THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Business Analytics (MM 5425)

L8. Association Rule

Dr. Yue (Katherine) FENG

Recap: Supervised vs. Unsupervised Learning

> **Supervised learning:** learns a model that predicts target outcome based on a set of other attributes/features (i.e., training data where target value is known).

- Stock price prediction (numerical target variable)
- Credit card default (binary target variable)

> **Unsupervised learning:** finds relationships in the data without reference to target variable.

- Beer and diaper

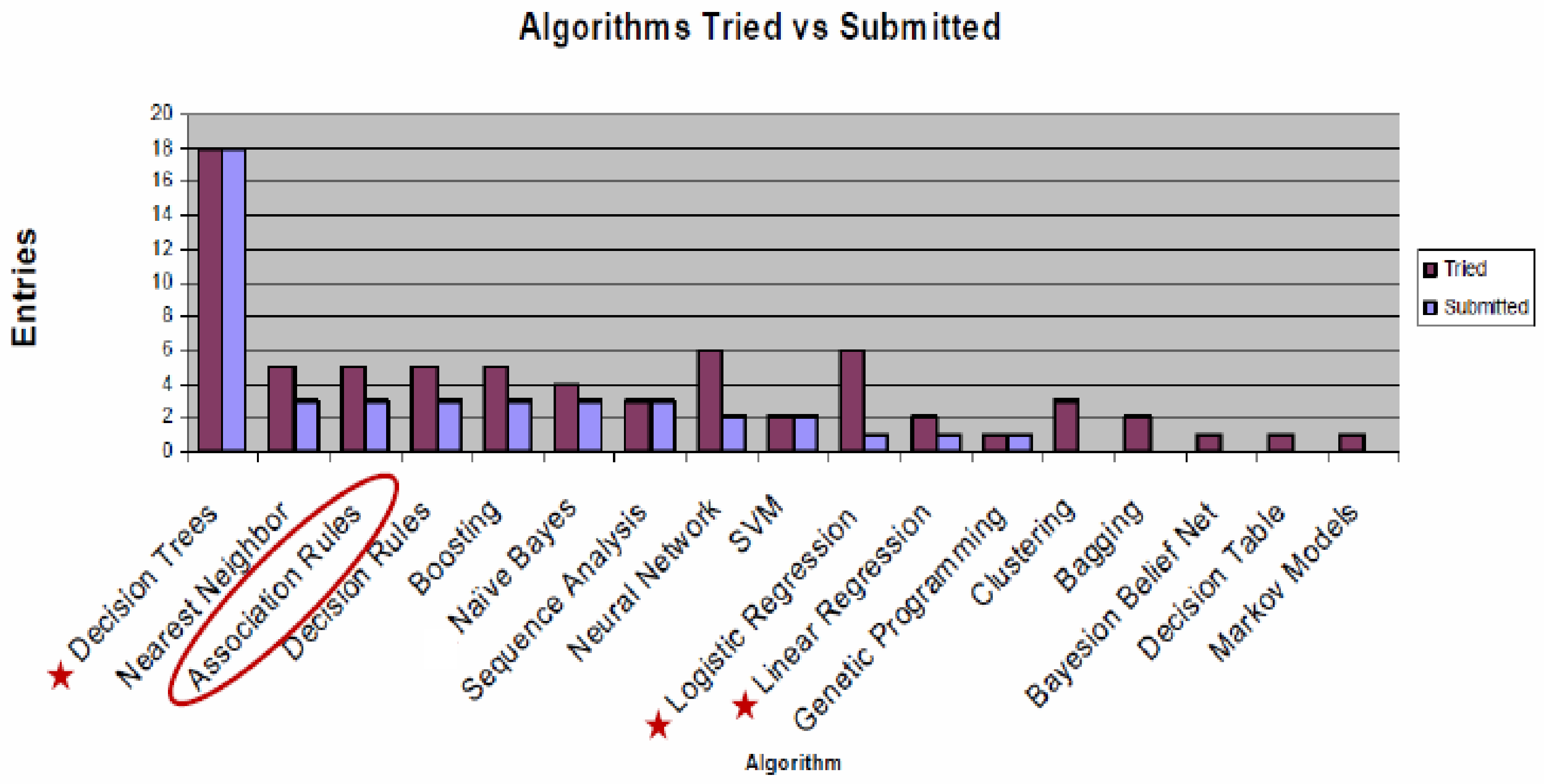
> Key: is there a specific **target variable** that we are trying to predict?

Unsupervised Learning

> How do I find items that occur together more than I might expect by chance?
- **Associations (relationship between columns)**

> How do I find natural groupings of data instances?
- **Clustering (relationship between rows/instances)**

Commonly Used Algorithms



Association Rule Learning

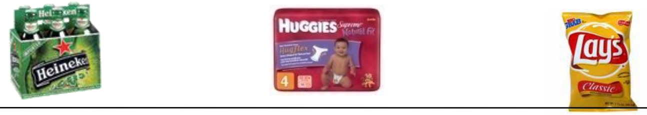
Market-basket Analysis (Associations Rule Learning)

> Are some items shopped together more than I might expect?

- With this information, I could:
  - Put them close to each other in the store
  - Make suggestions/bundles on a website

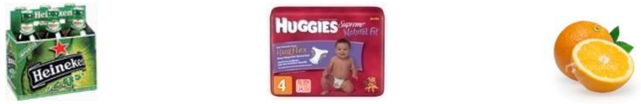
Market Basket Data

Transaction NO.	Item 1	Item 2	Item 3	…
1	Beer	Diapers	Chips	
2	Diaper	Orange		
3	Diaper	Milk		
4	Beer	Diaper	Orange	
5	Beer	Detergent		
…				

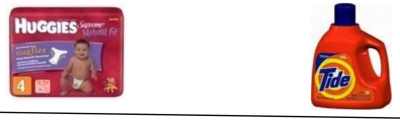
1. 

2.

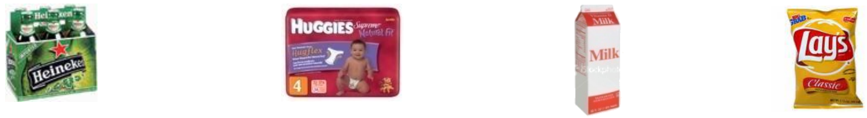
3.

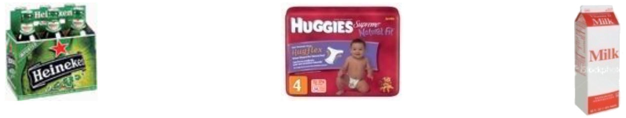
4. 

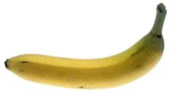
5.

6. 

7.

8. 

9. 

10. 

Association "Rules" – Standard Format

- Rule format: If {set of items} → Then {set of items}

IF

CONDITION (*C*)

RESULT (*R*)

If {beer} -> {diaper}

What is an Interesting Association?

> Some standard measures used for rule C -> R:

- **Support(R, C):** p(R&C)
  - proportion of transactions (“baskets”) that contain both R and C.
- **Confidence(C->R):** p(R|C)
  - proportion of transactions that R holds when C holds.
- **Lift and Leverage(C->R)**

How do We Calculate Probabilities?

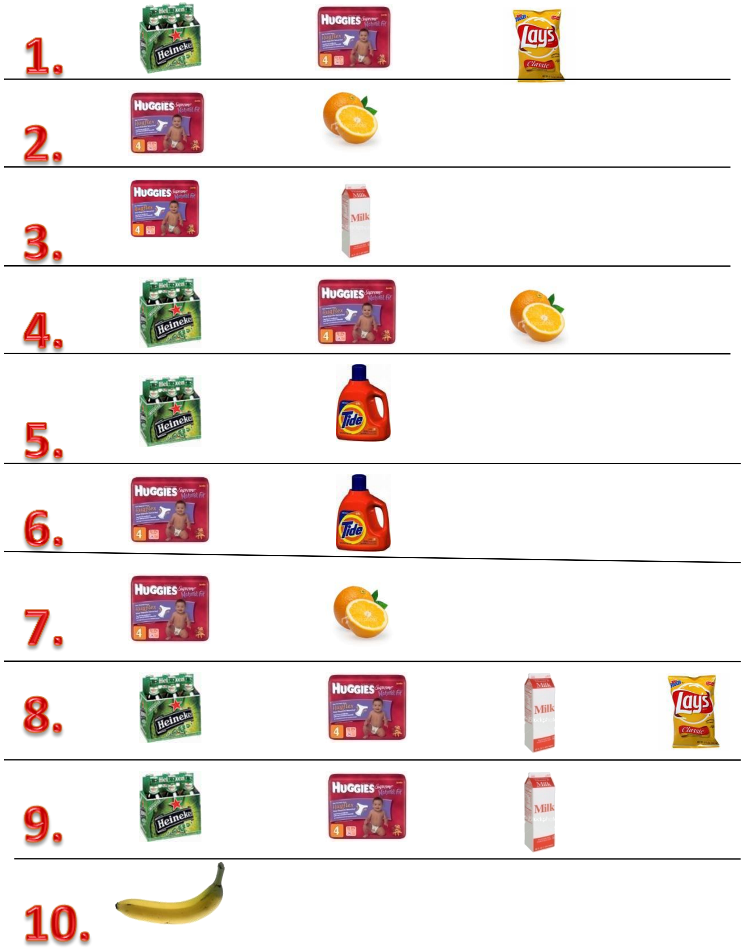
Again, count !!!

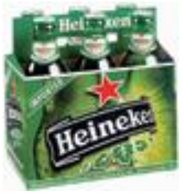
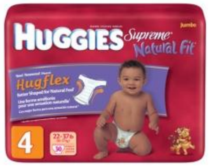
Support

> Support: how popular an item is, as measured by the proportion of transactions that contain an item.

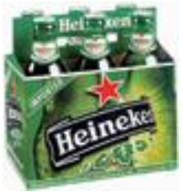
$$\text{Support}(X) = \frac{\text{\# transactions that contain X}}{\text{\# total transactions}}$$

Support



{  ,  } = 4

→ Support = 4/10 = 40%

{  } = 5

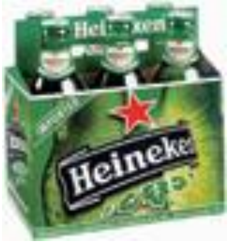
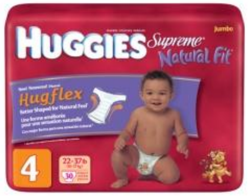
→ Support = 5/10 = 50%

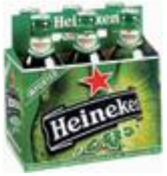
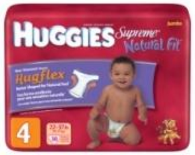
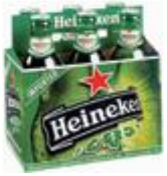
Confidence

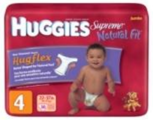
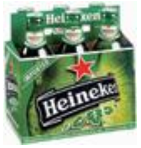
- Confidence (C->R): how often the association rule has been found to be true, as measured by the proportion of transactions that R holds when C holds.

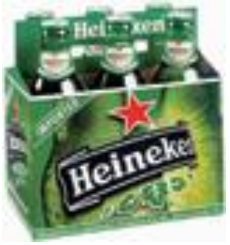
$$Confidence\ (C \rightarrow R) = \frac{Support\ (R, C)}{Support\ (C)}$$

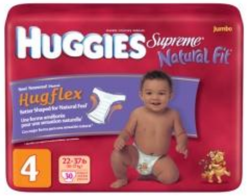
Confidence

IF  → 

Confidence = # {  ,  } / # {  } $= \frac{4}{5} = 80\%$

Confidence for this association rule is the likelihood that a transaction contains  given that it contains 

IF 

→ 

80% Confidence …… Any problems?

What if Many People buy diaper ?

{ } = 8

80% Prevalence of

……the confidence will be high for any item set (association) that **contains diapers as result**.

Important Measure: Lift (C->R)

Lift: measured by the ratio of the observed support to the expected support if C and R are independent. ➡

$$\text{Lift} = \frac{p(R\&C)}{p(R)\ p(C)} = \frac{40\%}{80\% * 50\%} = 1$$

For the association rule to be meaningful, the Lift must > 1

An Alternative: Leverage

Leverage: measured by the difference of the observed support to the expected support if C and R are independent.

➡

Leverage = p(R&C) - p(R) p(C) = 40% - 40% = 0

For the association rule to be meaningful, the Leverage must > 0

Exercise

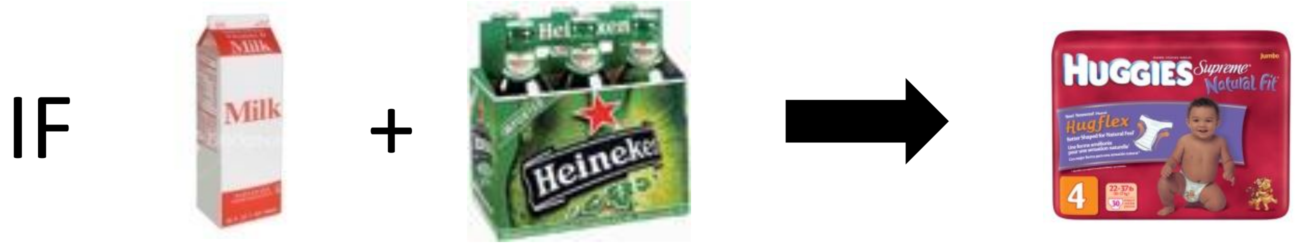
IF

→

What are the Confidence, Lift, and Leverage?

a) 50%, 1, 0
b) 50%, 1.2, 0
c) 70%, 0.8, 1
d) 40%, 0.75, 1
e) None of the above

Associations for More Than Two Items

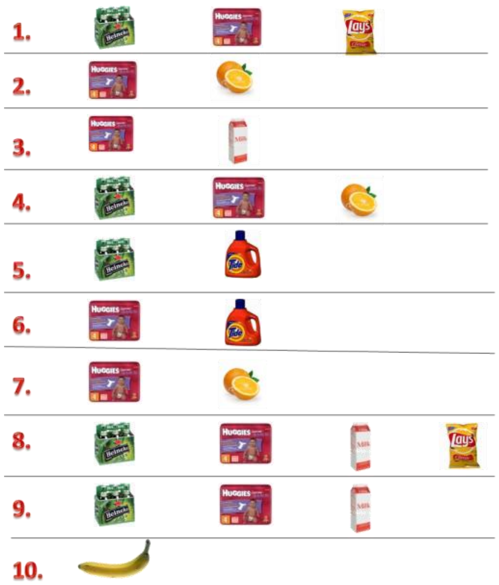


> Support = 2/10

> Confidence = 0.2/0.2 = 1

> Lift = 0.2/0.2*0.8=1.25

> Leverage = 0.2 – 0.2*0.8 = 0.04


1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

How to Find “Interesting” Associations?

> By setting threshold for being an “interesting” association

- e.g., $support \geq 0.3$, or $confidence \geq 0.5$, or both

> A common strategy in association rule learning algorithms has 3 steps:

1. Frequent itemset generation: find all itemsets with support that is greater than the minimum support threshold.
2. Rule generation: extract all high confidence rules from the frequent itemsets.
3. Rule examination: use lift/leverage to remove spurious rules (it is not just a coincidence).

Association Rule: Other Applications

> “Item” can be any features:

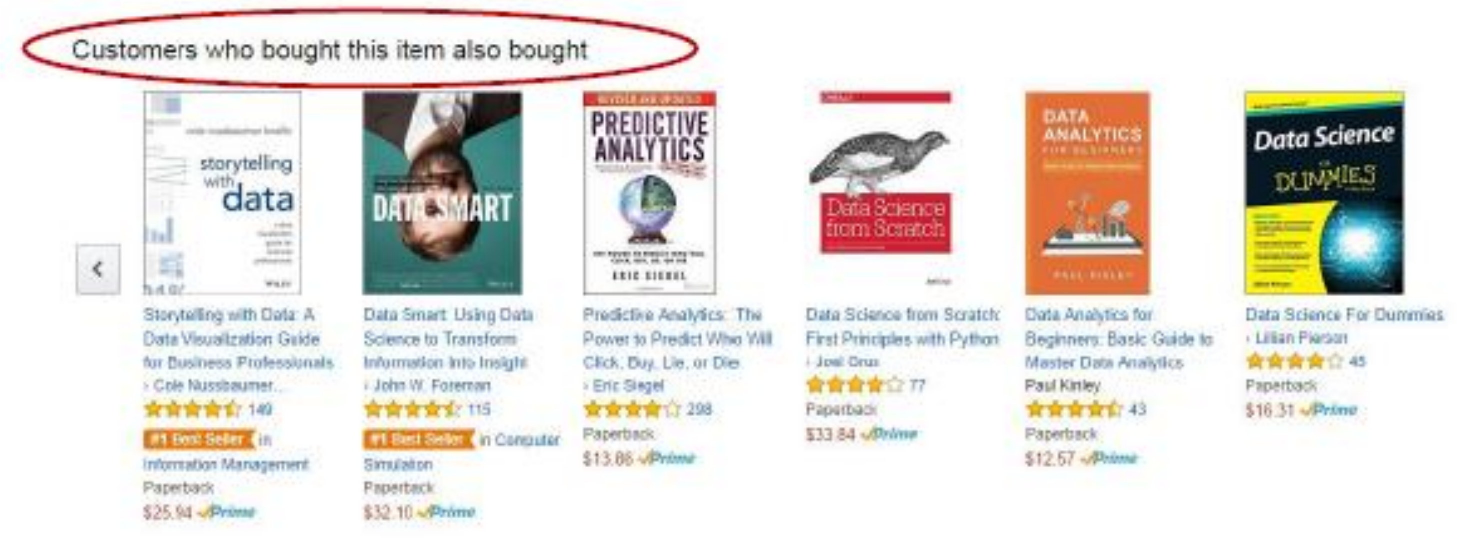
- Owns-luxury-vehicle => Frequent-purchaser
- age(“30 - 39”) & income(“42 - 48K”) => buys(“car”)

> Association mined from Facebook:

- Status=Undergrad & Political_Views=Liberal
  => Interested_in_Men <lift:(1.66)>

Discussion

> How to use association rule learning in recommender systems? What are the transactions and what are the items?



Customers who bought this item also bought

Storytelling with Data: A Data Visualization Guide for Business Professionals
› Cole Nussbaumer...
★★★★½ 149
#1 Best Seller in Information Management
Paperback
$25.94 ✓Prime

Data Smart: Using Data Science to Transform Information Into Insight
› John W. Foreman
★★★★½ 115
#1 Best Seller in Computer Simulation
Paperback
$32.10 ✓Prime

Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die
› Eric Siegel
★★★★☆ 298
Paperback
$13.86 ✓Prime

Data Science from Scratch: First Principles with Python
› Joel Grus
★★★★☆ 77
Paperback
$33.84 ✓Prime

Data Analytics for Beginners: Basic Guide to Master Data Analytics
Paul Kinley
★★★★½ 43
Paperback
$12.57 ✓Prime

Data Science For Dummies
› Lillian Pierson
★★★★☆ 45
Paperback
$16.31 ✓Prime

Associations: Pros and Cons

> Pros

- Can quickly mine patterns describing business/customers, etc. without major effort in problem formulation
- Unparalleled tool for hypothesis generation

> Cons

- Unfocused
  - Not clear exactly how to apply mined “knowledge”
- Can produce many, many rules!
  - May only be a few nuggets among them (or none)

Thank You !

IDEAS
Innovation-driven Education and Scholarship
Discover · Design · Deliver

Faculty of Business
工商管理學院

Department of
MANAGEMENT & MARKETING
管理及市場學系

Opening Minds • Shaping the Future
啟迪思維 • 成就未來