بسمه تعالی

Data Mining: Association Rule Discovery

Dr O. Pournik *MD, MPH, MSc, PhD*

pournik@gmail.com

Association Rule Discovery: Definition

- *Given a set of records each of which contain some number of items from a given collection;*
  - *Produce dependency rules which will predict occurrence of an item based on occurrences of other items.*

TID	*Items*
1	Bread, Cheese, Milk
2	Butter, Bread
3	Croissant, Jam, Milk
4	Croissant, Butter, Milk
5	Butter, Cheese, Bread, Milk
6	Bread, Cheese, Vegs

Association Rule Discovery: Application 1

- ***Supermarket shelf management.***
  - ***Goal:*** *To identify items that are bought together by sufficiently many customers.*
  - ***Approach:*** *Process the point-of-sale data collected with barcode scanners to find dependencies among items.*
  - *A classic rule --*
    - *If a customer buys Croissant, then he is very likely to buy milk.*
    - *So, don't be surprised if you find six-packs milk stacked next to Croissant!*

Association Rule Discovery: Application 2

- ***Inventory Management:***
  - ***Goal:*** *A consumer appliance repair company wants to anticipate the nature of repairs on its consumer products and keep the service vehicles equipped with right parts to reduce on number of visits to consumers.*
  - ***Approach:*** *Process the data on tools and parts required in previous repairs at different consumer locations and discover the co-occurrence patterns.*

بسمه تعالی

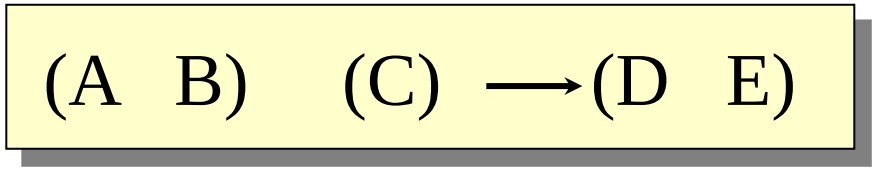
Data Mining: Others

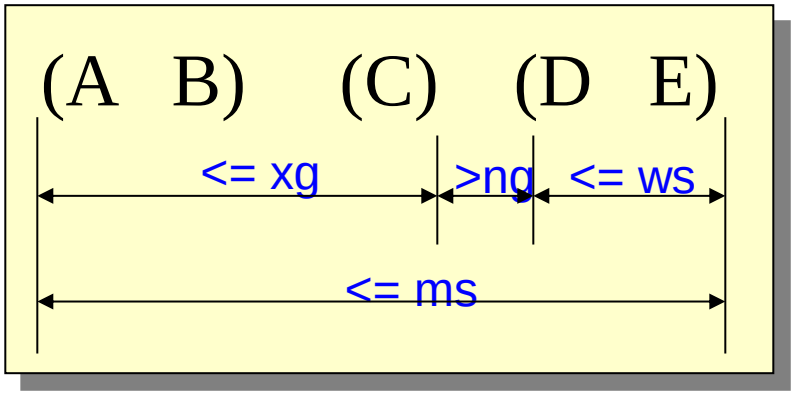
Dr O. Pournik *MD, MPH, MSc, PhD*

pournik@gmail.com

Sequential Pattern Discovery: Definition

- ***Given is a set of objects, with each object associated with its own timeline of events, find rules that predict strong sequential dependencies among different events.***
- ***Rules are formed by first discovering patterns. Event occurrences in the patterns are governed by timing constraints.***





Sequential Pattern Discovery: Examples

- ***In telecommunications alarm logs,***
  - ***(Inverter_Problem Excessive_Line_Current)***
    (Rectifier_Alarm) --> (Fire_Alarm)
- ***In point-of-sale transaction sequences,***
  - ***Computer Bookstore:***
    (Intro_To_Visual_C) (C++_Primer) -> (Perl_for_dummies)
  - ***Athletic Apparel Store:***
    (Shoes) (Racket, Racketball) --> (Sports_Jacket)

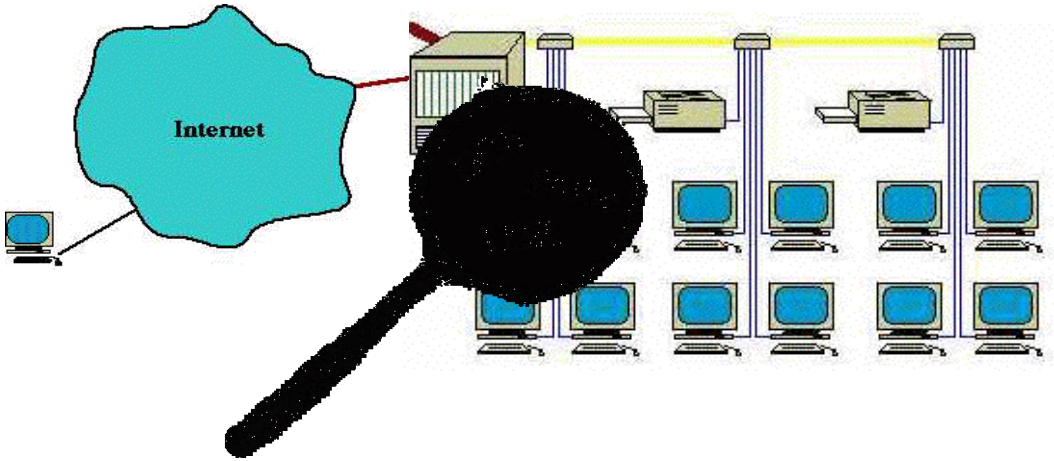
Regression

- ***Predict a value of a given continuous valued variable based on the values of other variables, assuming a linear or nonlinear model of dependency.***
- ***Greatly studied in statistics.***
- ***Examples:***
  - ***Predicting sales amounts of new product based on advertising expenditure.***
  - ***Predicting wind velocities as a function of temperature, humidity, air pressure, etc.***
  - ***Time series prediction of stock market indices.***

Deviation/Anomaly Detection

- *Detect significant deviations from normal behavior*
- *Applications:*
  - *Credit Card Fraud Detection*

  - *Network Intrusion Detection*



Challenges of Data Mining

- ***Scalability***
- ***Dimensionality***
- ***Complex and Heterogeneous Data***
- ***Data Quality***
- ***Data Ownership and Distribution***
- ***Privacy Preservation***
- ***Streaming Data***

Any Questions?