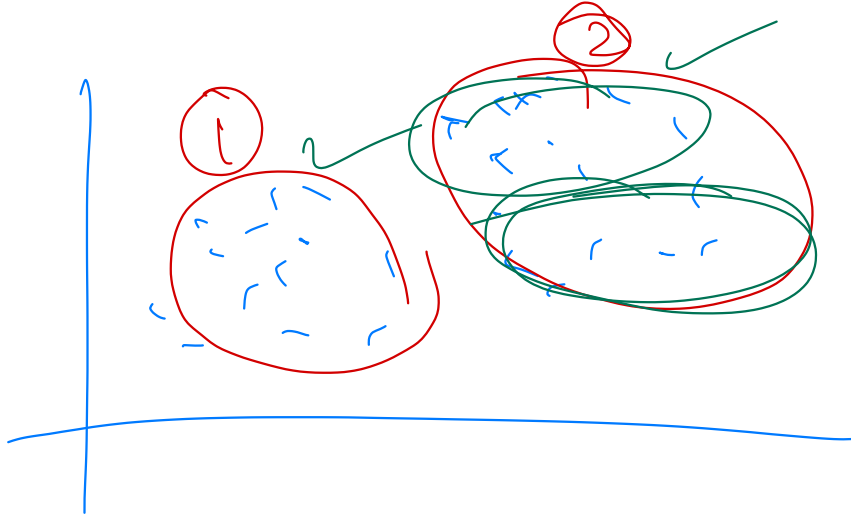


2021-12-20

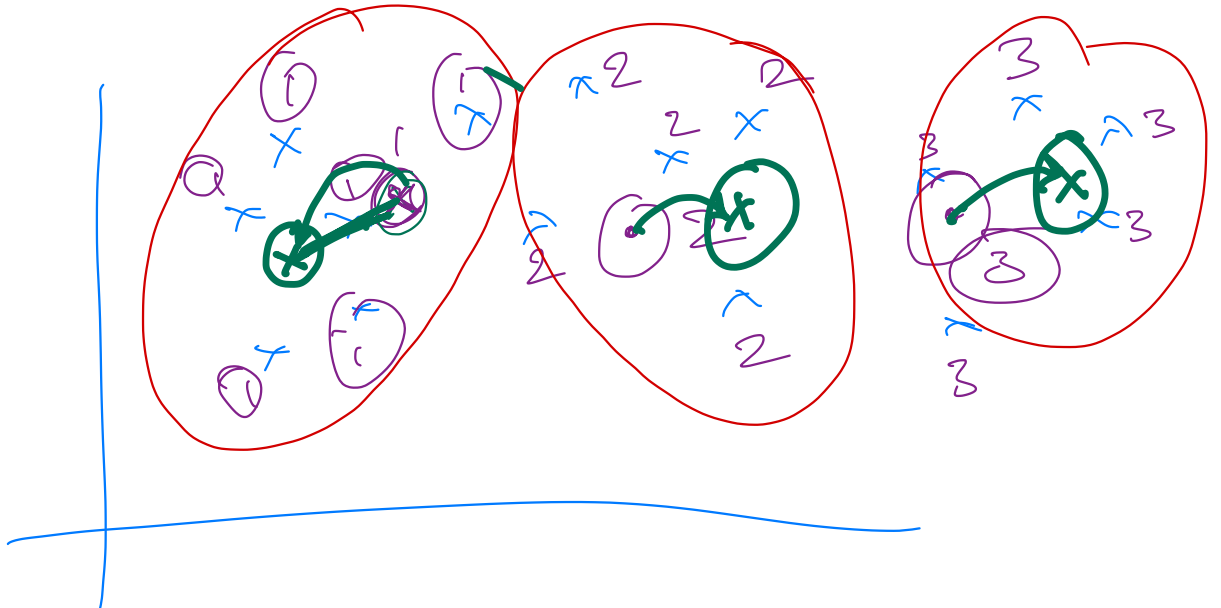
Cluster Analysis (Αναλυση κατά ομάδες)

Ομαδοποίηση παρατηρήσεων σε υποομάδες με "όμοιες" παρατηρήσεις σε κάθε υποομάδα



Μέτρηση ομοιότητας/απομοιότητας (κριτήρια)

Απόδοτος k-means (k ομάδες) (k: fixed)



Πλάτος Εγγύτητας (Proximity matrix)

$$D = (d_{ii'})_{i, i' = 1, \dots, N}$$

$N \times N$

$d_{ii'}$ = dissimilarity (ανομοιομορία) μεταξύ παρατηρήσεων i, i'

Μπορεί να αποκτήσει μη συμμετρικός

$$\tilde{D} = \frac{D + D^T}{2} \rightarrow \text{συμμετρικός}$$

$$d_{ii} \leq d_{ik} + d_{ki'}$$



Ανομοιομορία επί βάσει τιμών των μεταβλητών.

Δείγμα x_{ij} , $i = 1, \dots, N$, $j = 1, \dots, P$

N παραorp. P μεταβλητές

$$D(x_i, x_{i'}) = \sum_{j=1}^P w_j d_j(x_{ij}, x_{i'j})$$

Ειδικά $D(x_i, x_{i'}) = \rho(x_i, x_{i'}) = \frac{\sum_j (x_{ij} - \bar{x}_i)(x_{i'j} - \bar{x}_{i'})}{\sqrt{\sum_j (x_{ij} - \bar{x}_i)^2 \sum_j (x_{i'j} - \bar{x}_{i'})^2}}$

Πως ορίζεται η d_j ?

Ποσοτικές μεταβλητές

(x_j : ποσοτική)

$$d_j(x_i, x_{i'}) = d(|x_i - x_{i'}|)$$

$$x_i, x_{i'} \in \mathbb{R}$$

$$\text{αν } d(v) = v \Rightarrow d_j = |x_i - x_{i'}|$$
$$d(v) = v^2 \Rightarrow d_j = \underline{(x_i - x_{i'})^2}$$

Ordinal Variable

$$X_j \in \{1, 2, \dots, M\}$$

δ. άραξη

αλ κλίματα

Συνίδως

Rescaling

$$\text{Αν } X_j = i \Rightarrow X_j = \frac{i - 1/2}{M}, i = 1, \dots, M$$

$$i = 1, 2, 3, 4$$

$$\left(\frac{1}{8}\right) \quad \frac{3}{8} \quad \frac{5}{8} \quad \frac{7}{8}$$

κρίση
σαν quantitative
(ποσοτικές)

$$\sum \left(\frac{i - 1/2}{M}\right) = \frac{M}{2}$$

$$X_1 \in \{1, 2, 3, \cancel{4}\} \quad \frac{1}{8} \quad \frac{2}{8} \quad \frac{3}{8} \quad \frac{7}{8}$$

$$X_2 \in \{1, \dots, \cancel{20}\} \quad \frac{1}{40} \quad \frac{2}{40}, \frac{3}{40}, \dots, \frac{39}{40}$$
