Hierarchical and K-means Clustering

RUCHI KAPADIWALA
002772936

Hierarchical

K-means

**01** Hierarchical methods can be either divisive or agglomerative.

k-means, using a pre-specified number of clusters, the method assigns records to each cluster to find the mutually exclusive cluster of spherical shape based on distance.

02 In hierarchical clustering one can stop at any number of clusters, one find appropriate by interpreting the dendrogram.

K Means clustering needed advance knowledge of K i.e. no. of clusters one want to divide your data.

03 Agglomerative methods begin with 'n' clusters and sequentially combine similar clusters until only one cluster is obtained.

One can use median or mean as a cluster centre to represent each cluster.

04 In Hierarchical Clustering, results are reproducible in Hierarchical clustering.

In K Means clustering, since one start with random choice of clusters, the results produced by running the algorithm many times may differ.

05 A hierarchical clustering is a set of nested clusters that are arranged as a tree.

K Means clustering is found to work well when the structure of the clusters is hyper spherical (like circle in 2D, sphere in 3D).

06 Hierarchical clustering don't work as well as, k means when the shape of the clusters is hyper spherical.
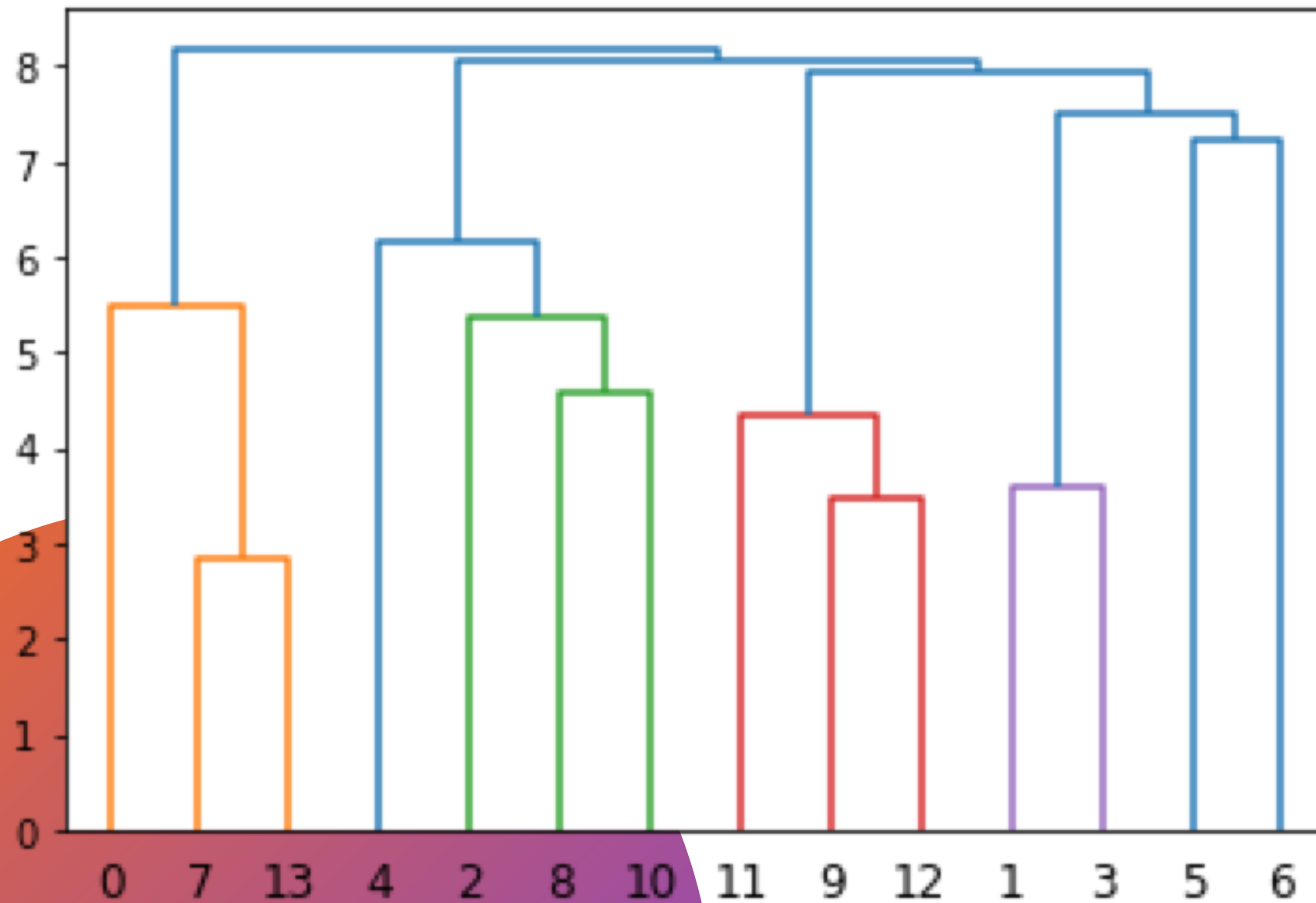
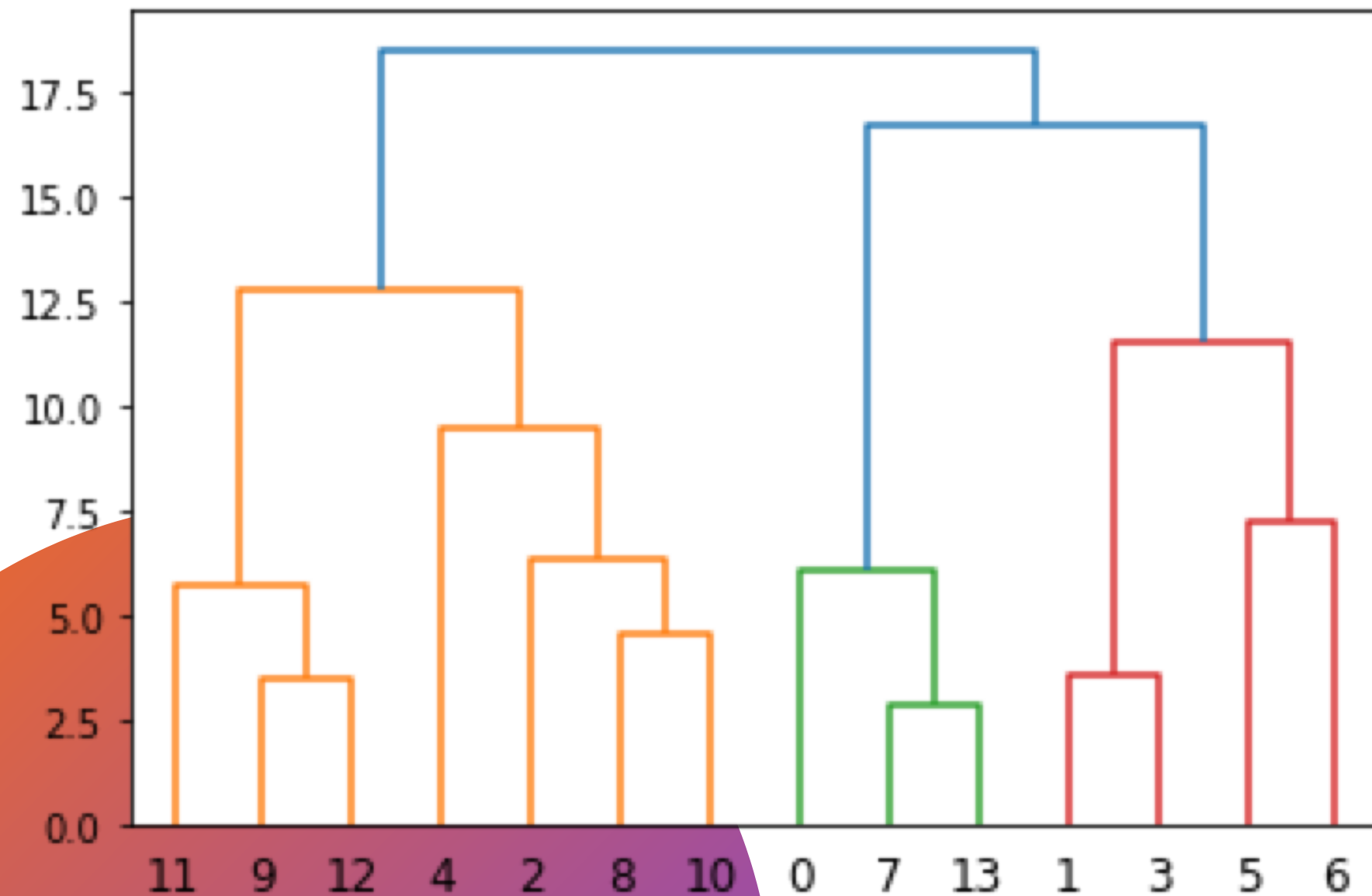# Types of Linkages in Clustering

Single Linkage:

For two clusters R and S, the single linkage returns the minimum distance between two points i and j such that i belongs to R and j belongs to S.
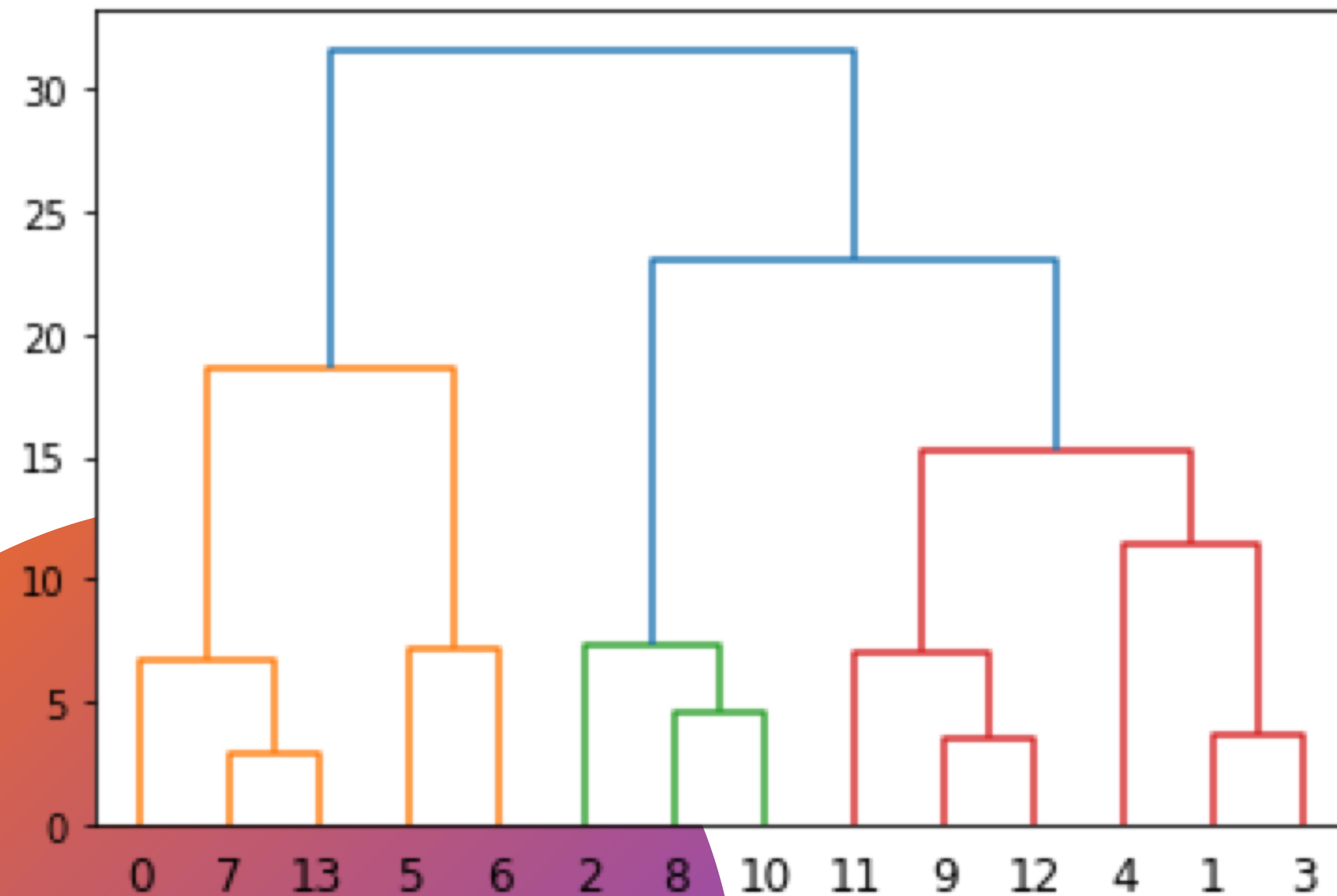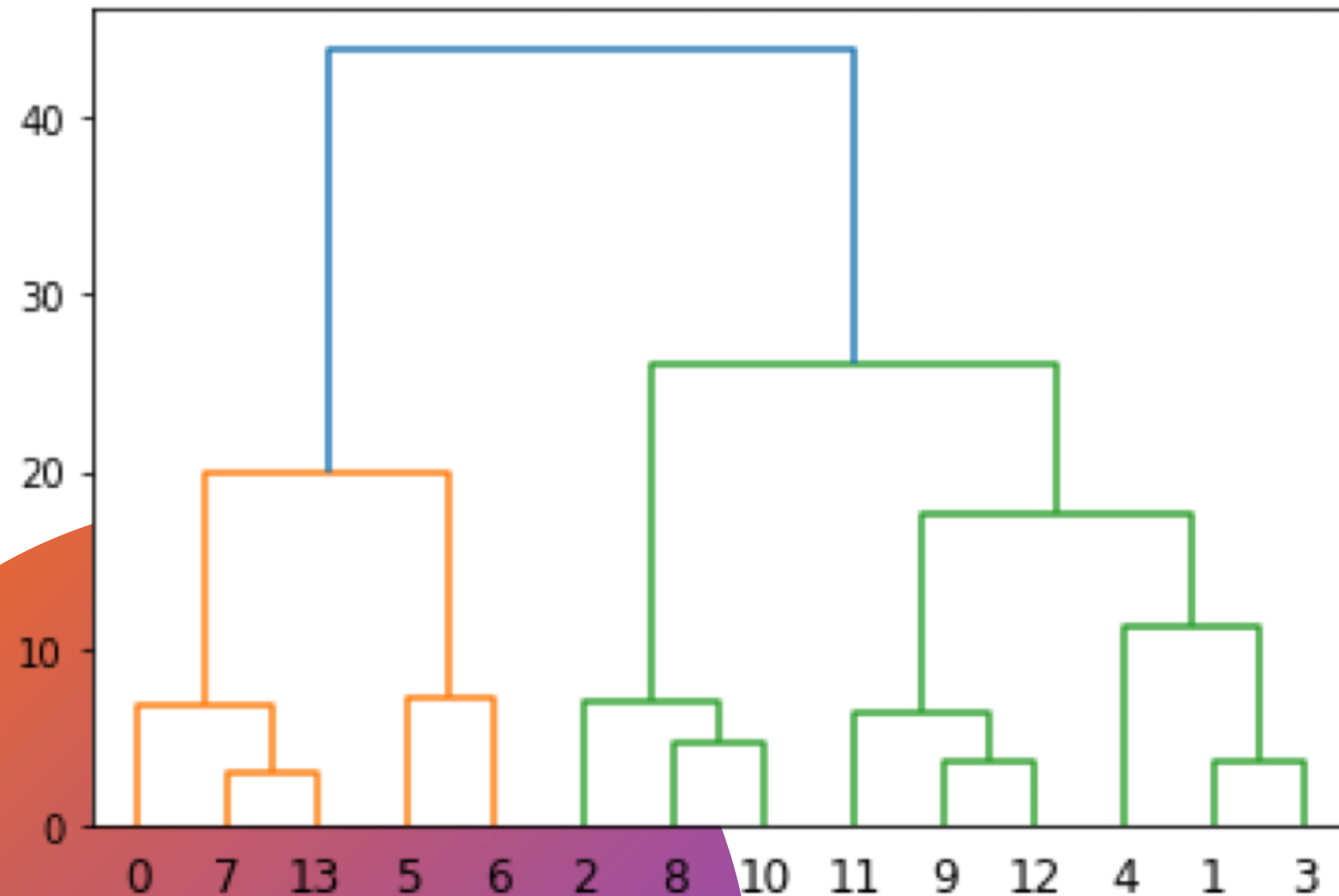
Types of Linkages in Clustering

Complete Linkage:

For two clusters R and S, the complete linkage returns the maximum distance between two points i and j such that i belongs to R and j belongs to S.

Types of Linkages in Clustering

Average Linkage:

For two clusters R and S, first for the distance between any data-point i in R and any data-point j in S and then the arithmetic mean of these distances are calculated. Average Linkage returns this value of the arithmetic mean.

Types of Linkages in Clustering

Ward linkage:

The Ward approach analyzes the variance of the clusters rather than measuring distances directly, minimizing the variance between clusters.
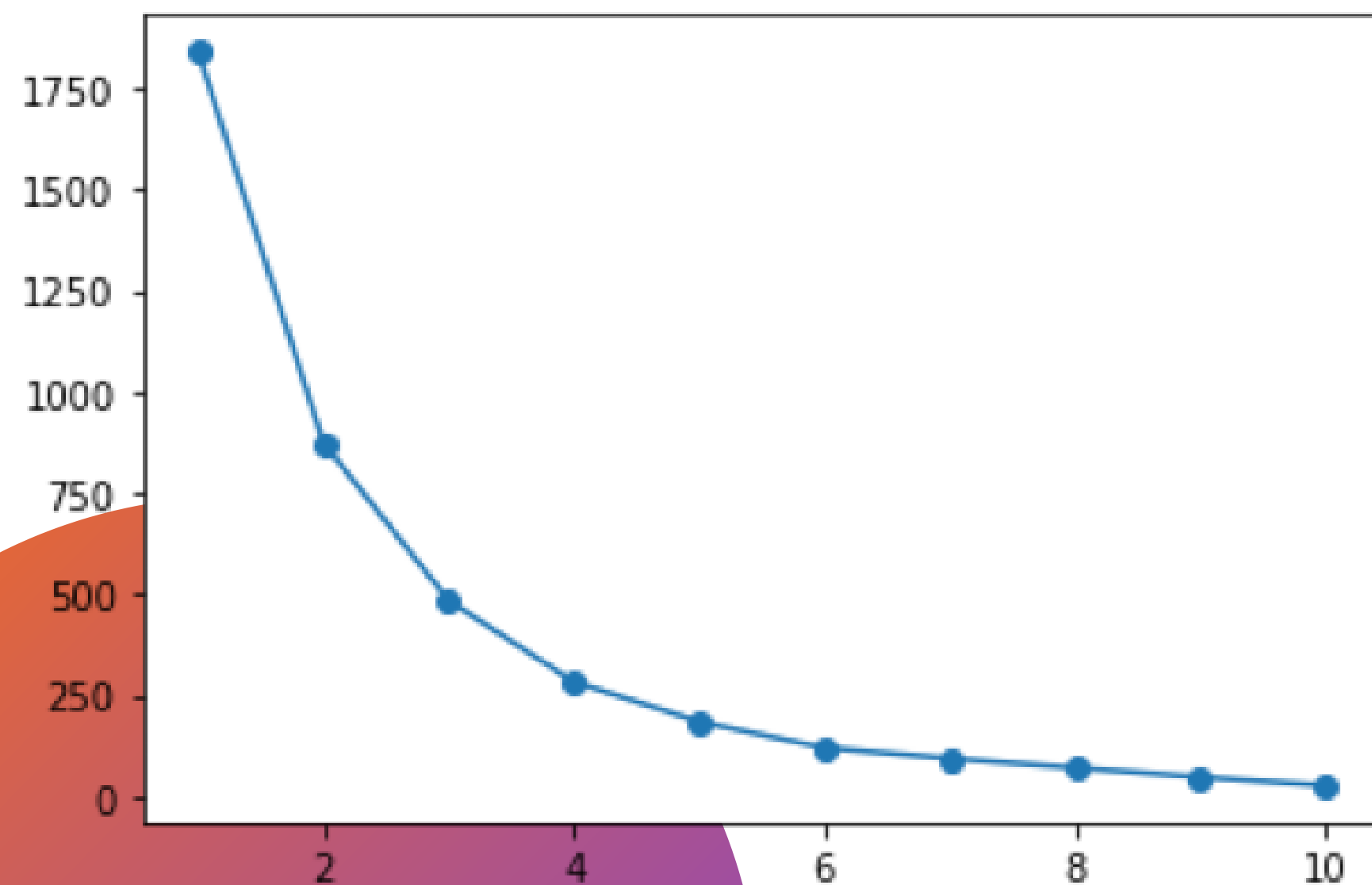
Elbow Method

The Elbow method is a very popular technique, and the idea is to run k-means clustering for a range of clusters k (let's say from 1 to 10) and for each value, we are calculating the sum of squared distances from each point to its assigned center(distortions).