



Lecture 8: Association Rule Mining

Pilsung Kang

School of Industrial Management Engineering

Korea University

Type of Machine Learning/Data Mining

- According to the existence of target (Y) variable
 - ✓ Supervised learning vs. Unsupervised learning

Supervised Learning

A given dataset **X** & **Y**

	Var. 1	Var. 2	...	Var. d		Y
Ins. 1	$y = f(x)$..
Ins. 2
...
Ins. N

Semi-supervised Learning

A given dataset **X** & **Y**

	Var. 1	Var. 2	...	Var. d		Y
Ins. 1	$y = f(x)$..
Ins. 2
...
Ins. N
...		
...		
...		
...		
...		
Ins. M		

Unsupervised Learning

A given dataset **X**

	Var. 1	Var. 2	...	Var. d
Ins. 1
Ins. 2
...
Ins. N

Type of Machine Learning/Data Mining

- Unsupervised Learning

$$\mathcal{X} = \{\mathbf{x}_i | i = 1, \dots, n, \mathbf{x}_i \in \mathbb{R}^d\}$$



Unsupervised learning

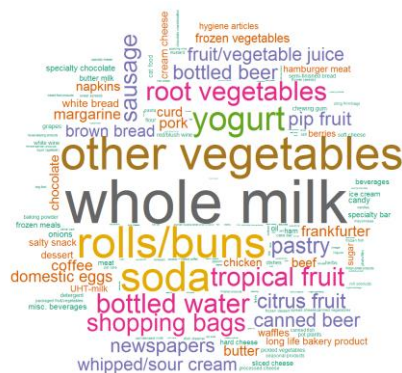
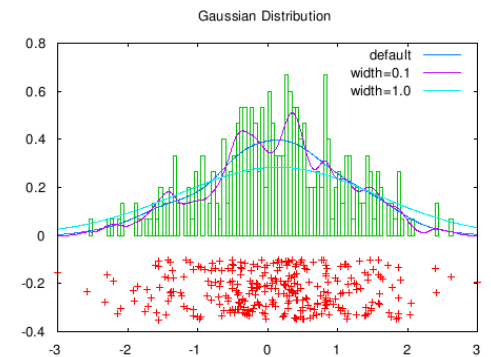
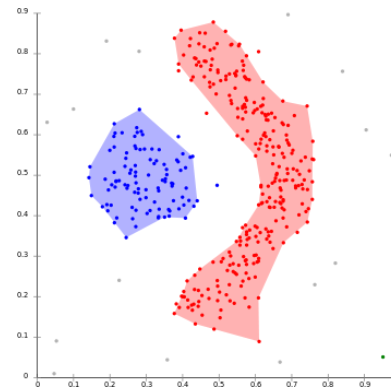
- Explores intrinsic characteristics.
- Estimates underlying distribution.
- Density estimation, clustering, novelty detection, etc.

Type of Machine Learning/Data Mining

- Unsupervised Learning

A given dataset X

	Var. 1	Var. 2	...	Var. d
Ins. 1
Ins. 2
...
Ins. N



이 책과 함께 구매한 도서



데이터마이닝 방법론(빅데이터 분석을 위한)

25,000원

[0%*1%P]



R로 배우는 데이터 분석 기본기 데이터

28,000원

[10%*10%P]



데이터마이닝 입문(SAS Enterprise)

20,000원

[0%*1%P]



Type of Machine Learning/Data Mining

- Supervised Learning

$$\mathcal{X} = \{\mathbf{x}_i | i = 1, \dots, n, \mathbf{x}_i \in \mathbb{R}^d\}$$

Supervised learning

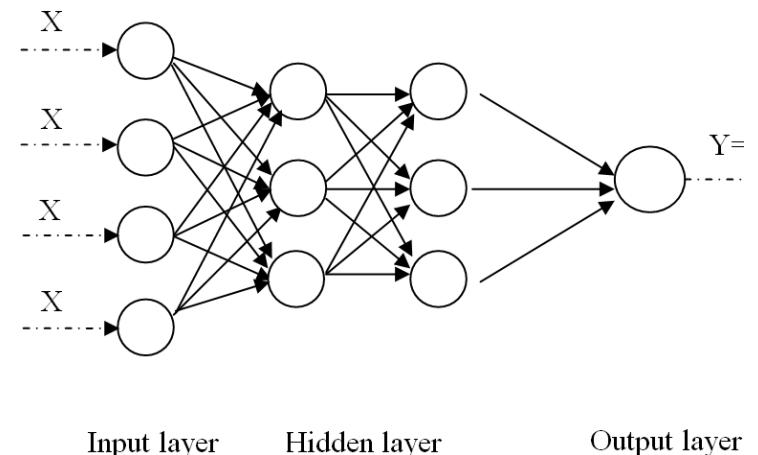
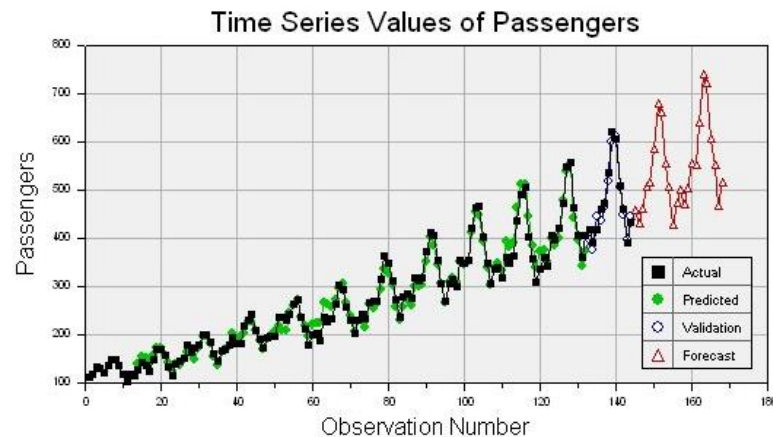
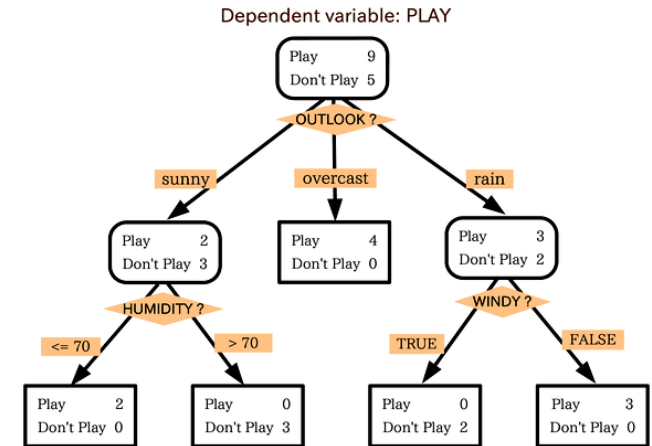
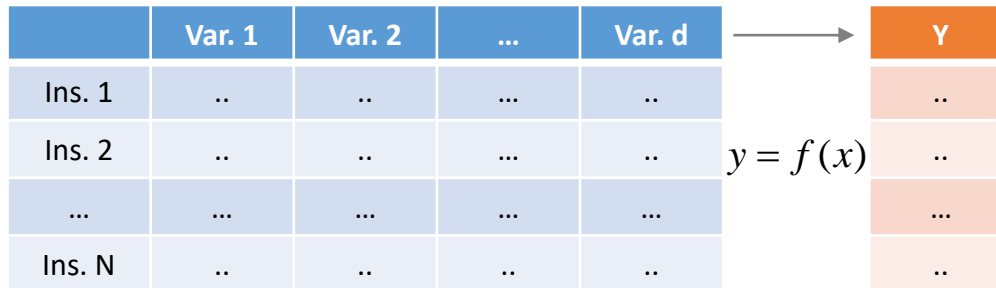
- Finds relations between X and Y.
- Estimate the underlying function $y = f(x)$.
- Classification, regression.

$$\mathcal{Y} = \{y_i | i = 1, \dots, n, y_i = f(\mathbf{x}_i)\}$$

$$y = f(x)$$

Type of Machine Learning/Data Mining

- Supervised Learning



Association Rule Mining



크게보기 | 미리보기

매장 재고 · 위치 >

키워드 Pick

안내

양로원

갱단

트렁크

데뷔작

율리우스

핵폭탄

이 책의 다른 상품 정보

sam : 한달 3권 9,900원 >

eBook : 9,000원 >

원서/번역서 :
[보유] The Hundred-Year-Old
Man Who Climbed Out of the
Window and Disappeared >

오늘의책 | 무료배송 | 소득공제

창문 넘어 도망친 100세 노인 요나스 요나손 장편소설

요나스 요나손 지음 | 임호경 옮김 | 열린책들 | 2013년 07월 25일 출간

★★★★★ 리뷰 112개 | 리뷰쓰기 | 9.0(137)

KBS TV책 -김창완과 책읽기 >

정가 : 13,800원

판매가 : **12,420원** [10%↓ 1,380원 할인]

통합포인트 : [기본적립] 690원 적립 [5% 적립] 안내

[추가적립] 5만원 이상 구매 시 2천원 추가적립

[회원혜택] 우수회원 5만원 이상 구매 시 2~3% 추가적립

추가혜택 : 카드/포인트 안내 | 도서소득공제 안내 | 추가혜택 더보기

배송비 : 무료 | 배송비 안내

배송일정 : 서울특별시 종로구 세종대로 기준 지역변경

03월 04일 출고 예정 | 배송일정 안내

바로드림 : 인터넷으로 주문하고 매장에서 직접 수령 안내

주문수량 1 + -

장바구니 담기

바로구매

바로드림 주문

선물하기

보관함 담기

Association Rule Mining

이 책을 구매하신 분들이 함께 구매하신 상품입니다

전체선택

장바구니 담기



참을 수 없는 존재의 가벼움(양장본)

13,500원



셈을 할 줄 아는 까막눈이 여자(큰글자판)

13,320원



셈을 할 줄 아는 까막눈이 여자

13,320원



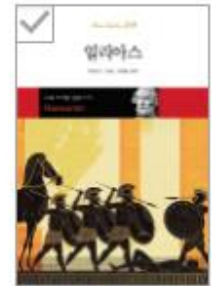
자신을 행성이라 생각한 여자

13,320원



마리아(Maria)(고려대학교 청소년문학 시리즈)

11,000원



일리아스(클래식 투게더 23)

10,620원

이 상품의 꾸러미



+



창문 넘어 도망친 100세 노인

The 100-Year-Old Man Who Climbed Out

창문 넘어 도망친 100세 노인 한영판 세트 (도서 2종)


25,640원

18,460원 [28%할인] | 690원 [4%적립]

자세히 보기

장바구니 담기

Association Rule Mining



All ▾ nespresso









Q

Tax Cen

Departments ▾ Prime ▾ Video ▾ Music ▾

Kitchen & Dining ▾ Best Sellers ▾ Wedding Registry ▾ Small Appliances ▾ Kitchen Tools ▾ Cookware ▾ Bakeware ▾ Cutlery ▾ Dining & Entertaining ▾ Storage & Organization ▾ Event & Party Supplies ▾ Shop by Room

[Back to search results for "nespresso"](#)



Nespresso VertuoLine Coffee and Espresso Maker with Aeroccino Plus Milk Frother, Black

by Nespresso

★★★★★ ▾ 684 customer reviews | 163 answered questions

List Price: \$249.00

Price: **\$199.05** & **FREE Shipping**. [Details](#)


You Save: **\$49.95** (20%)


In Stock.


Want it Wednesday, March 1? Order within **6 hrs 36 mins** and choose **Two-Day Shipping** at checkout. [Details](#)


Ships from and sold by Amazon.com. Gift-wrap available.

Color: **Black**

 **\$199.05**

 \$161.85











 \$229.13

 from \$150.00

- Includes Aeroccino Plus milk frother: rapid one touch preparation of hot or cold milk froth; Items sold separately valued at 398
- New revolutionary Centrifusion technology to gently brew both Coffee and Espresso with one touch of a button
- Capsule recognition and code reading technology for blend-specific parametric brewing; Two capsule sizes, large for Coffee and small for Espresso
- Easy insertion and ejection of capsules; For use with Nespresso VertuoLine capsules only; Not compatible with Nespresso Original Line capsules
- Removable used capsule container holds 13-20 used capsules; Automatic off mode after 9 minutes of inactivity; Fast heat up time 15 seconds

Customers Who Bought This Item Also Bought

Page 1 of 6



DecoBros Crystal Tempered Glass Nespresso VertuoLine Storage Drawer Holder...
★★★★★ 737
\$29.99 ✓Prime

Nespresso VertuoLine Best Seller Assortment, 30 Count
★★★★★ 15
\$42.46 ✓Prime

Nespresso VertuoLine Coffee Capsules Assortment - The Best Sellers: 1 Sleeve of...
★★★★★ 81
\$44.92 ✓Prime

Nespresso VertuoLine Voltesso Espresso, 10 Count
★★★★★ 17
\$11.00 ✓Prime

Nespresso VertuoLine Espresso Assortment, 50 Count
★★★★★ 16
\$48.85 ✓Prime

Nespresso VertuoLine Caramelizio Coffee, 10 Count
★★★★★ 30
\$11.00 ✓Prime

Nespresso VertuoLine Odacio Coffee, 10 Count
★★★★★ 21
\$11.00 ✓Prime

Nespresso VertuoLine Altissimo Espresso, 10 Count
★★★★★ 19
\$11.00 ✓Prime

Nespresso VertuoLine Diavolitto Espresso, 10 Count
★★★★★ 16
\$8.60

Nespresso VertuoLine Intense Assortment, 10 Count (Pack of 4)
★★★★★ 13
\$44.77 ✓Prime

Association Rule Mining

- Also known as “Market Basket Analysis”



Wall Mart (USA)

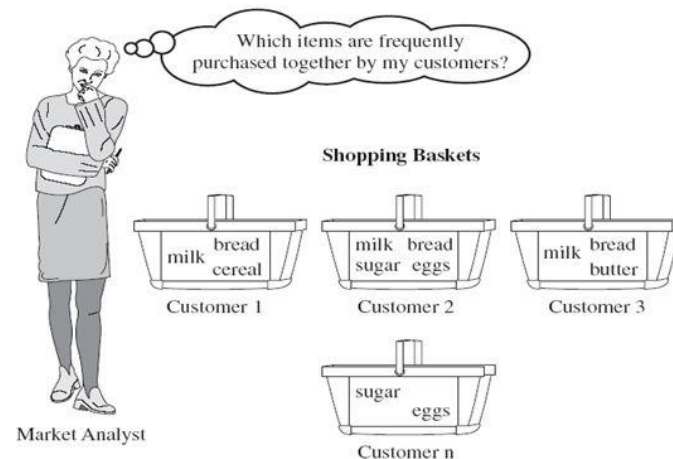


E-Mart (Korea)



Association Rule Mining

- Goal:
 - ✓ Produce rules that define “what goes with what”
 - ✓ “If X was purchased, then Y was also purchased”
- Features
 - ✓ Rows are transactions
 - ✓ Used in recommendation systems – “Our records show that you bought X, thus you may also like Y”
 - ✓ Also called “affinity analysis” or “market basket analysis”



Association Rule Mining

- Dataset for association rule mining
 - ✓ Each transaction is represented as a record
 - ✓ Two representations are possible: (1) item list and (2) item matrix

[Item list]

























Transaction ID	Items
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

[Item matrix]

Transaction ID	Bread	Milk	Diaper	Beer	Eggs	Coke
1	1	1	0	0	0	0
2	1	0	1	1	1	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	1

Association Rule Mining

- A toy example: a tiny retail market data

Transaction	Item 1	Item 2	Item 3	Item 4
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

Association Rule Mining

- Terminology
 - ✓ Antecedent – “IF” part
 - ✓ Consequent – “THEN” part
 - ✓ Item set – the items comprising the antecedent or consequent
 - ✓ Antecedent and consequent are **disjoint** (have no items in common)
- Generating rules
 - ✓ Many rules are possible (e.g., for transaction 1)
 - If egg is bought, then noddle is also bought
 - If egg and noddle are bought, then tuna is also bought
 - If tuna is bought, then egg is also bought, etc.

























Association Rule Mining

Performance Measures for the rule $A \rightarrow B$

- Support

$$\text{support}(A \rightarrow B) = P(A) \text{ or } P(A, B)$$

- ✓ Used to find the frequent item sets
- ✓ The higher the support, the higher the chance of applying the rule

Transaction	Item 1	Item 2	Item 3	Item 4
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				


























Association Rule Mining

Performance Measures for the rule $A \rightarrow B$

- Confidence

$$\text{confidence}(A \rightarrow B) = \frac{P(A, B)}{P(A)}$$

- ✓ The conditional probability of B given A
- ✓ Used to generate meaningful rules

Transaction	Item 1	Item 2	Item 3	Item 4
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

Association Rule Mining

























Performance Measures for the rule $A \rightarrow B$

- Confidence

$$\text{lift}(A \rightarrow B) = \frac{P(A, B)}{P(A) \cdot P(B)}$$

✓ Used to determine the usefulness of generated rules

- Lift = 1: A and B are statistically independent
- Lift > 1: Positive relationship between A and B
- Lift < 1: Negative relationship between A and B

Transaction	Item 1	Item 2	Item 3	Item 4
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

Association Rule Mining

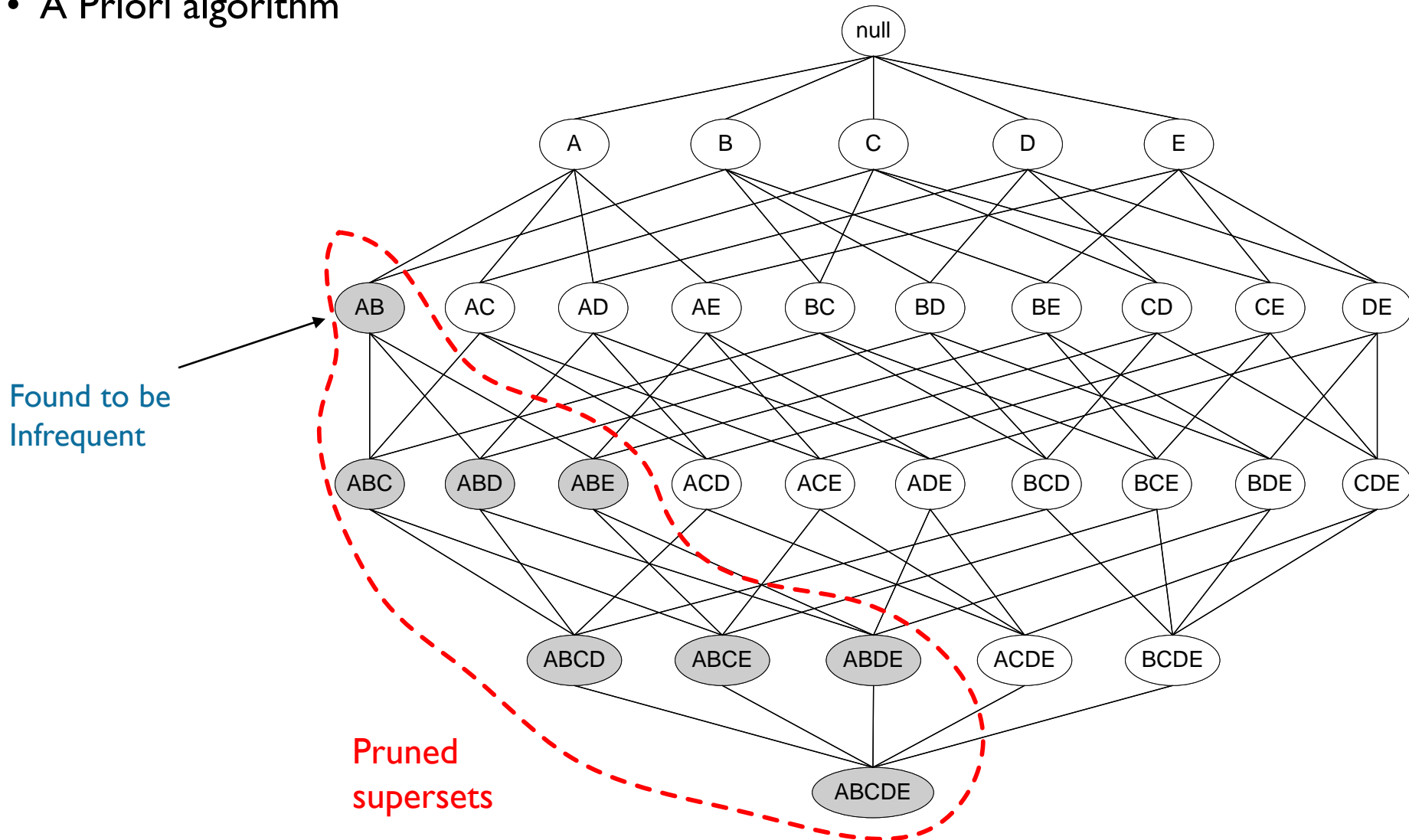
- How to generate an effective association rules?
 - ✓ Ideally, create all possible combinations of items and see what rules are effective and what rules are not.
 - ✓ Computation time grows exponentially as the number of items increases.
- Brute-force approach
 - ✓ List all possible association rules
 - ✓ Compute the support and confidence for each rule
 - ✓ Prune rules that fail the **minsup** and **minconf** threshold.
 - ✓ **Computationally prohibitive!**

Association Rule Mining

- A priori algorithm
 - ✓ Consider only “frequent item sets”
 - ✓ “support”
 - Criterion for item set frequency $P(A)$
 - # (%) of transactions that include both the antecedent and the consequent
 - Support for the item set {egg, noodle} is 4 out of transactions, or 40%
 - ✓ Support of an itemset never exceeds the support of its subsets, which is known as **anti-monotone** property of support.

Association Rule Mining

























- A Priori algorithm



Association Rule Mining

- Generating frequent item sets

✓ Users set a minimum support criterion: e.g. 2 transactions or 20%

Transaction	Item 1	Item 2	Item 3	Item 4
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

Association Rule Mining

- Generating frequent item sets
 - ✓ Generate the list of one-item sets that meets the support criterion



8 (80%)



5 (50%)



5 (50%)



3 (30%)



2 (20%)



1 (10%)

- ✓ Onion is removed because it does not meet the minimum support criterion

Association Rule Mining

- Generating frequent item sets

- ✓ Use the life of one-item sets to generate list of two-item sets that meet the support criterion

	noodle	egg	cola	rice	tuna
noodle		40%	40%	20%	20%
egg			30%	0%	20%
cola				0%	10%
rice					0%
tuna					

- ✓ {noodle, egg}, {noodle, cola}, {noodle, rice}, {noodle, tuna}, {egg, cola}, {egg, tuna} are frequent two-item sets

Association Rule Mining

- Generating frequent item sets
 - ✓ Use the list of two-item sets to generate the three-item sets.
 - ✓ Continue up through k-item sets.

Set-size	Item 1	Item 2	Item 3	...	Item 6
1	noodle				
1	egg				
1	cola				
1	rice				
1	tuna				
2	noodle	egg			
2	noodle	cola			
2	noodle	rice			
...			

Association Rule Mining

- A Priori algorithm
 - ✓ Let $k=1$
 - ✓ Generate frequent itemsets of length 1
 - ✓ Repeat until no new frequent itemsets are identified
 - Generate length $(k+1)$ candidate itemsets from length k frequent itemsets
 - Prune candidate itemsets containing subsets of length k that are infrequent
 - Count the support of each candidate by scanning the DB
 - Eliminate candidates that are infrequent, leaving only those that are frequent

Association Rule Mining

- Confidence

- ✓ The % of antecedent transactions that also have the consequent item set
- ✓ E.g. “if noodle is purchased, then egg is also purchased”

$$\text{support}(\text{noodle}) = P(\text{noodle}) = \frac{8}{10}, \quad \text{support}(\text{egg}) = P(\text{egg}) = \frac{5}{10}$$

$$\text{confidence}(\text{noodle} \rightarrow \text{egg}) = \frac{P(\text{noodle}, \text{egg})}{P(\text{noodle})} = \frac{4/10}{8/10} = 0.5(50\%)$$

$$\text{lift}(\text{noodle} \rightarrow \text{egg})$$

$$\begin{aligned} &= \frac{\text{confidence}(\text{noodle} \rightarrow \text{egg})}{\text{support}(\text{egg})} = \frac{\frac{P(\text{noodle}, \text{egg})}{P(\text{noodle})}}{P(\text{egg})} = \frac{P(\text{noodle}, \text{egg})}{P(\text{noodle}) \times P(\text{egg})} \\ &= \frac{\frac{4}{10}}{\frac{8}{10} \times \frac{5}{10}} = 1 \end{aligned}$$

Association Rule Mining

- Generated rules
 - ✓ Set the support to 20%.
 - ✓ Set the confidence to 70%.

Rule #	Antecedent (a)	Consequent	Support	Confidence	Lift
1	tuna=>	egg, noodle	2	100	2.5
2	tuna=>	egg	2	100	2
3	noodle, tuna=>	egg	2	100	2
4	rice=>	noodle	3	100	1.25
5	egg, tuna=>	noodle	2	100	1.25
6	tuna=>	noodle	2	100	1.25
7	cola=>	noodle	5	80	1
8	egg=>	noodle	5	80	1

Association Rule Mining

- Summary
 - ✓ Produce rules on associations between items from a database of transactions
 - ✓ Widely used in recommender systems
 - ✓ Most popular method is A-priori algorithm
 - ✓ To reduce computation, consider only “frequent” item sets (=support)
 - ✓ Performance is measured by confidence and lift

