Ταξινόμηση ασαφών c-μέσων

Έστω $X = \{x_1, x_2, ..., x_n\}$ ένα σύνολο δεδομένων

Ασαφής διαμέριση $P = \{A_1, A_2, ..., A_c\}$

με $\sum_{i=1}^{c} A_i(x_k) = 1 \qquad 0 < \sum_{k=1}^{n} A_i(x_k) < n$

$\Rightarrow$ A_i κλάσεις της διαμέρισης

Σύνολο δεδομένων $\mathbf{x}_k = (x_{k1}, x_{k2}, ..., x_{kp}) \in \mathbf{R}^p$

$\Rightarrow$ εύρεση της κατάλληλης ασαφούς διαμέρισης η οποία αναπαριστά τη δομή των δεδομένων

Κέντρα των κλάσεων $\mathbf{v}_i = \dfrac{\sum_{k=1}^{n} [A_i(\mathbf{x}_k)]^m \mathbf{x}_k}{\sum_{k=1}^{n} [A_i(\mathbf{x}_k)]^m}$

Αλγόριθμος Bezdek

Ελαχιστοποίηση δείκτη επίδοσης

$$J_m(P) = \sum_{k=1}^{n} \sum_{i=1}^{c} [A_i(\mathbf{x}_k)]^m \|\mathbf{x}_k - \mathbf{v}_i\|^2$$

ΒΗΜΑ 1: Θέτουμε $t = 0$. Επέλεξε μία αρχική τυχαία ασαφή διαμέριση $P^{(0)}$.

ΒΗΜΑ 2: Υπολόγισε τα c κέντρα των κλάσεων για την $P^{(t)}$ και τη δοσμένη τιμή του m.

ΒΗΜΑ 3: Ανανέωσε το $P^{(t+1)}$ από τη σχέση

$$A_i^{(t+1)}(\mathbf{x}_k) = \left[\sum_{j=1}^{c} \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i^{(t)}\|^2}{\|\mathbf{x}_k - \mathbf{v}_j^{(t)}\|^2} \right)^{\frac{1}{m-1}} \right]^{-1};$$

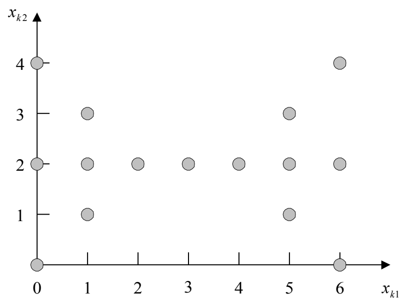
ΒΗΜΑ 4: Αν $|P^{(t+1)} - P^{(t)}| \le \varepsilon$, σταμάτησε; αλλιώς αύξησε το t και επανέλαβε

Ο Αλγόριθμος συγκλίνει για κάθε $m > 1$

Παράδειγμα

Σύνολο δεδομένων ($\varepsilon = 0.01$, $m = 1.25$)

k	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
x_{k_1}	0	0	0	1	1	1	2	3	4	5	5	5	6	6	6
x_{k_2}	0	2	4	1	2	3	2	2	2	1	2	3	0	2	4



Παράδειγμα

$\mathcal{P}^{(o)} = \{A_1, A_2\}$ με

$A_1 = .854 / x_1 + \cdots + .854 / x_{15}$

$A_1 = .146 / x_1 + \cdots + .146 / x_{15}$

ο αλγόριθμος τελειώνει όταν $t = 6$.

Η ασαφής ψευδοδιαμέριση που υπολογίζεται είναι η εξής:

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$A_1(x_k)$	.99	1	.99	1	1	1	.99	.47	.01	0	0	0	.01	0	.01
$A_2(x_k)$	.01	0	.01	0	0	0	.01	.33	.99	1	1	1	.99	1	.99

Τα κέντρα των κλάσεων είναι τα:

$v_1 = (0.88, 2) \quad v_2 = (5.14, 2)$

Ασαφείς σχέσεις ισοδυναμίας

Μία ασαφής δυαδική σχέση $R(X,X)$ θα λέγεται *ανακλαστική*, αν και μόνο αν

$$R(x,x)=1,\ \forall x \in X,$$

συμμετρική, αν και μόνο αν

$$R(x,y)=R(y,x),\ \forall x,y \in X$$

και *max-min μεταβατική* αν και μόνο αν

$$R(x,z)\geq R(x,y)\circ R(y,z),\ \forall x,y,z \in X.$$

Μία ασαφής δυαδική σχέση με τις τρεις ιδιότητες θα λέγεται σχέση *ισοδυναμίας*

$$R(x,y)=\begin{array}{c} a\\ b\\ c\\ d\\ e\\ f\\ g \end{array}\begin{bmatrix} 1 & .8 & 0 & .4 & 0 & 0 & 0\\ .8 & 1 & 0 & .4 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 1 & .9 & .5\\ .4 & .4 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 1 & .9 & .5\\ 0 & 0 & .9 & 0 & .9 & 1 & .5\\ 0 & 0 & .5 & 0 & .5 & .5 & 1 \end{bmatrix}$$

Μέθοδος ταξινόμησης με χρήση ασαφών σχέσεων

Ορισμός ασαφούς σχέσης από απόσταση δεδομένων

$$R(x_i,x_k)=1-\delta\left(\sum_{j=1}^{p}\left|x_{ij}-x_{kj}\right|^{q}\right)^{1/q}$$

Η R είναι σχέση συμβατότητας

Πρόταση

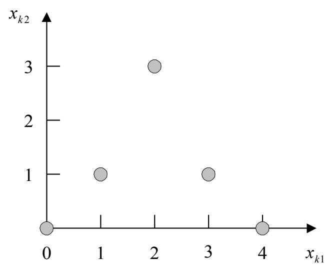
Έστω *R* μία σχέση συμβατότητας σε ένα σύνολο *X* με πληθικότητα $|x|=n$.

Τότε, το max-min μεταβατικό κλείσιμο της *R* είναι η σχέση $R^{(n-1)}$.

Παράδειγμα

Σύνολο δεδομένων

k	1	2	3	4	5
x_{k_1}	0	1	2	3	4
x_{k_2}	0	1	3	1	0



Παράδειγμα

Βρίσκουμε

$$R=\begin{bmatrix} 1 & .65 & .1 & .21 & 0\\ .65 & 1 & .44 & .5 & .21\\ .1 & .44 & 1 & .44 & .1\\ .21 & .5 & .44 & 1 & .65\\ 0 & .21 & .1 & .65 & 1 \end{bmatrix} \qquad R_T=\begin{bmatrix} 1 & .65 & .44 & .5 & .5\\ .65 & 1 & .44 & .5 & .5\\ .44 & .44 & 1 & .44 & .44\\ .5 & .5 & .44 & 1 & .65\\ .5 & .5 & .44 & .65 & 1 \end{bmatrix}$$

Ασαφείς διαμερίσεις

$a \in [0,.44]$: $\{\{x_1,x_2,x_3,x_4,x_5\}\}$

$a \in [.44,.5]$: $\{\{x_1,x_2,x_4,x_5\},\{x_3\}\}$

$a \in [.5,.65]$: $\{\{x_1,x_2\},\{x_3\},\{x_4,x_5\}\}$

$a \in [.65,.1]$: $\{\{x_1\},\{x_2\},\{x_3\},\{x_4\},\{x_5\}\}$

Αναγνώριση προτύπων

n κλάσεις προτύπων

αντιπροσωπεύονται από n παραδειγματικά πρότυπα

Είσοδος $\mathbf{u} = \left(u_1, u_2, \ldots, u_p\right)$

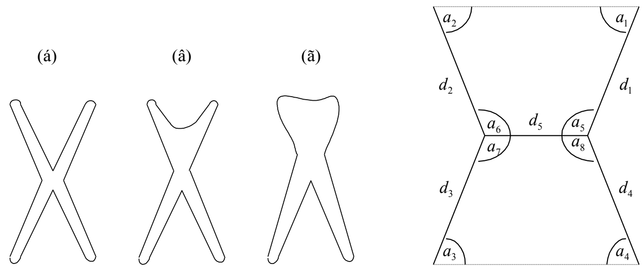
Σε ποια κλάση ταιριάζει περισσότερο

Συγκρίνουμε το u με τα n παραδειγματικά πρότυπα και παίρνουμε το διάνυσμα $A_k(\mathbf{u})$

Αναγνωρίζουμε το u ως μέλος της κλάσεις με τη μεγαλύτερη τιμή

Παράδειγμα

$\mathbf{u} = \left(a_1, a_2, \ldots, a_8, d_1, d_2, \ldots, d_5\right)$



Σχήμα 6: Εικόνες χρωμοσωμάτων και μοντέλο συμμετρικού χρωμοσώματος

(α) μεσοκεντρικά (β) υπομεσοκεντρικά (γ) ακροκεντρικά

Παράδειγμα

Ορίζουμε $S(\mathbf{u}) = 1 - \dfrac{1}{720^o} \sum_{i=1}^{4} \left|a_{2i-1} - a_{2i}\right|$

$$M(\mathbf{u}) = S(\mathbf{u}) \left[1 - \frac{\left|d_1 - d_4\right| + \left|d_2 - d_3\right|}{d_T}\right]$$

$$SM(\mathbf{u}) = S(\mathbf{u}) \left[1 - \frac{d_{SM}}{2d_T}\right]$$

$$AC(\mathbf{u}) = S(\mathbf{u}) \left[1 - \frac{d_{AC}}{4d_T}\right]$$

$$d_T = d_1 + d_2 + d_3 + d_4 + d_5$$

$$d_{SM} = \min\left(\left|d_1 - 2d_4\right| + \left|d_2 - 3d_3\right| + \left|2d_1 - d_4\right| + \left|2d_2 - d_3\right|\right)$$

$$d_{AC} = \min\left(\left|d_1 - 4d_4\right| + \left|d_2 - 4d_3\right| + \left|4d_1 - d_4\right| + \left|4d_2 - d_3\right|\right)$$

Αναγνώριση προτύπων

Έστω $\mathcal{P}_i,\ i \in N_p$ ασαφείς διαμερίσεις στο πεδίο ορισμού X_i των χαρακτηριστικών

Κάθε στοιχείο των συνόλων $\mathcal{P}_i$ έχει ένα γλωσσικό περιεχόμενο που περιγράφει το χαρακτηριστικό (π.χ. “ΜΙΚΡΗ ΓΩΝΙΑ”, “ΜΕΓΑΛΗ ΑΠΟΣΤΑΣΗ” κλπ)

Κατασκευάζονται *n* ασαφή συστήματα, ένα για κάθε κλάση, τα οποία παριστάνονται με τη βοήθεια *n* *p*-διάστατων ασαφών σχέσεων $R_1, \ldots, R_p\left(X_1 \times \cdots \times X_p \to [0,1]\right)$

Ο βαθμός συμβατότητας του κάθε προς ταξινόμηση προτύπου με το i παραδειγματικό πρότυπο που χαρακτηρίζει την κλάση i, δίνεται από τη σχέση

$$A_i(\mathbf{u}) = \mathbf{u} \circ R_i,\ i \in N_p$$

όπου $\circ$ η σύνθεση max-min

Το πρότυπο ταξινομείται, όπως και προηγούμενα, στην κλάση που παρουσιάζει το μεγαλύτερο βαθμό συμβατότητας, αν αυτός είναι αρκετά μεγάλος