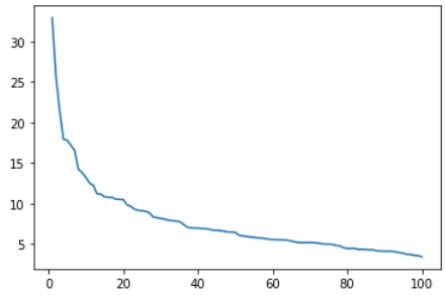
Hierarchical clustering using IRIS dataset

1. Input Iris data (except the labels)
2. Perform “single”, “average” linkage with the help of scipy.cluster.hierarchy package
3. Selection of the number of clusters
    a. Implement “elbow” method



Identify the clustering step where the acceleration of distance growth is the biggest.

4. Display dendogram
5. Evaluate the dendogram

Cophenetic correlation coefficient of the dendogram

$$c = \frac{\sum_{i<j}(x(i,j) - \bar{x})(t(i,j) - \bar{t})}{\sqrt{[\sum_{i<j}(x(i,j) - \bar{x})^2][\sum_{i<j}(t(i,j) - \bar{t})^2]}}.$$

x(i,j) is the Euclidean distance between i and j observations and t(i, j) is the height when i and j are first joined together.

Use scipy.cluster.hierarchy.cophenet

6. Truncation of the dendogram using “truncate_mode”
7. Decide the criteria: distance-cutoff or number of clusters
8. Assign clusters to each sample depending on the criteria