ECLT 5810/SEEM5750
E-Commerce Data Mining Techniques - Introduction

Prof. Lam Wai

Data Opportunities

- Business infrastructure have improved the ability to collect data
- Virtually every aspect of business is now open to data collection:
  - customer behavior, marketing campaign performance, supply-chain management, workflow procedures, etc
- Information is now widely available on external events
  - market trends, industry news, etc
- Broad availability of data
  - increasing interest in methods for extracting useful information and knowledge from data

Data Opportunities

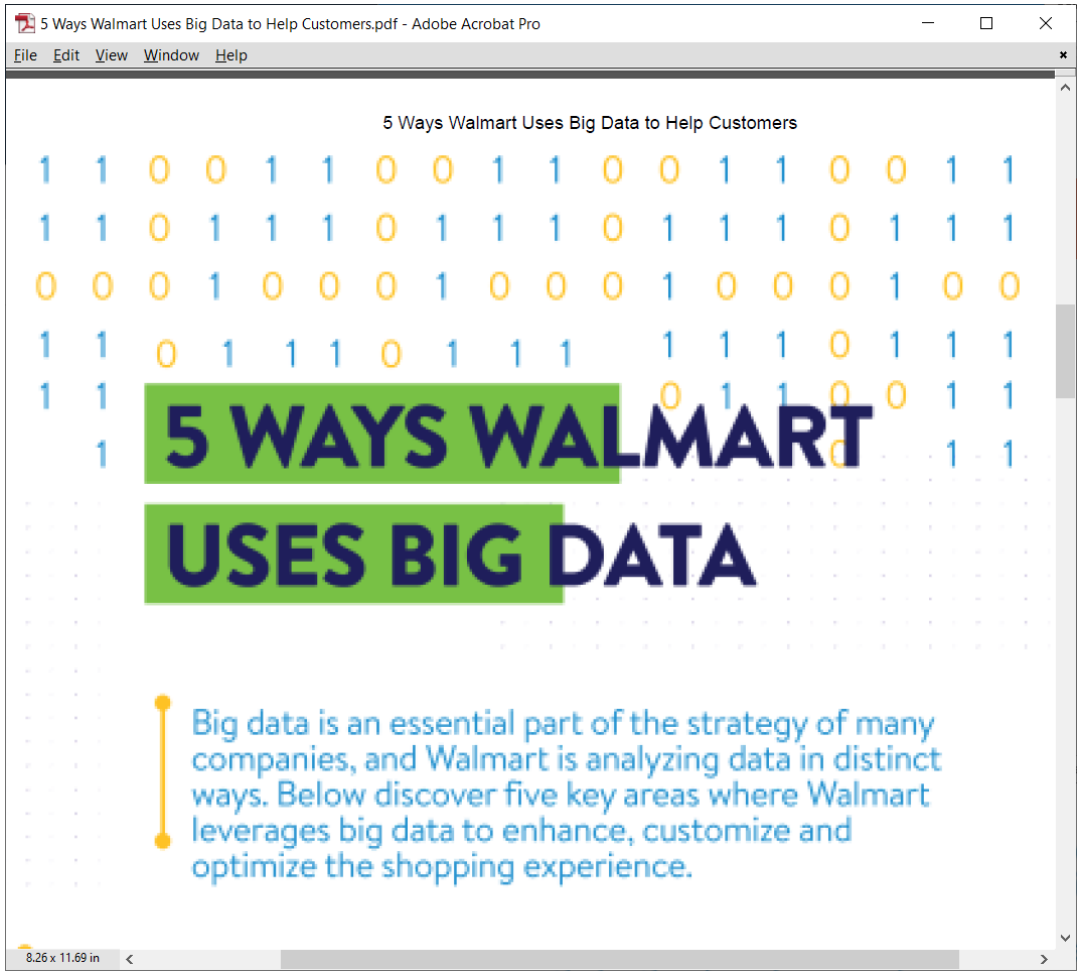
- Companies focus on exploiting data for competitive advantage
- In the past, statisticians and analysts explore datasets manually
  - but the volume and variety of data have far outstripped the capacity of manual analysis
- Computers have become far more powerful
  - algorithms have been developed that can connect datasets to enable broader and deeper analyses
- Given rise to the increasing widespread business application of data mining techniques

Data Mining Adoption

- Data Mining – extraction of useful knowledge from data
- Knowledge may refer to - models, rules, regularities, patterns
  - non-trivial, implicit, previously unknown
- Used for general customer relationship management
  - analyze customer behavior in order to manage attrition and maximize expected customer value
- Used for credit scoring and trading, fraud detection, and workforce management
- Major retailers from Walmart to Amazon apply data mining from marketing to supply-chain management.

A Data Mining Case in Retail Industry - Walmart

- Largest retailer in the world
- Over ten thousand stores and about a trillion dollars in annual revenue



A Data Mining Case in Retail Industry - Walmart

- Collects huge amount of information from millions of customers every hour
- Employs data analytics to improve operational efficiency and marketing campaigns

A Data Mining Case in Retail Industry - Walmart

- Collects petabytes of information from millions customers every hour
- Employs data analytics to improve operational efficiency and marketing campaigns
  - Correlates sales trends with external events such as trending in Twitter, local weather, etc.
    - Sales pattern finding: Strawberry pop-tarts sales increased by 7 times before a hurricane (storm)

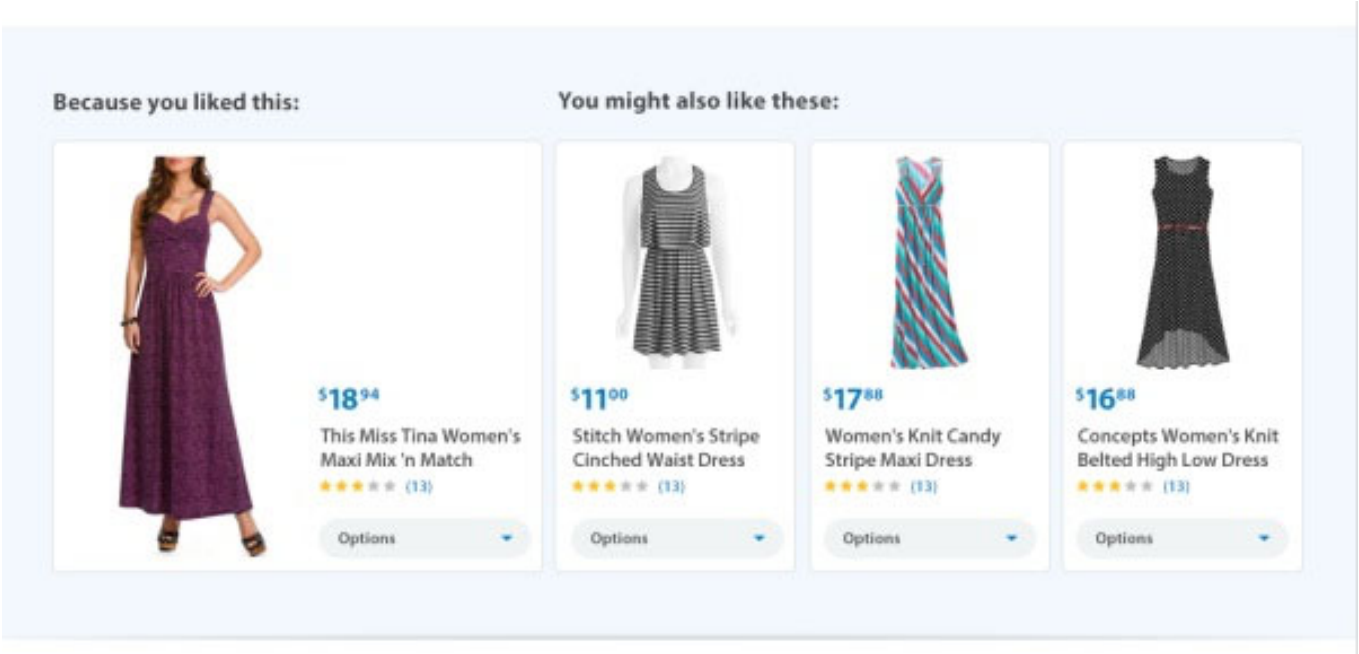
  - Find patterns that can be used to provide personalized product recommendations and promotional offers

A Data Mining Case in Retail Industry - Walmart

- Connecting in-store and online customer behavior
  - Combine information from online and in-store sources
  - Track when shoppers are looking at a product online using in-store Wi-Fi.
- Personalized Recommendation
  - Social Genome: Reach customers who have mentioned something on social media to inform them about that product and include a discount.

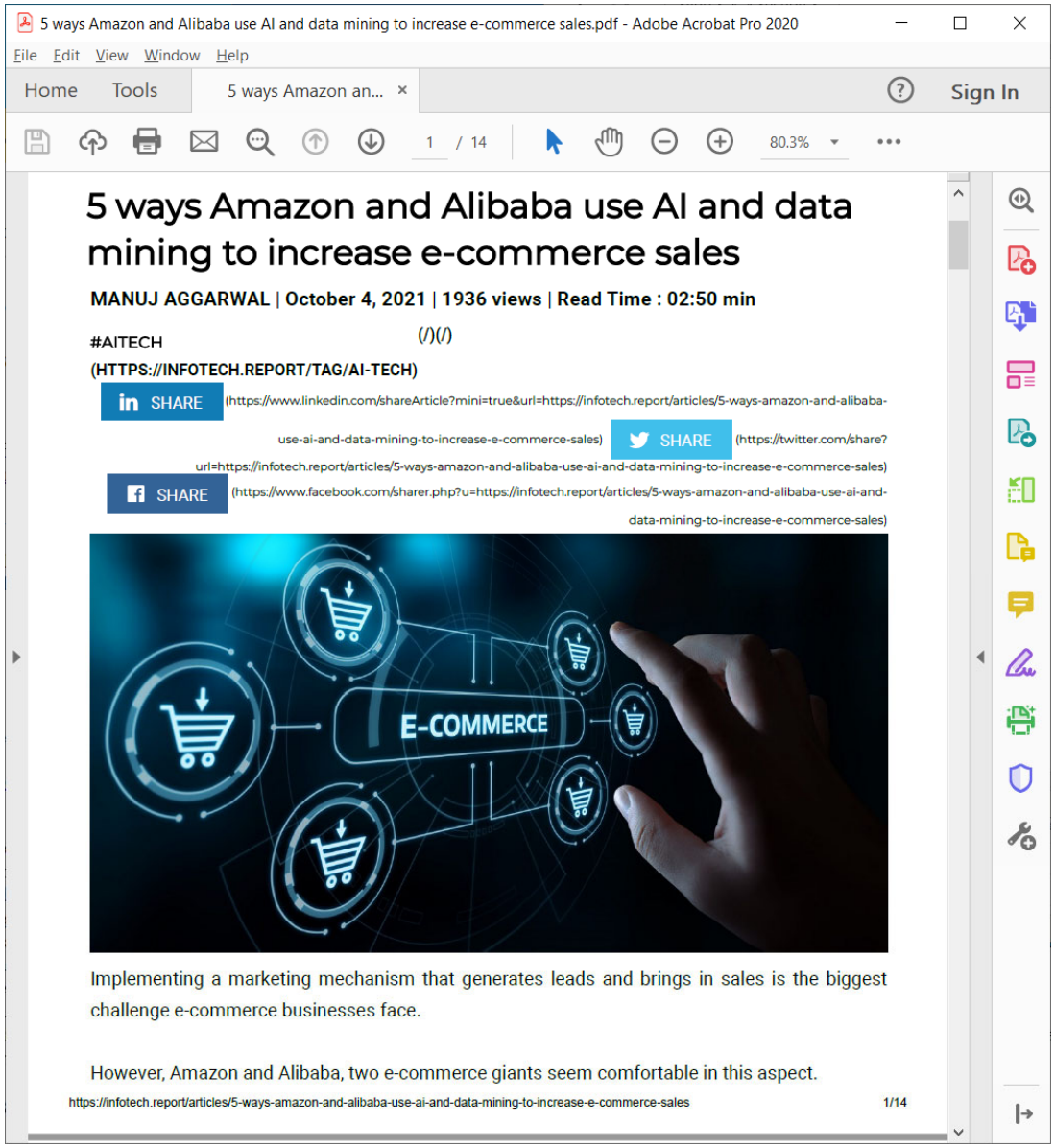
A Data Mining Case in Retail Industry - Walmart

- Personalized Recommendation (cont')
  - Send tailored product offers based on customer behavior online and inside their stores.



  - Track campaign's email open rate and re-align delivery times based on user patterns.
  - Email subject line personalization

Alibaba uses AI and Data Mining



Alibaba uses AI and Data Mining

- Product Recommendation
  - Alibaba has developed a software called “E-commerce Brain” which uses real-time online data to predict consumer wants.
- Smarter Pricing
  - Source different types of products worldwide, compare their prices and come up with the best price among all.
- Build Brand Loyalty with Consumers

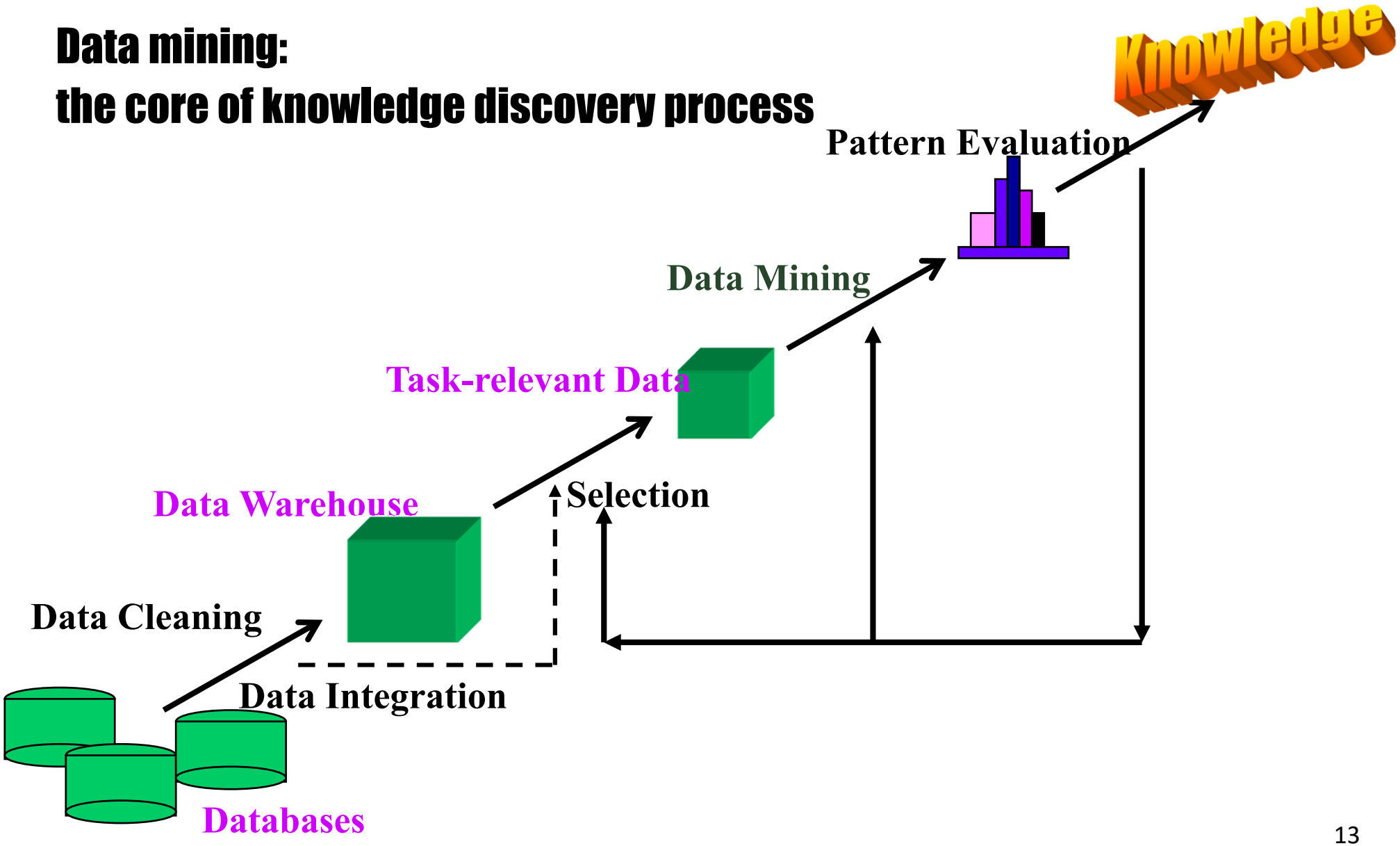
Data-Driven Decision Making (DDD)

- In Walmart and Alibaba examples, the data analysis was not just testing a simple hypothesis.
- Instead, the data were explored with the hope that something useful would be discovered.
- Also referred to as data mining, predictive analytics, business intelligence

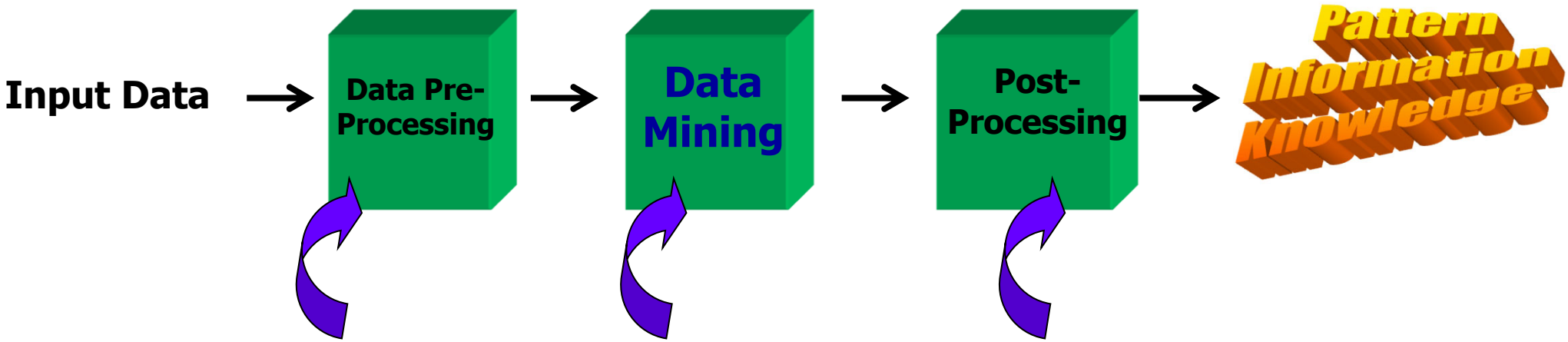
Video: What is Data Mining (from IBM Technology)
https://www.youtube.com/watch?v=7rs0i-9nOjo

Data Mining: A General KDD Process

Data mining:
the core of knowledge discovery process



Data Mining Procedures



Data integration
Normalization
Feature selection
Dimension reduction

Pattern discovery
Association & correlation
Classification
Clustering
Outlier analysis
… … … …

Pattern evaluation
Pattern selection
Pattern interpretation
Pattern visualization

- This is a view from typical machine learning and statistics communities

Steps of a General KDD Process

- Learning the application domain:
  - relevant prior knowledge and goals of application
- Creating a target data set: data selection
- Data cleaning and preprocessing: (may take 60% of effort!)
- Data reduction and transformation:
- Find useful features, dimensionality/variable reduction, invariant representation.
- Choosing functions of data mining
  - summarization, classification, regression, association, clustering.
- Choosing the mining algorithm(s)
- Pattern evaluation and knowledge presentation

Data Mining: Functionalities (1)

- Classification and Prediction
  - Finding models (functions) that describe and distinguish classes or concepts for future prediction
    - e.g., classify countries based on climate, or identify good clients
  - Model: decision-tree, classification rule, neural network

Data Mining: Functionalities (2)

- Cluster analysis
  - Class label is unknown: Group data to form new classes
    - e.g., cluster houses to find distribution patterns
  - Clustering based on the principle: maximizing the intra-class similarity and minimizing the interclass similarity

Data Mining: Functionalities (3)

- Association (correlation and causality)
  - age(X, “20..29”) ^ income(X, “20K..29K”)
    → buys(X, “PC”)
    [support = 2%, confidence = 60%]

Data Mining: Relationship to Other Disciplines