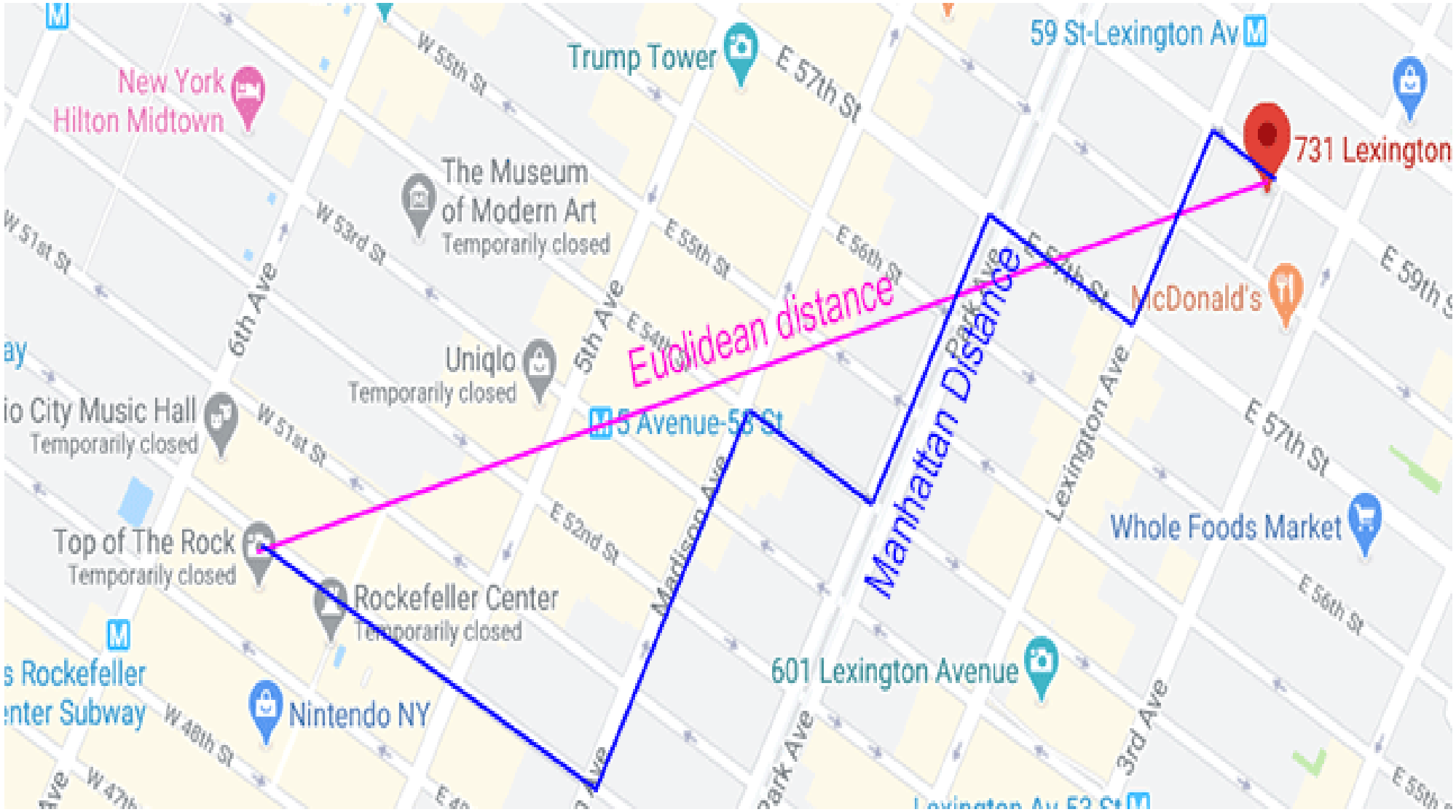
$$d(A,B)=(x2-x1)2+(y2-y1)2$$

Manhattan distance (taxicab distance)

Measures the distance between two points in a grid-based system

For two points A(x1,y1) and B(x2,y2), the Manhattan distance d is calculated as:

$$d(A,B)=|x1-x2|+|y1-y2|$$


New York Hilton Midtown
W 55th St
Trump Tower
E 57th St
59 St-Lexington Av
731 Lexington
The Museum of Modern Art
Temporarily closed
W 53rd St
E 55th St
E 56th St
W 51st St
6th Ave
5th Ave
Euclidean distance
Manhattan Distance
Park Ave
McDonald's
E 59th St
Uniqlo
Temporarily closed
5 Avenue-53 St
Lexington Ave
E 57th St
io City Music Hall
Temporarily closed
W 51st St
Top of The Rock
Temporarily closed
Rockefeller Center
Temporarily closed
E 52nd St
Madison Ave
Whole Foods Market
E 56th St
601 Lexington Avenue
3rd Ave
s Rockefeller
nter Subway
W 48th St
Nintendo NY
E 55th St

Linking methods between clusters

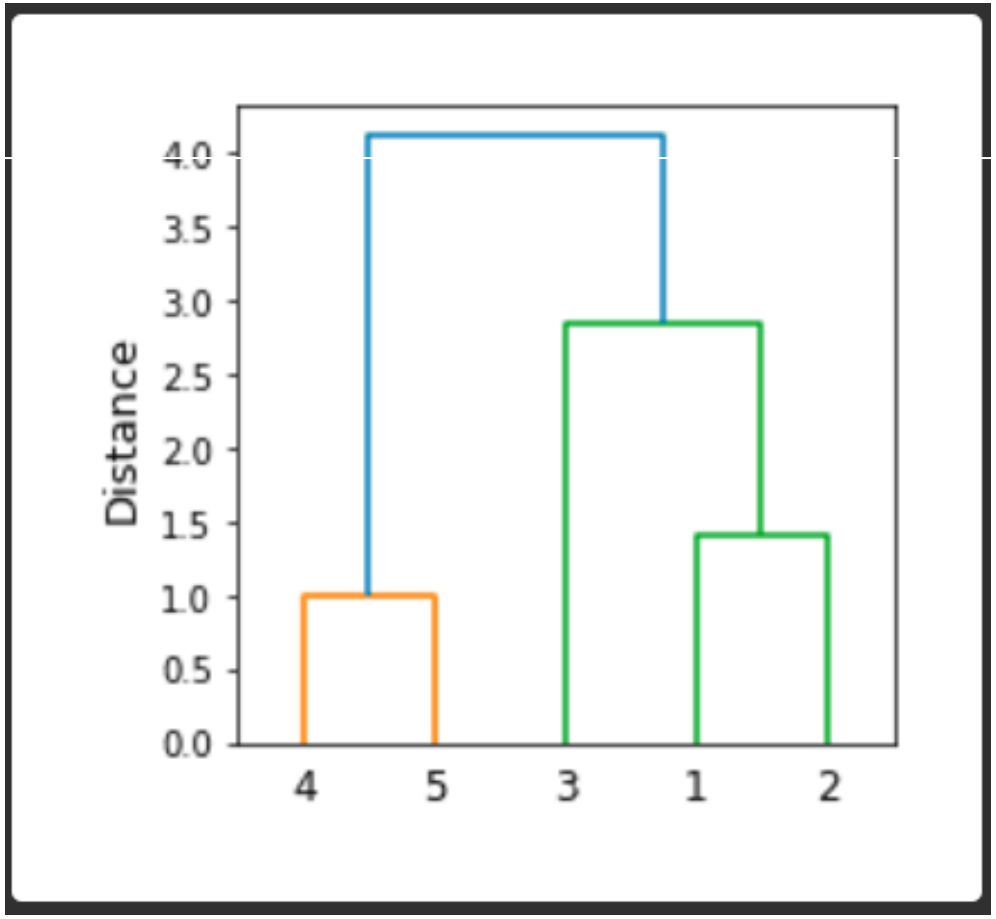
Single linkage minimal distance between a data point in cluster A and a data point in cluster B

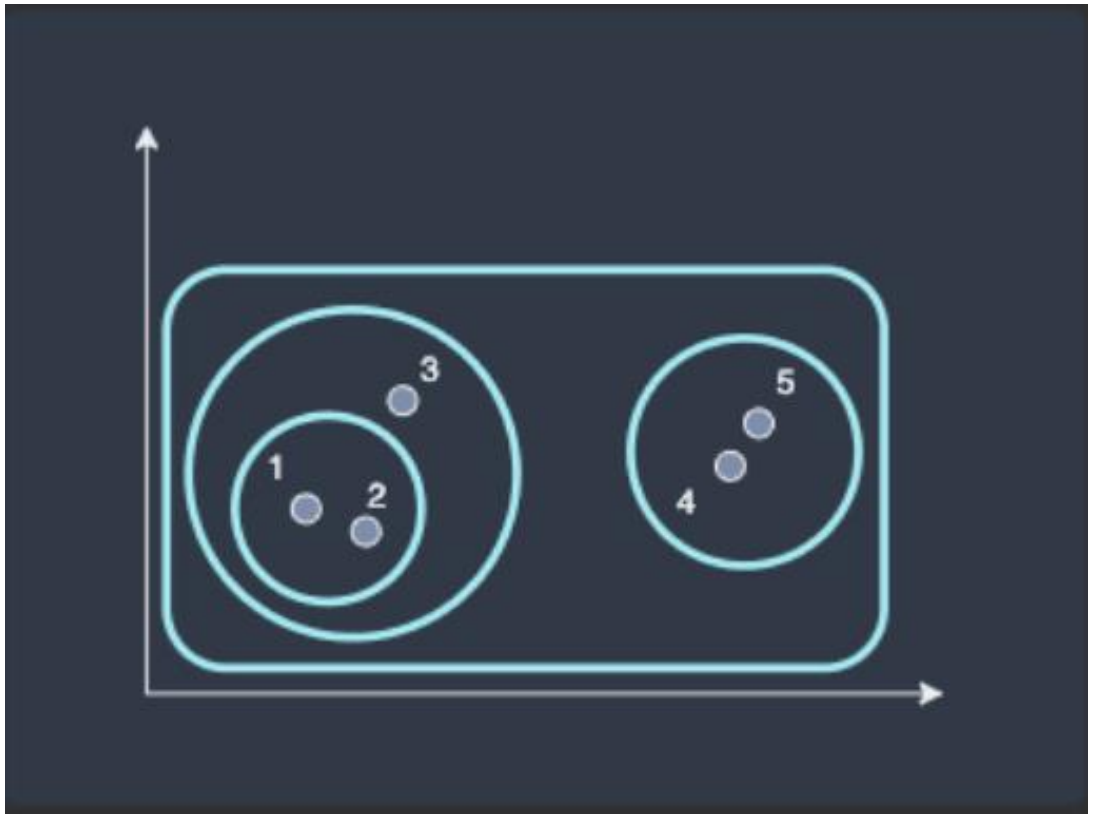
Maximum linkage maximum distance between a data point in cluster A and a data point in cluster B

Average linkage the average distance between all pairs of points in clusters A and B:

Ward Linkage the within-cluster sum of squares (WSS) of the merged cluster (A,B):

Choosing cluster using Dendogram





Choosing cluster using Dendogram