

Online-Test 4: Association Rules

Test und Assessment – Druckansicht

Online-Test 4: Association Rules

Datum: Tue Nov 2 07:57:18 2021 Maximale Punktezahl: 10

Frage 1 - Frequent, Maximal und Closed (1 Punkt) [ID: 931928]

Gegeben seien folgende Transaktionen:

ID Items

- 1 A, B
- 2 A
- 3 A, B, C

Bewerten Sie folgende Aussagen unter Annahme eines minimalen Supports von 2:

Für jede Aussage muss entschieden werden: [richtig] oder [falsch]

richtig

- ☐
- ☒
- ☐
- ☒

falsch

- ☒
- ☐
- ☒
- ☐

A ist ein Itemset.

{A} ist frequent.

{A} ist maximal.

{A} ist closed.

A ist natürlich kein Itemset, sondern ein Item.

{A} übertrifft den geforderten minimalen Support und ist damit frequent.

Es gibt keine Obermenge mit demselben Support, insofern ist {A} auch closed.

{A} ist allerdings nicht maximal, da es mit {A, B} eine Obermenge gibt, die auch frequent ist.

Frage 2 - Association Rules - Support (1 Punkt) [ID: 932116]

Gegeben sei eine Menge an Transaktionen:

ID Items

- 1 fruits, milk, vegetables
- 2 soda, vegetables, milk, fruits
- 3 vegetables, fruits, soda, beer, milk
- 4 fruits, soda, milk
- 5 fruits, beer, milk
- 6 fruits, vegetables, soda, milk, beer
- 7 soda, milk, beer, vegetables, fruits
- 8 soda, fruits

Welche der folgenden Association Rules haben einen Mindestsupport von 50%?

Für jede Aussage muss entschieden werden: [richtig] oder [falsch]

richtig

- ☒

falsch

- ☐

{fruits} => {milk}

<input type="radio"/>	<input checked="" type="radio"/>	{soda} => {beer}
<input checked="" type="radio"/>	<input type="radio"/>	{vegetables} => {milk}
<input checked="" type="radio"/>	<input type="radio"/>	{milk, soda} => {vegetables}

Der Support einer Association Rule der Form $LHS \Rightarrow RHS$ ist definiert als die Häufigkeit der (Item-)Mengenvereinigung $LHS \cup RHS$.

Dementsprechend ergeben sich:

$\text{Support}(\{ \text{fruits} \} \Rightarrow \{ \text{milk} \}) = \text{Support}(\{ \text{fruits}, \text{milk} \}) = 7/8$

$\text{Support}(\{ \text{soda} \} \Rightarrow \{ \text{beer} \}) = \text{Support}(\{ \text{soda}, \text{beer} \}) = 3/8$

$\text{Support}(\{ \text{vegetables} \} \Rightarrow \{ \text{milk} \}) = \text{Support}(\{ \text{vegetables}, \text{milk} \}) = 5/8$

$\text{Support}(\{ \text{milk}, \text{soda} \} \Rightarrow \{ \text{vegetables} \}) = \text{Support}(\{ \text{milk}, \text{soda}, \text{vegetables} \}) = 4/8$

Frage 3 - Association Rules - Confidence (1 Punkt) [ID: 932118]

Gegeben sei eine Menge an Transaktionen:

ID Items

- 1 fruits, milk, vegetables
- 2 soda, vegetables, milk, fruits
- 3 vegetables, fruits, soda, beer, milk
- 4 fruits, soda, milk
- 5 fruits, beer, milk
- 6 fruits, vegetables, soda, milk, beer
- 7 soda, milk, beer, vegetables, fruits
- 8 soda, fruits

Welche der folgenden Association Rules haben eine Confidence von mindestens 90%?

Für jede Aussage muss entschieden werden: [richtig] oder [falsch]

richtig	falsch	
<input type="radio"/>	<input checked="" type="radio"/>	{fruits} => {milk}
<input type="radio"/>	<input checked="" type="radio"/>	{soda} => {beer}
<input checked="" type="radio"/>	<input type="radio"/>	{vegetables} => {milk}
<input type="radio"/>	<input checked="" type="radio"/>	{milk, soda} => {vegetables}

Die Confidence einer Association Rule der Form $LHS \Rightarrow RHS$ ist definiert als $\frac{\text{Support}(LHS \cup RHS)}{\text{Support}(LHS)}$.

Dementsprechend ergeben sich:

$\text{Confidence}(\{ \text{fruits} \} \Rightarrow \{ \text{milk} \}) = 7/8$

$\text{Confidence}(\{ \text{soda} \} \Rightarrow \{ \text{beer} \}) = 3/6$

$\text{Confidence}(\{ \text{vegetables} \} \Rightarrow \{ \text{milk} \}) = 5/5$

$\text{Confidence}(\{ \text{milk}, \text{soda} \} \Rightarrow \{ \text{vegetables} \}) = 4/5$

Frage 4 - A-Priori - Schritte (1 Punkt) [ID: 931360]

Bringen Sie die Schritte des A-Priori-Algorithmus in die richtige Reihenfolge!

- Erzeugen der einelementigen Frequent Itemsets

- [BEGINN] Schleife über k
- Join
- Prune
- Support Counting
- [ENDE] Schleife über k
- Bestimmen der Association Rules

Siehe Vorlesung.

Frage 5 - A-Priori - Pruning (1 Punkt) [ID: 932120]

Gegeben seien folgende Frequent Itemsets der Größe 2:

{A, B}

{A, C}

{A, D}

{B, D}

{C, D}

Welche der folgenden Itemsets der Größe 3 können im A-Priori-Algorithmus im Pruning-Schritt entfernt werden?

Für jede Aussage muss entschieden werden: [ja] oder [nein]

ja

nein



{A, B, C}



{A, B, D}



{A, C, D}



{B, C, D}

Es können alle die Itemsets der Größe 3 entfernt werden, die ein Itemset der Größe 2 enthalten, welches nicht frequent ist. Für jedes Itemset der Größe 3 müssen 3 solche Teilmengen überprüft werden. Das Itemset {B, C} ist nicht frequent,

Frage 6 - FP-Baum - Reihenfolge der Items (1 Punkt) [ID: 931369]

Gegeben sei eine Menge an Transaktionen. Ordnen Sie die Items entsprechend ihrer Reihenfolge im FP-Baum an (angenommen, die Items befinden sich auf einem Pfad)!

ID Items

- 1 fruits, milk, vegetables
- 2 soda, vegetables, milk, fruits
- 3 vegetables, fruits, soda, beer, milk
- 4 fruits, soda, milk
- 5 fruits, beer, milk
- 6 fruits, vegetables, soda, milk, beer
- 7 soda, milk, beer, vegetables, fruits
- 8 soda, fruits

- fruits
- milk
- soda
- vegetables
- beer

Die Items werden entsprechend ihrer Häufigkeit im Gesamtdatensatz geordnet. Häufigere Items tauchen im Baum weiter oben auf.

Frage 7 - FP-Baum - Aufbau (1 Punkt) [ID: 931378]

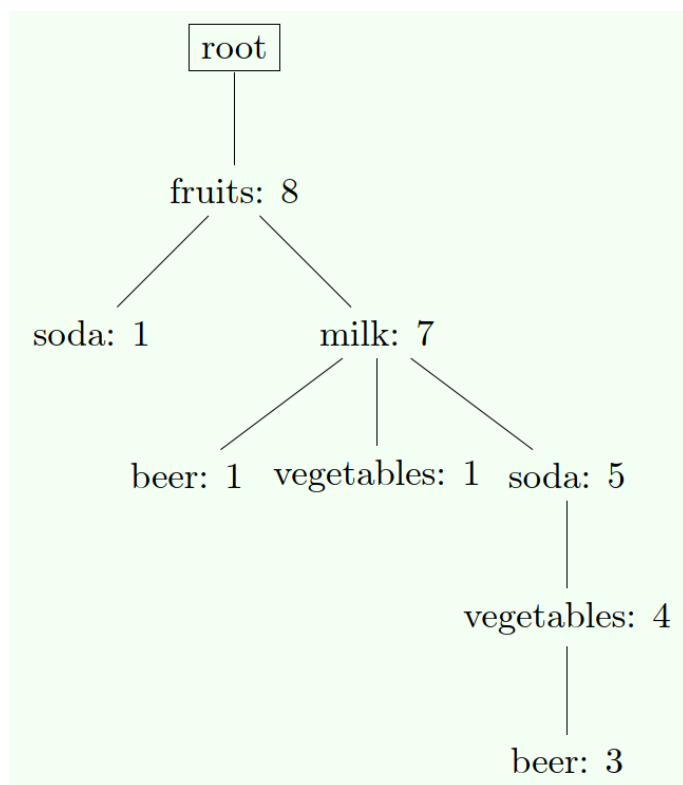
Gegeben sei eine Menge an Transaktionen. Wie viele Kanten enthält der zugehörige FP-Baum (die erste Kante startet vom Knoten "root", die restlichen Knoten entsprechen den Items)?

ID Items

- 1 fruits, milk, vegetables
- 2 soda, vegetables, milk, fruits
- 3 vegetables, fruits, soda, beer, milk
- 4 fruits, soda, milk
- 5 fruits, beer, milk
- 6 fruits, vegetables, soda, milk, beer
- 7 soda, milk, beer, vegetables, fruits
- 8 soda, fruits

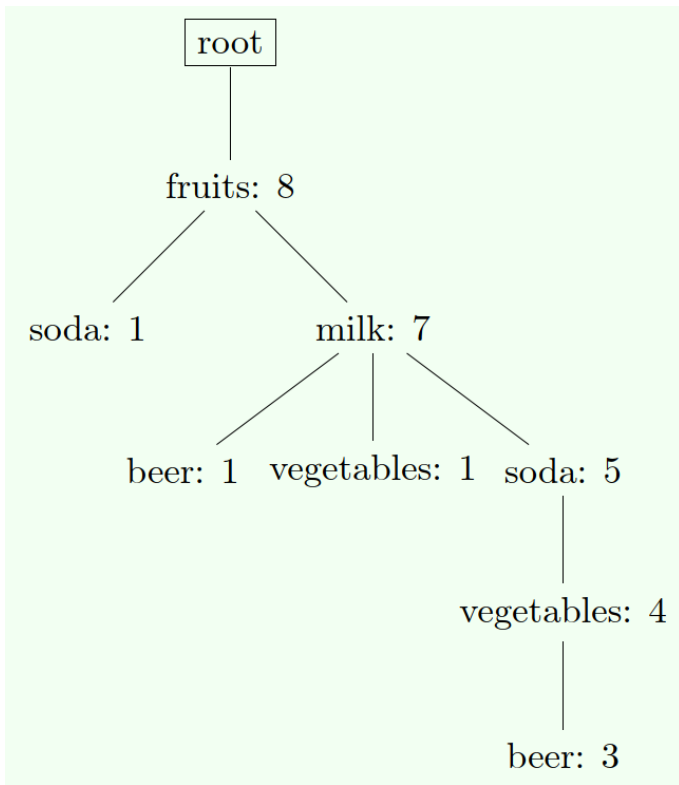
Der Wert muss zwischen 8 und 8 liegen

Ein Beispielbaum könnte so aussehen (die horizontale Reihenfolge von Kindknoten wurde hier willkürlich gewählt):



Frage 8 - FP-Baum - Extraktion der Regeln (1 Punkt) [ID: 931448]

Gegeben sei folgender FP-Baum:



Welche Items sind zusammen mit *vegetables* in mindestens einem Frequent Itemset, wenn der minimale Support 5 beträgt?

Für jede Aussage muss entschieden werden: [ja] oder [nein]

ja	nein	
<input type="radio"/>	<input checked="" type="radio"/>	beer
<input checked="" type="radio"/>	<input type="radio"/>	fruits
<input checked="" type="radio"/>	<input type="radio"/>	milk
<input checked="" type="radio"/>	<input checked="" type="radio"/>	soda

Es gibt zwei Pfade, auf denen *vegetables* vorkommt. Der Pfad über *fruits* und *milk* hat einen Support von 1. Der Pfad über *fruits*, *milk* und *soda* hat einen Support von 4 bis zu *vegetables* bzw. 3 bis zu *beer*. Insgesamt kommt *vegetables* daher fünfmal zusammen mit *fruits*, fünfmal zusammen mit *milk*, viermal zusammen mit *soda* und dreimal zusammen mit *beer* vor.

Frage 9 - Anti-Monotonizität (1 Punkt) [ID: 932122]

Gegeben sei eine Menge von Attributwerten A. Welche der folgenden Constraints ist anti-monoton?

Für jede Aussage muss entschieden werden: [ja] oder [nein]

ja	nein	
<input type="radio"/>	<input checked="" type="radio"/>	$\text{Median}(A) > 5$
<input type="radio"/>	<input checked="" type="radio"/>	$\text{Modalwert}(A) < 10$
<input checked="" type="radio"/>	<input type="radio"/>	$\text{Max}(A) - \text{Min}(A) < 18$
<input checked="" type="radio"/>	<input checked="" type="radio"/>	$\text{Varianz}(A) > 4$

Median, Modalwert und Varianz können sich durch Hinzunahme oder Weglassen von Werten sowohl nach unten als auch nach oben verschieben, sodass die Constraints nicht anti-monoton sind (im Übrigen auch nicht monoton). Die Spannbreite (Maximum - Minimum) kann durch Hinzunahme weiterer Werte nur größer werden, sodass eine Ungleichung mit $<$ für größere Itemsets verletzt bleibt, wenn sie für Teilmengen schon verletzt war. (Oder umgekehrt: Wenn die Constraint für ein Itemset erfüllt ist, bleibt sie auch für jede Teilmenge davon erfüllt). Dadurch ist die Constraint anti-monoton.

Frage 10 - Succinctness (1 Punkt) [ID: 932124]

Gegeben sei eine Menge von Attributwerten A. Welche der folgenden Constraints sind succinct?

Für jede Aussage muss entschieden werden: [richtig] oder [falsch]

richtig	falsch	
<input checked="" type="radio"/>	<input type="radio"/>	$\min(A) < 10$
<input checked="" type="radio"/>	<input type="radio"/>	$\min(A) > 10$
<input type="radio"/>	<input checked="" type="radio"/>	$\text{sum}(A) < 10$
<input type="radio"/>	<input checked="" type="radio"/>	$\text{sum}(A) > 10$

Zum Vergleichen mit dem Minimum-Operator kann man zunächst für jeden einzelnen Wert (jedes einzelne Item) feststellen, ob die Constraint erfüllt ist. Um zu einer Gesamtaussage für eine Menge von Werten/Items (wie sie im Laufe des Mining-Prozesses entsteht) zu kommen, muss man diese Einzelaussagen nur noch aggregieren, ohne dass man sich die ursprünglichen Werte in der Datenbank anschauen muss:

- $\min(A) < x$: Sobald ein Element kleiner als x ist (also die Constraint für eine einelementige Teilmenge erfüllt ist), ist die Constraint insgesamt erfüllt.
- $\min(A) > x$: Nur wenn alle Elemente größer als x sind (also die Constraint für alle einelementigen Teilmengen erfüllt ist), ist die Constraint insgesamt erfüllt.

Für den $\text{sum}()$ -Operator müssen dagegen die tatsächlichen Werte aus der Datenbank gelesen werden. Insofern ist der Operator nicht succinct.