

# Clustering

## 1 AHC

Given the data  $X = \{1, 2, 9, 12, 20\}$  in 1-D space.

1. Apply the AHC clustering using Single Linkage / Complete Linkage for the  $X$  dataset.
2. Draw the clustering result using available functions (`dendrogram()` in Matlab or `hclust()` in R, etc.).
3. Apply on 2 more datasets from UCI. Make a study of data features. Observe the dendrogram and comment on results.
4. Conclude on the advantages and drawbacks of AHC.

## 2 K-means

Select two datasets from UCI and for each of them:

1. Run experiments with k-means. Explain the experimental protocol.
2. Explain the centroid initialization of the k-means used in the first question.
3. Analyze and compare the results with different value of  $k$ .
4. Calculate the clustering quality (any criteria in slides).
5. Use PCA or SVD to visualize the data distribution in 2D/3D.
6. Apply k-means for the selected datasets in 2D/3D after the use of PCA/SVD. Compare the performance of k-means before and after the dimensionality reduction.