

Week 3: Clustering

Clustering organizes things that are close into groups.

Hierarchical Clustering

Simple and intuitive technique to group and visualize multi-dimensional data.

Method

- Find two things which are close.
- Merge them together to form a new data point.
- Find the two such new datapoints which are close and merge them together.

Requires

- Distance Metric
- Merging technique

Output

- A tree showing how data points are close to each other.

Distance metric examples

- Euclidean distance (continuous)
- Correlation distance (continuous)
- Manhattan distance (binary)