

Knowledge Discovery and Data Mining I
 WS 2019/20

Exercise 11: Association Rules, Prefix Span, Interestingness

Exercise 11-1 Association Rules

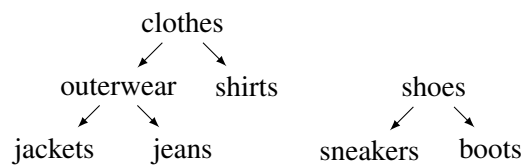
Given the following frequent itemsets extract all strong association rules with a minimum confidence of $minConf = 80\%$. Which candidates can be pruned based on anti-monotonicity?

Itemset	Support
A	1.00
B	1.00
D	0.75
AB	1.00
AD	0.75
BD	0.75
ABD	0.75

Exercise 11-2 R-Interestingness

Given the following item hierarchy and frequent itemsets decide whether the these association rules are R-interesting using $R = 1.6$ and explain why.

Itemset	Support
{clothes}	20
{outerwear}	10
{jackets}	4
{shoes}	15
{clothes, shoes}	10
{outerwear, shoes}	9
{jackets, shoes}	4



- (a) clothes \Rightarrow shoes
- (b) outerwear \Rightarrow shoes
- (c) jackets \Rightarrow shoes

Exercise 11-3 Sequential Pattern Mining

Let D be a database that contains the following five sequences.

SID	Sequence
1	ABBA
2	BBACA
3	CBAA
4	ACA
5	BAAB

In addition let $min_sup = 40\%$, i.e. there need to be 2 sequences supporting a pattern.

- Find all frequent sequence patterns using the *PrefixSpan* algorithm.
- Which patterns are maximal? Which are closed?