Hierarchial Clustering:

- We build a clustering tree
- the last level, level N (no. of obs) each element is its own cluster

* Bottoms up approach
  - start off with independent observations
  - → we then merge "which 2 should be merged now"
  - until there is one cluster

Top down Approach
↓
Splitting each time, using best way to split each cluster up
- choose best cluster to split
- choose best way to split
- until each obs its own cluster

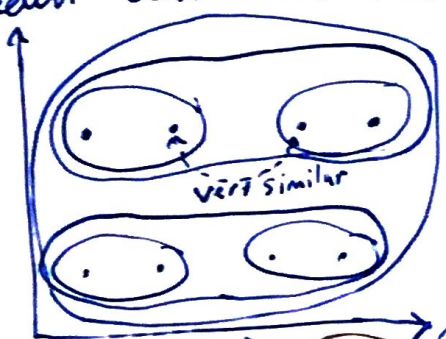
* Closest Pair of Clusters
  - when deciding which two clusters to merge, we usually need to determine which two clusters are the closest.

Single link
Similarity of most similar

Complete link
- Similarity of furthest pair

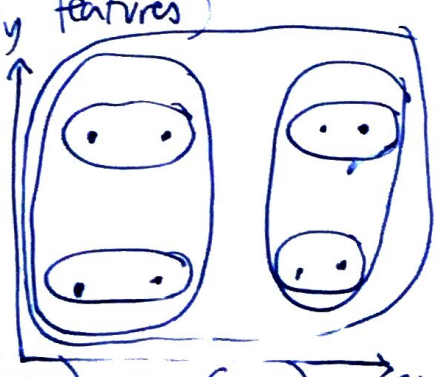
Avg. link ← most used.
- Avg. pairwise similarity of clusters

Each observations has many features



$Sim(C_i, C_j) = \max_{x \in C_i, y \in C_j} Sim(x, y)$

Single link (fat strong sausage)



$Sim(C_i, C_j) = \min_{x \in C_i, y \in C_j} Sim(x, y)$

Complete link

Compromise b/w Single & Complete link

Avg. over all pairs b/w the two clusters

$sim(C_i, C_j) = \frac{\sum_{x \in C_i} \sum_{y \in C_j} Sim(x, y)}{|C_i||C_j|}$