2021-12-15

Εργασία 3 : Εφαρμογή natural splines

Ασκ 2

$$\boxed{\int_{x_{min}}^{x_{max}} N_i''(t)\, N_j''(t)\, dt = (\Omega_N)_{ij}}$$

$$N_{i+2}(t) = d_i(t) - d_{K-1}$$

$$d_k(t) = \frac{(t-\xi_k)_+^3 - (t-\xi_K)_+^3}{\xi_K - \xi_k}$$

$\xi_k = x_k$.

$N \to \Omega_N \to K$

Ω_N 150×150

1. Symbolic Computation {π.χ. SageMath}
2. N ← προσοχή στην αντιστροφή
3. $\int N_j'' N_k''$ ← αριθμητική ολοκλήρωση

δοκιμάστε με μικρό $N \approx 10$

Association Rules

$X = (X_1, \dots, X_p)$

S_j = support της X_j = πεδίο τιμών της X_j

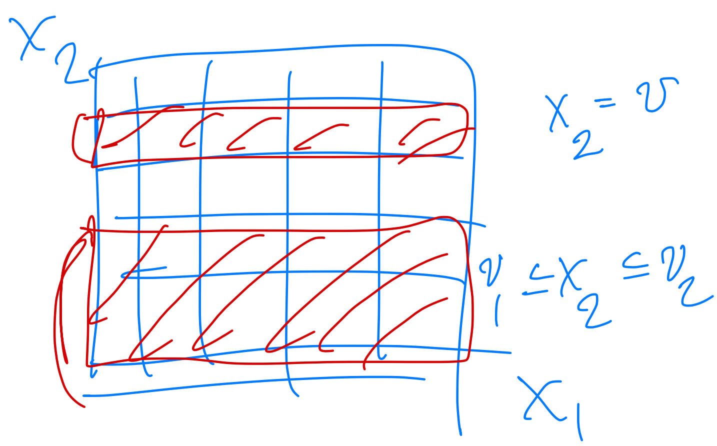
Θέλουμε $s_j \subset S_j$, $j = 1, \dots, p$.

$$\prod_{j=1}^{p} s_j \subseteq S_1 \times S_2 \times \dots S_p.$$

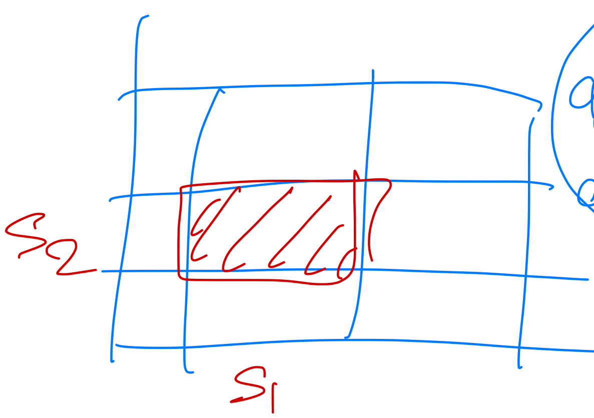
τ.ω. $P\left(\bigcap_{j=1}^{p} (X_j \in s_j)\right)$: "μεγάλη"

if $s_j = S_j$

X_j δεν εμφανίζεται



$S_1 \times S_2 \times \dots S_p \leftarrow$ association rule | κανόνας συσχέτισης



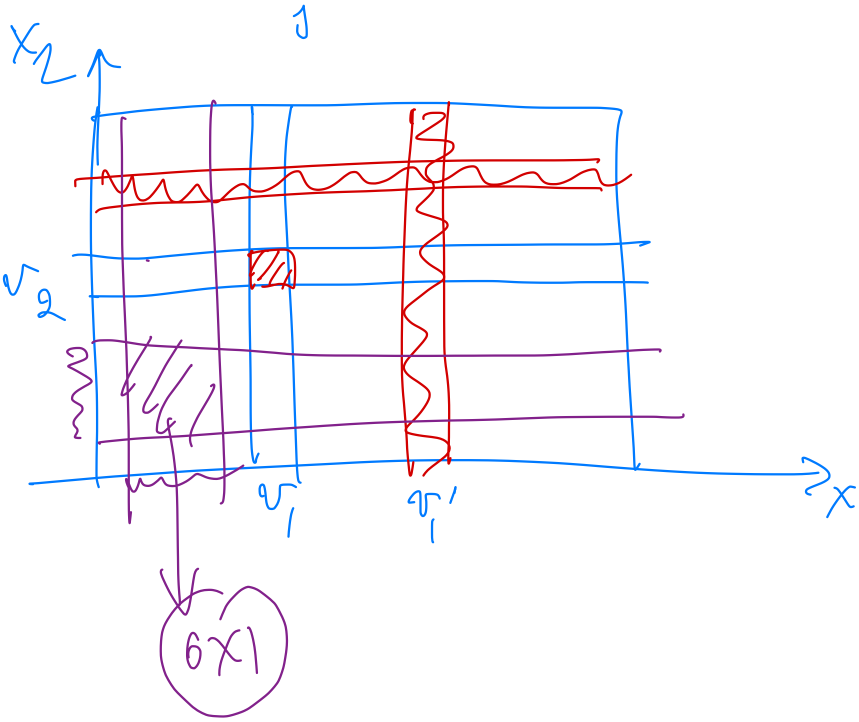
$a_1 \le x_1 \le b_1$ και $a_2 \le x_2 \le b_2$

Market Basket Approach.

Γενικά!

$S_j \to \{v_{0j}\}$, $S_j \to S_j$

$X_j = v_{0j}$



Ισοδύναμα οι κανόνες που επιτρέπονται είναι:

$$J \subset \{1, \dots, p\} \quad (v_{0j},\ j \in J) \implies \bigcap_{j \in J} \{X_j = v_{0j}\}$$

Αν X_j κατηγορικές $\quad X_j \in \{x_{j1}, x_{j2}, \dots, x_{j,m_j}\} = S_j$

Ορίζουμε $\quad Z_{jk} = 1(X_j = x_{jk}) \quad j = 1, \dots, P \quad k = 1, \dots, m_j$

$X \to Z \in \{0,1\}^K, \quad K = \sum_{j=1}^{P} |S_j| \implies$

Κανόνας $\iff \mathcal{K} \subset \{1, 2, \ldots K\}$

Μεταβλητές $Z_1, \ldots, Z_K$

$$P\left(\bigcap_{k \in \mathcal{K}} (Z_k = 1)\right) = P\left(\prod_{k \in \mathcal{K}} Z_k = 1\right) \quad \text{"large"}$$

Συνολικός αρ. κανόνων $= 2^K$

$\mathcal{K} \subset K$

$\uparrow$ item set

Για ένα item set $\mathcal{K}$, $i = 1, \ldots, N$ $\Big\}$ Z_{ik} = τιμή της Z_k στην παρατήρηση i

$= 1$ (αν ο πελάτης i αγόρασε το k)

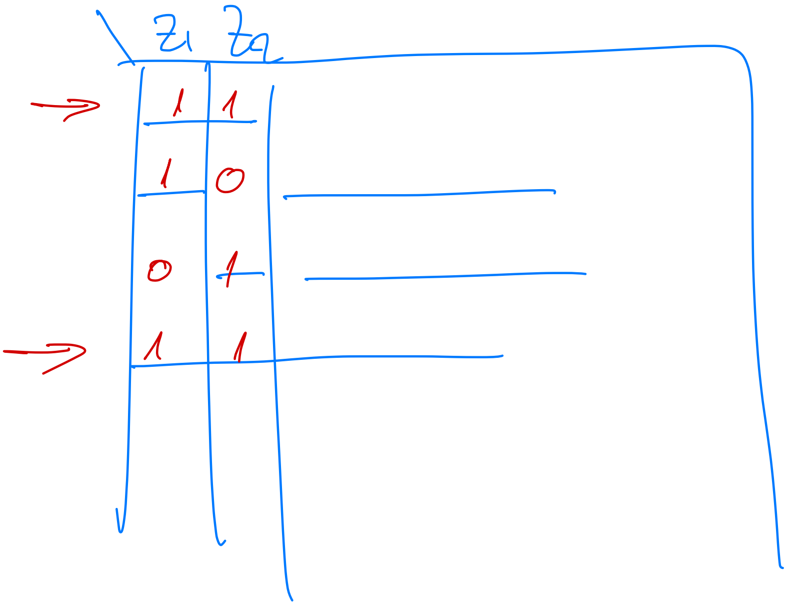
Dataset:

i	Z_1	.	.	.	.	.	Z_K
1	0	0	1	0	0	1	1
2	1	0	1	0	.	.	.
⋮							
N	1	1	0	. 1	0	0	

$$\hat{P}\left(\prod_{k\in\mathcal{K}} z_k = 1\right) = \frac{1}{N}\sum_{i=1}^{N} \underline{\prod_{k\in\mathcal{K}}} z_{ik} = T(\mathcal{K})$$

support / prevalence

στήριγμα / "επιπολασμός"



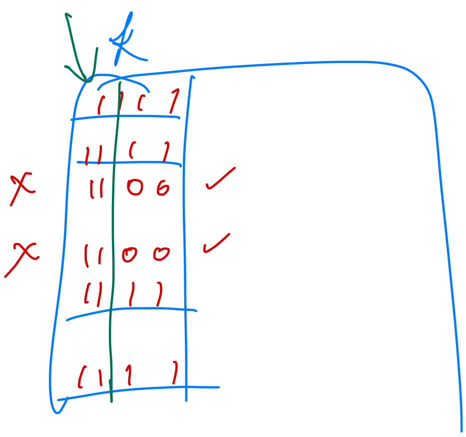
$\mathcal{K} = \{1, 2\}$

given threshold

Θέλουμε να εντοπίσουμε item sets $\mathcal{K}$: $T(\mathcal{K}) > t$

Apriori Algorithm για εντοπισμό των : $\mathcal{K} : T(\mathcal{K}) > t$

αν $T(\mathcal{K}) > t$, $L \subset \mathcal{K} \Rightarrow T(L) \geq T(\mathcal{K})$



Επαναληπτικός Αλγόριθμος

1. Ένα πέρασμα, υπολογίζει support για όλα τα μονοσύνολα.

   Επιλέγω μόνο αυτά που έχουν support $> t$

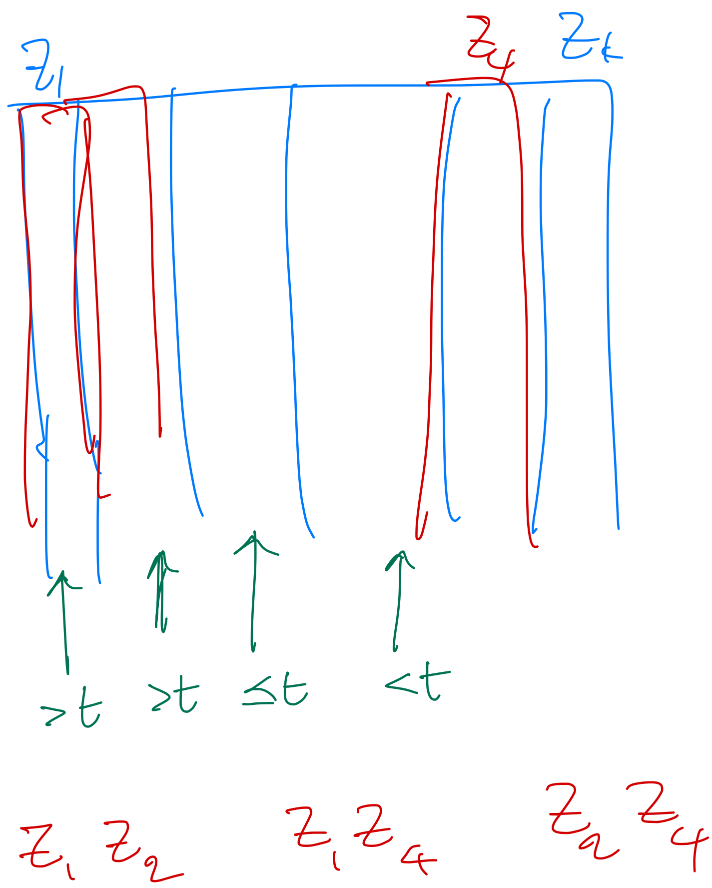
2. Πέρασμα 2: Από τα μονοσύνολα που έμειναν δημιουργούμε όλα τα δυνατά δισύνολα

   υπολ. το support για καθένα από αυτά.

   Επιλέγω μόνο αυτά που έχουν support $> t$.

⋮

K. Πέρασμα K



Association Rule:

Έστω $\mathcal{k} \subset K \quad : \quad T(\mathcal{k}) > t$

$\mathcal{k} = A \cup B \quad , \quad A \cap B = \emptyset$

Association rule : $\underline{\underline{A \Rightarrow B}}$

(αν αγοράσουν τα προϊόντα στο A τότε αγοράζουν και αυτά στο B)

$\downarrow$ antecendent — consequent

$$T(A \Rightarrow B) = T(A \cup B) = T(\mathcal{k})$$

$= \%$ παρατηρ. που $Z_k = 1 \ , k \in A \cup B$

Confidence/Predictability του $A \Rightarrow B$

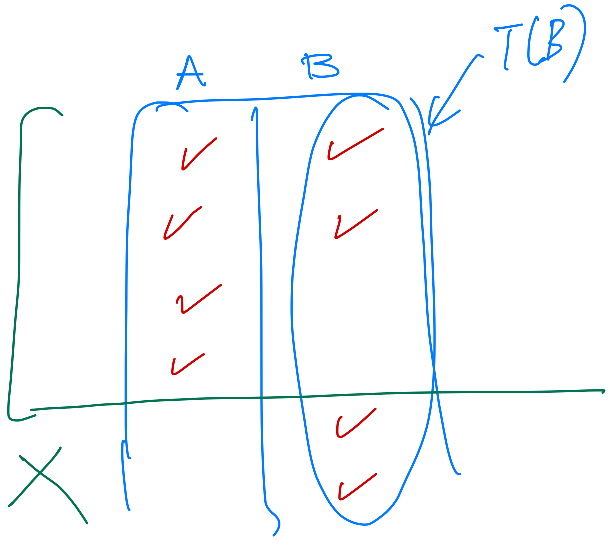
$$C(A \Rightarrow B) = \frac{T(A \Rightarrow B)}{T(A)} = \frac{\hat{P}(Z_k = 1, k \in A \cup B)}{\hat{P}(Z_k = 1, k \in A)}$$

$$= \hat{P}(\text{αγορ} B \,/\, \text{αγοράζει } A)$$

$T(B)$ = expected confidence $(= \hat{P}(B))$

$$\frac{C(A \Rightarrow B)}{T(B)} = \frac{P(B|A)}{P(B)}$$

lift του κανόνα



Market Basket Analysis as Supervised Method.

Έστω $X \sim g(x)$ άγνωστη σππ. $\to \exists$ δείγμα μεγέθους N

Έστω $g_0(x)$ μια γνωστή σππ

Δημιουργώ N_0 παρατηρήσεις τυχαία από g_0

Νέο σύνθετο δείγμα

$N+N_0$

X	Y	
⋮	1	$\leftarrow X \sim g$
	0	$\leftarrow X \sim g_0$
	1	
	1	
⋮	0	
	0	
	⋮	

Μπορεί να θεωρηθεί ως δείγμα από μια μείξη κατανομών $g(x)$, $g_0(x)$

$$\mu(x) = E(Y|X=x) = P(Y=0|X=x)\cdot 0 + P(Y=1|X=x)\cdot 1$$

$$= \frac{g(x)}{g(x)+g_0(x)} = \frac{g(x)/g_0(x)}{1+g(x)/g_0}$$

Μπορούμε να εκτιμήσουμε το $\mu(x)$ μέσω logistic regression $\hat{\mu}(x)$ $\Rightarrow$ $\hat{g}(x) = \frac{\hat{\mu}(x)}{1-\hat{\mu}(x)} \cdot g_0(x)$

$$\log \frac{g(x)}{g_0(x)}$$

$$\frac{C(A \Rightarrow B)}{T(B)} = \frac{T(A \Rightarrow B)}{T(A)\,T(B)} = \frac{C(B \Rightarrow A)}{T(A)}$$

$$= \frac{T(A \cup B)}{T(A)\,T(B)}$$

$A = \{$[illegible] [illegible]$\}$

$B = \{$[illegible], [illegible]$\}$

$P(A|B)$

$P(B|A)$