

Exercise 3

Data Mining Algorithms 1 - WS 2015/16

Davide Pedranz
(362504)

Jakub Amanowicz
(362594)

Hongmei Liang
(292520)

November 13, 2015

Exercise 3.1

(a) Let $S \subseteq I$ be a frequent itemset. Then every non-empty subset $S' \subseteq S$ must also be frequent.

Let $D = \{T \mid T \subseteq I\}$ be the database of transactions and $T \in D$ a transaction in it. Let $S \subseteq T$ be an itemset of T and $S' \subseteq S$ a subset of S . Then we have

$$\{T \in D \mid S' \subseteq T\} \stackrel{S' \subseteq S \subseteq T}{\supseteq} \{T \in D \mid S \subseteq T\} \quad (1)$$

It follows:

$$\text{support}(S') \stackrel{\text{def}}{=} |\{T \in D \mid S' \subseteq T\}| \stackrel{(1)}{\geq} |\{T \in D \mid S \subseteq T\}| \stackrel{\text{def}}{=} \text{support}(S)$$

Thus:

$$\text{support}(S') \geq \text{support}(S) \quad (2)$$

Since S is frequent:

$$\text{support}(S') \stackrel{(2)}{\geq} \text{support}(S) \stackrel{S \text{ frequent}}{\geq} \text{minFreq}$$

We conclude that S' is frequent. \square

(b) Let $S \subseteq I$ be an arbitrary itemset. Then $\text{support}(S') \geq \text{support}(S)$ holds for any non-empty subset $S' \subseteq S$.

See (2) in the previous proof.