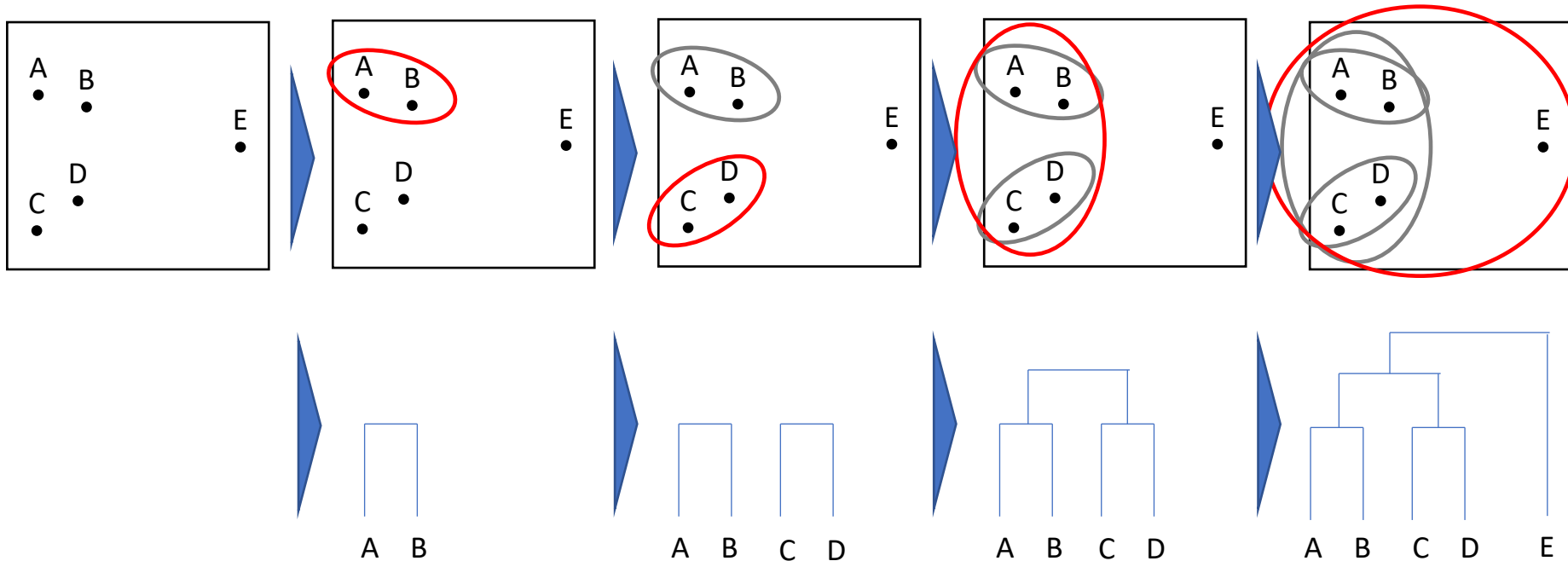


階層的クラスタリング

階層的クラスタリングとは？

- 距離の情報をもとに近いサンプルをグループ化して束ねていく。凝集型クラスタリング、ボトムアップ型クラスタリングなどとも呼ばれる
- 最終的に、樹形図（デンドログラム）ができあがる

■ 階層的クラスタリングの過程

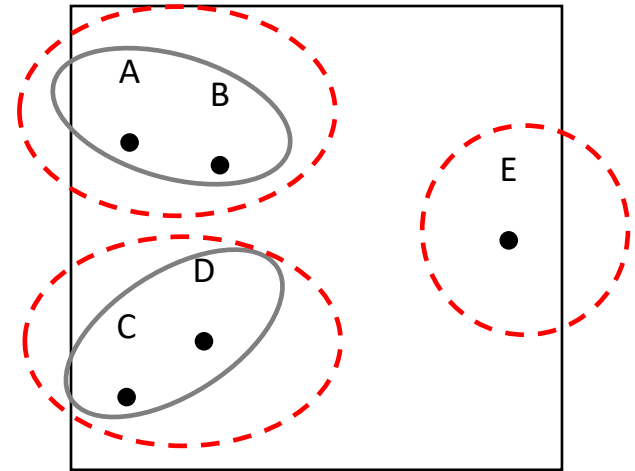
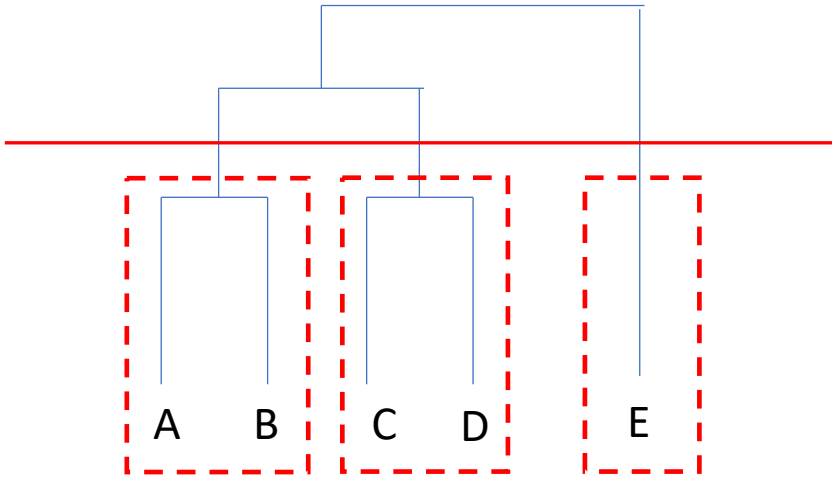


引用元)

• https://www.albert2005.co.jp/knowledge/data_mining/cluster/hierarchical_clustering

樹形図をもとにN個にグループ化することもできる

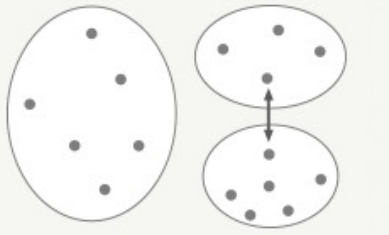
- 赤線の箇所では樹形図を区切ると、3つのグループに分割される



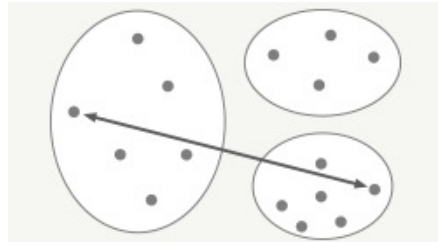
複数サンプルのクラスター間の距離の算出

- サンプルが1つずつの距離は、ユークリッド距離等の通常の計算でよいが、複数サンプルのクラスター間の距離の算出はどうか？

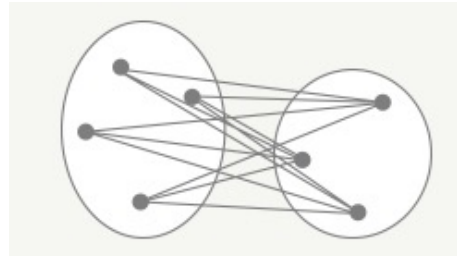
最短距離法



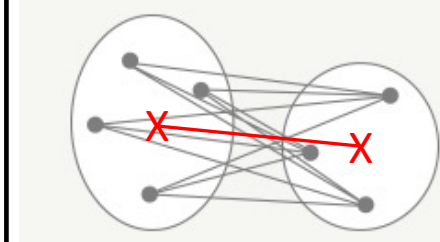
最長距離法



群平均法



重心法



ward (ウォード法)

「クラスターの重心と各サンプルとの距離の2乗和」を考慮する。クラスタリングの前後でのこの2乗和の変化が最小になるようにクラスタリングを実施する。

例) クラスタP、Q、Rの3つあったとする

PとQをクラスタリングしたとき

- 前：「クラスターPの2乗和」 + 「クラスターQの2乗和」
- 後：「クラスターWの2乗和」
- 差：「クラスターWの2乗和」 - 「クラスターPの2乗和」 - 「クラスターQの2乗和」

同様に、PとR、QとRについても前後の差を求めて、差が最小のものをクラスタリングする