DATA MINING
Introductory and Advanced Topics
Part I

Source :
Margaret H. Dunham
Department of Computer Science and Engineering
Southern Methodist University

Companion slides for the text by Dr. M.H.Dunham, *Data Mining, Introductory and Advanced Topics*, Prentice Hall, 2002.

Data Mining Outline

- ***PART I***
  - ***Introduction***
  - ***Related Concepts***
  - ***Data Mining Techniques***
- PART II
  - Classification
  - Clustering
  - Association Rules
- PART III
  - Web Mining
  - Spatial Mining
  - Temporal Mining

Introduction Outline

Goal: Provide an overview of data mining.

- Define data mining
- Data mining vs. databases
- Basic data mining tasks
- Data mining development
- Data mining issues

Introduction

- Data is growing at a phenomenal rate
- Users expect more sophisticated information
- How?

UNCOVER HIDDEN INFORMATION

DATA MINING

Data Mining Definition

- Finding hidden information in a database
- Fit data to a model
- Similar terms
  - Exploratory data analysis
  - Data driven discovery
  - Deductive learning

Database Processing vs. Data Mining Processing

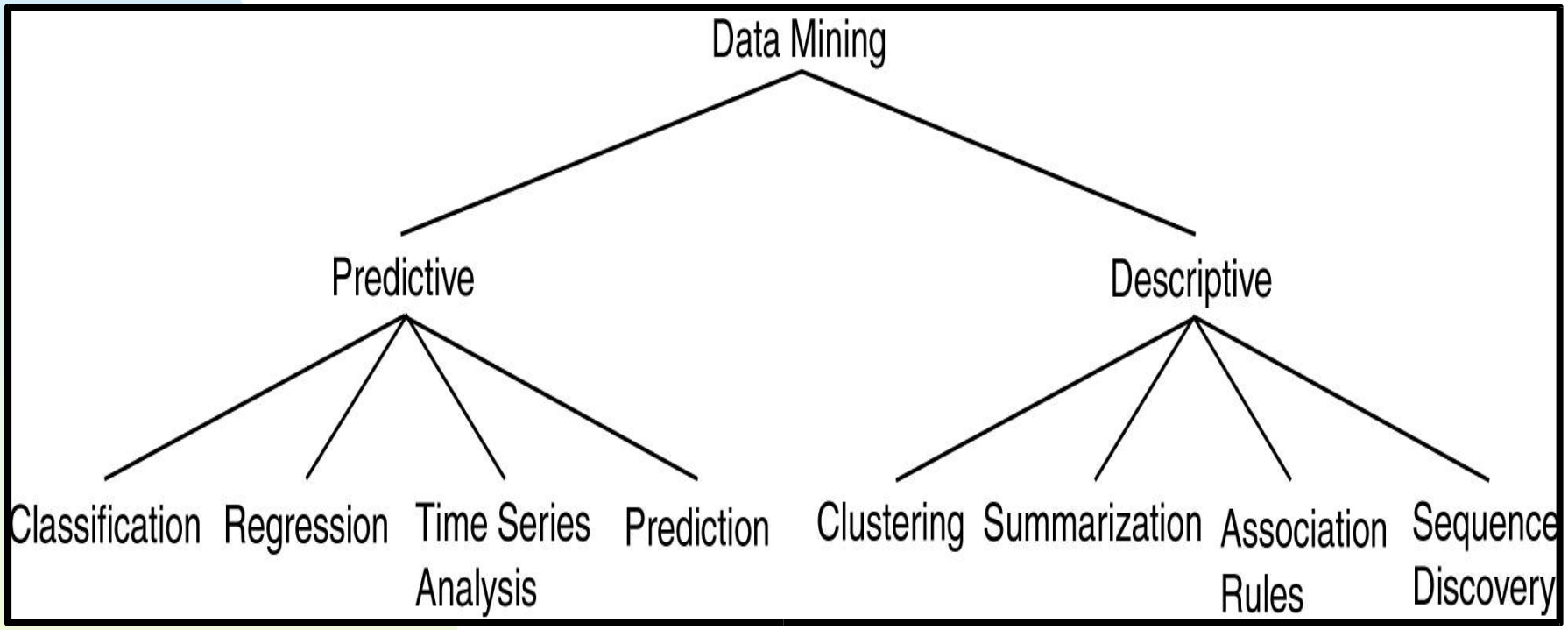
- Query
  - Well defined
  - SQL
- Data
  - Operational data
- Output
  - Precise
  - Subset of database

- Query
  - Poorly defined
  - No precise query language
- Data
  - Not operational data
- Output
  - Fuzzy
  - Not a subset of database

Query Examples

- Database
  - Find all credit applicants with last name of Smith.
  - Identify customers who have purchased more than $10,000 in the last month.
  - Find all customers who have purchased milk
- Data Mining
  - Find all credit applicants who are poor credit risks. (classification)
  - Identify customers with similar buying habits. (Clustering)
  - Find all items which are frequently purchased with milk. (association rules)

Data Mining Models and Tasks



Basic Data Mining Tasks

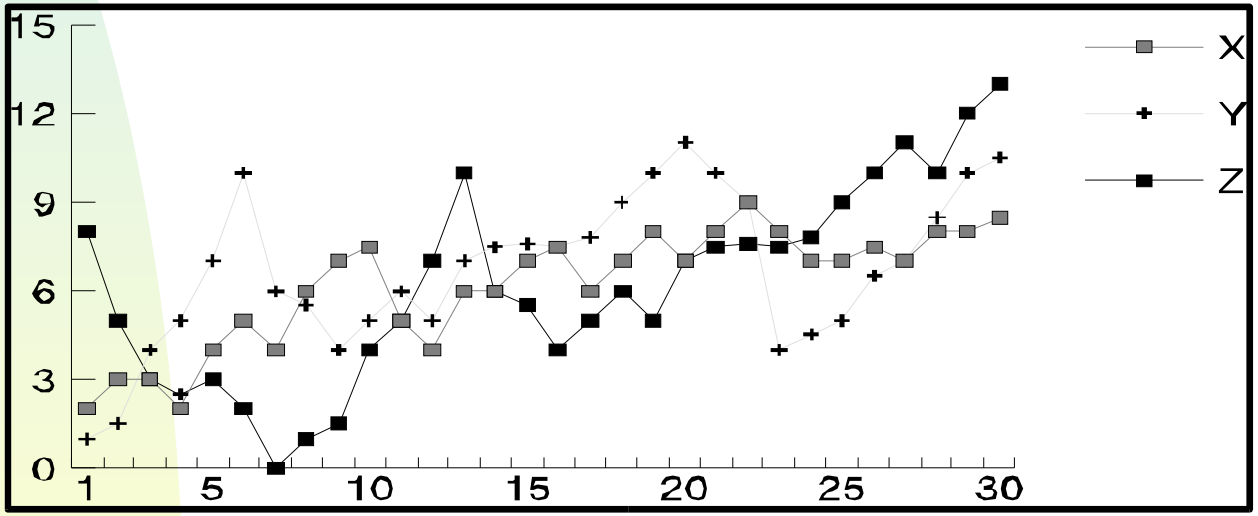
- ***Classification*** maps data into predefined groups or classes
  - Supervised learning
  - Pattern recognition
  - Prediction
- ***Regression*** is used to map a data item to a real valued prediction variable.
- ***Clustering*** groups similar data together into clusters.
  - Unsupervised learning
  - Segmentation
  - Partitioning

Basic Data Mining Tasks (cont'd)

- ***Summarization*** maps data into subsets with associated simple descriptions.
  - Characterization
  - Generalization
- ***Link Analysis*** uncovers relationships among data.
  - Affinity Analysis
  - Association Rules
  - Sequential Analysis determines sequential patterns.

Ex: Time Series Analysis

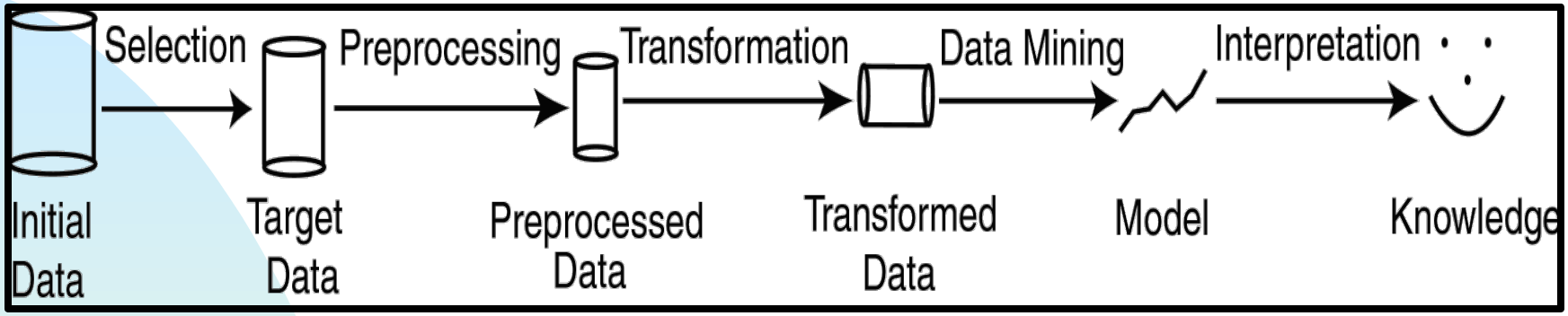
- Example: Stock Market
- Predict future values
- Determine similar patterns over time
- Classify behavior



Data Mining vs. KDD

- ***Knowledge Discovery in Databases (KDD):*** process of finding useful information and patterns in data.
- ***Data Mining:*** Use of algorithms to extract the information and patterns derived by the KDD process.

KDD Process



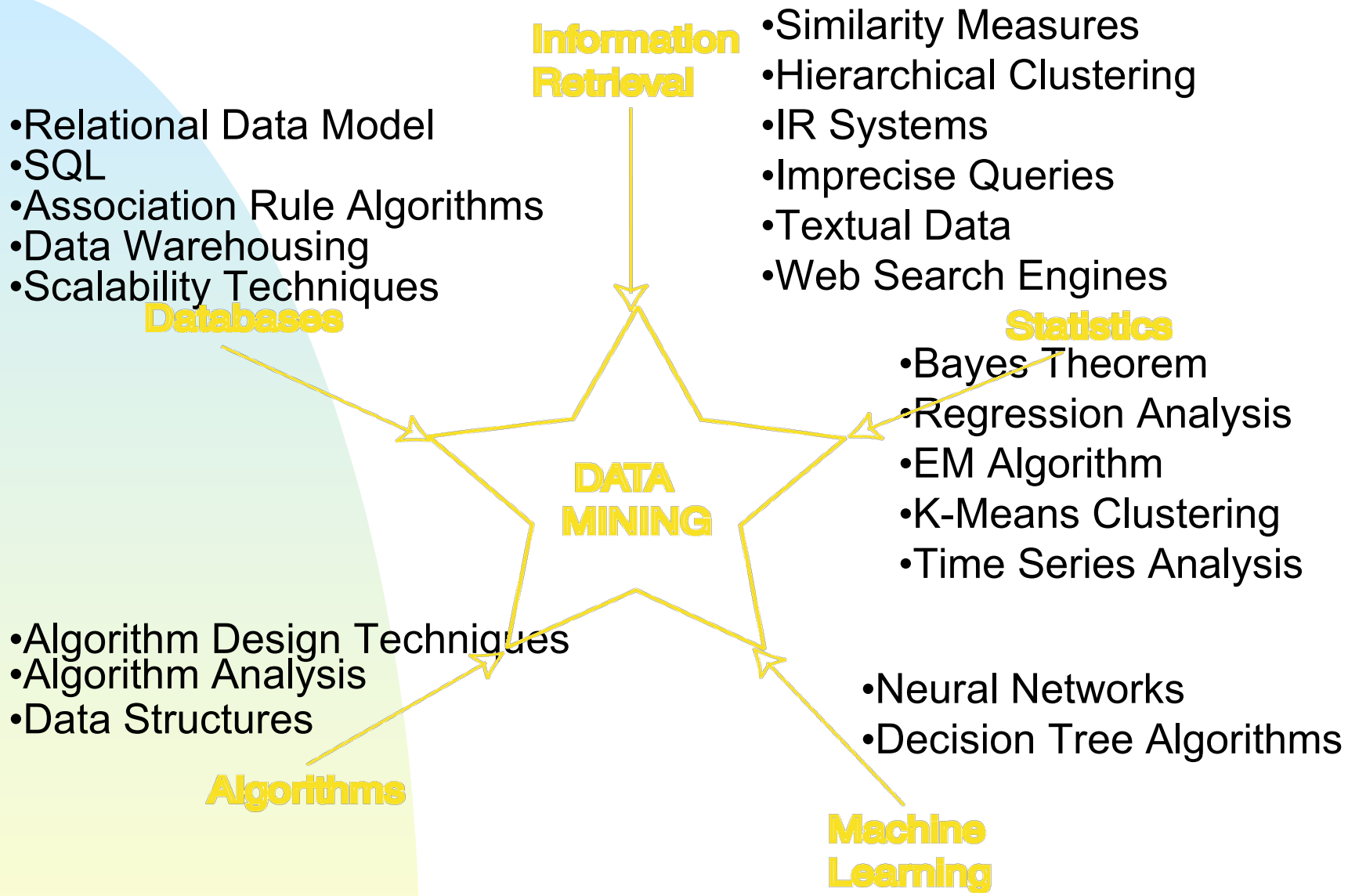
Modified from [FPSS96C]

- ***Selection:*** Obtain data from various sources.
- ***Preprocessing:*** Cleanse data.
- ***Transformation:*** Convert to common format. Transform to new format.
- ***Data Mining:*** Obtain desired results.
- ***Interpretation/Evaluation:*** Present results to user in meaningful manner.

KDD Process Ex: Web Log

- ***Selection:***
  - Select log data (dates and locations) to use
- ***Preprocessing:***
  - Remove identifying URLs
  - Remove error logs
- ***Transformation:***
  - Sessionize (sort and group)
- ***Data Mining:***
  - Identify and count patterns
  - Construct data structure
- ***Interpretation/Evaluation:***
  - Identify and display frequently accessed sequences.
- ***Potential User Applications:***
  - Cache prediction
  - Personalization

Data Mining Development



KDD Issues

- **Human Interaction**
- **Overfitting**
- **Outliers**
- **Interpretation**
- **Visualization**
- **Large Datasets**
- **High Dimensionality**

KDD Issues (cont'd)

- **Multimedia Data**
- **Missing Data**
- **Irrelevant Data**
- **Noisy Data**
- **Changing Data**
- **Integration**
- **Application**

Social Implications of DM

- Privacy
- Profiling
- Unauthorized use

Data Mining Metrics

- Usefulness
- Return on Investment (ROI)
- Accuracy
- Space/Time

Database Perspective on Data Mining

- Scalability
- Real World Data
- Updates
- Ease of Use

Visualization Techniques

- Graphical
- Geometric
- Icon-based
- Pixel-based
- Hierarchical
- Hybrid

Related Concepts Outline

Goal: Examine some areas which are related to data mining.

- Database/OLTP Systems
- Fuzzy Sets and Logic
- Information Retrieval(Web Search Engines)
- Dimensional Modeling
- Data Warehousing
- OLAP/DSS
- Statistics
- Machine Learning
- Pattern Matching

DB & OLTP Systems

- Schema
  - (ID,Name,Address,Salary,JobNo)
- Data Model
  - ER
  - Relational
- Transaction
- Query:

```
SELECT Name
FROM T
WHERE Salary > 100000
```

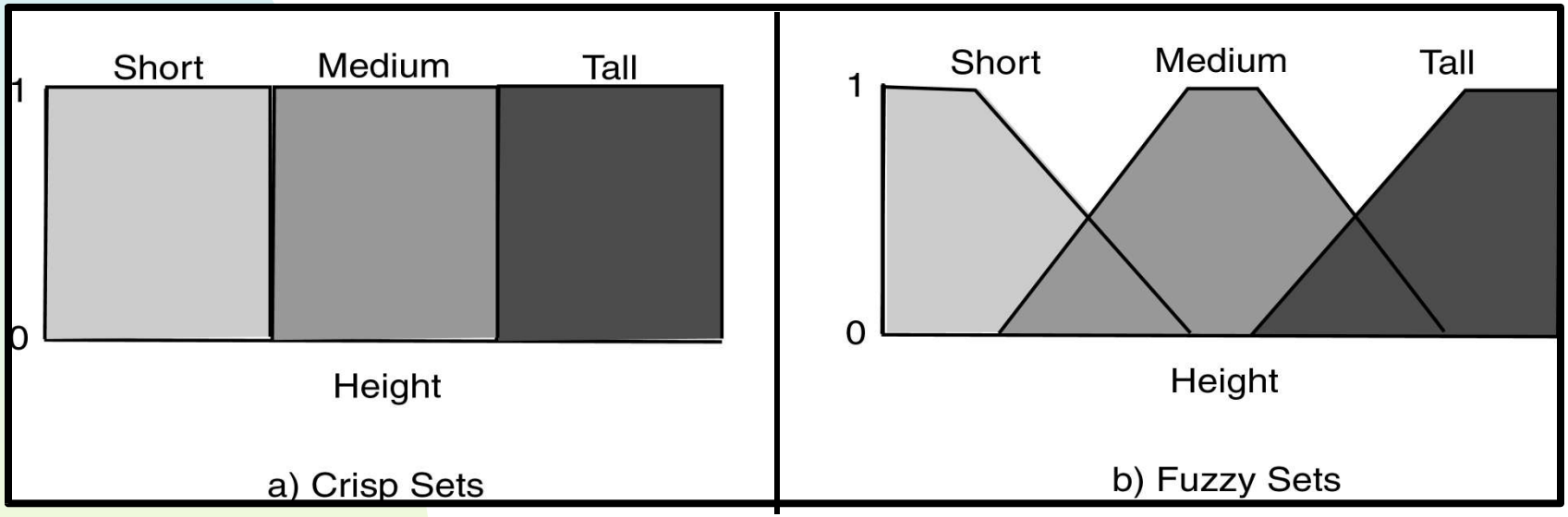
DM: Only imprecise queries

Fuzzy Sets and Logic

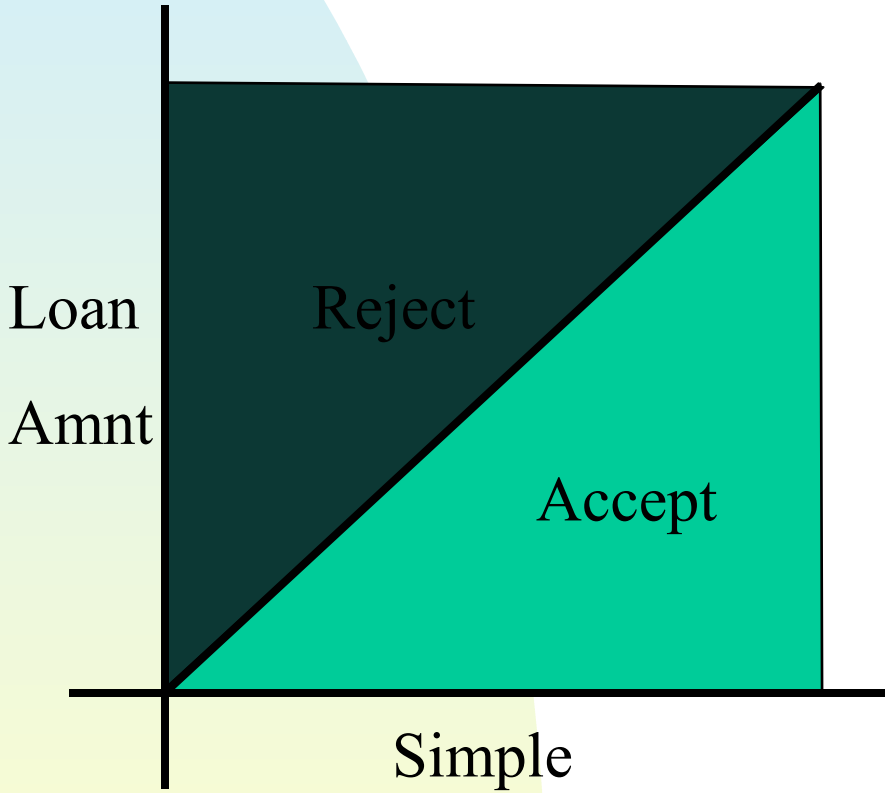
- ***Fuzzy Set:*** Set membership function is a real valued function with output in the range [0,1].
- f(x): Probability x is in F.
- 1-f(x): Probability x is not in F.
- EX:
  - T = {x | x is a person and x is tall}
  - Let f(x) be the probability that x is tall
  - Here f is the membership function

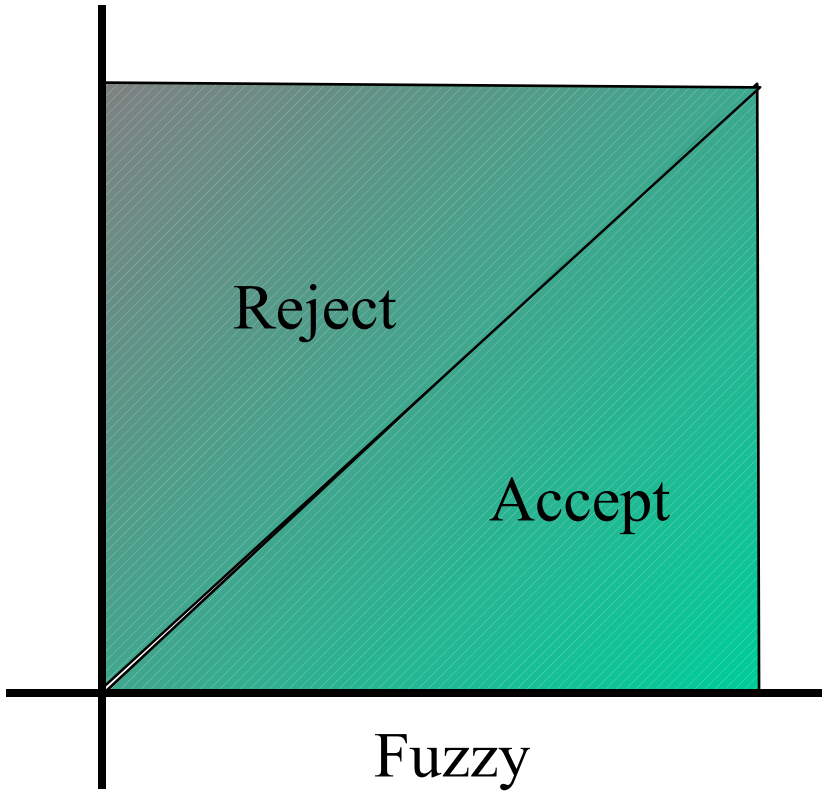
DM: Prediction and classification are fuzzy.

Fuzzy Sets


Short
Medium
Tall
1
0
Height
a) Crisp Sets
Short
Medium
Tall
1
0
Height
b) Fuzzy Sets

Classification/Prediction is Fuzzy





Information Retrieval

- ***Information Retrieval (IR):*** retrieving desired information from textual data.
- Library Science
- Digital Libraries
- Web Search Engines
- Traditionally keyword based
- Sample query:

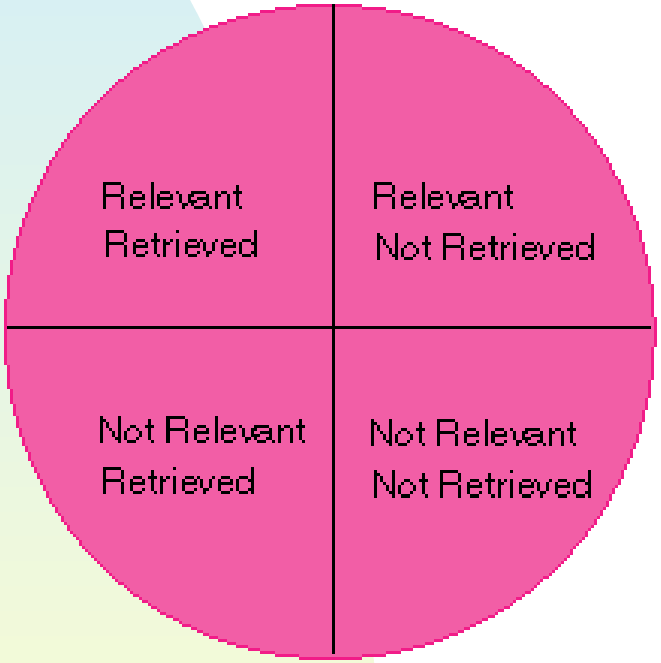
  Find all documents about “data mining”.

DM: Similarity measures;
Mine text/Web data.

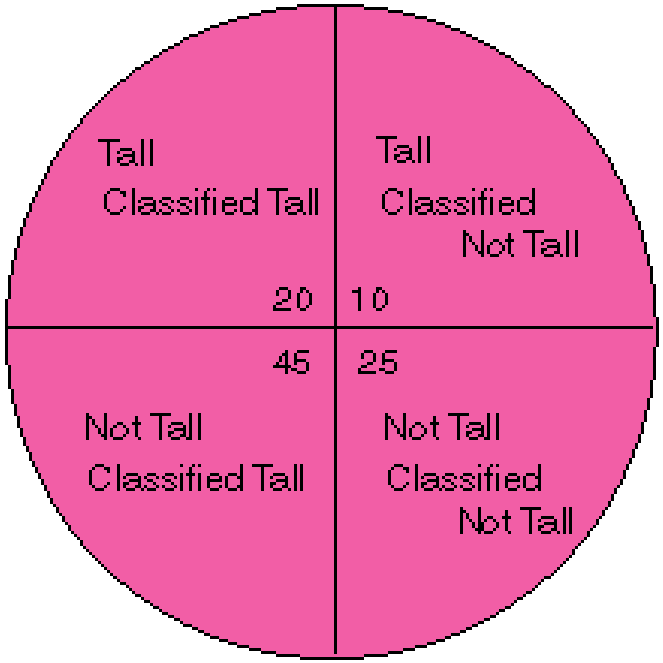
Information Retrieval (cont'd)

- ***Similarity:*** measure of how close a query is to a document.
- Documents which are "close enough" are retrieved.
- Metrics:
  - ***Precision*** = $\frac{|\text{Relevant and Retrieved}|}{|\text{Retrieved}|}$
  - ***Recall*** = $\frac{|\text{Relevant and Retrieved}|}{|\text{Relevant}|}$

IR Query Result Measures and Classification



IR



Classification

Dimensional Modeling

- View data in a hierarchical manner more as business executives might
- Useful in decision support systems and mining
- ***Dimension:*** collection of logically related attributes; axis for modeling data.
- ***Facts:*** data stored
- Ex: Dimensions – products, locations, date

  Facts – quantity, unit price

DM: May view data as dimensional.

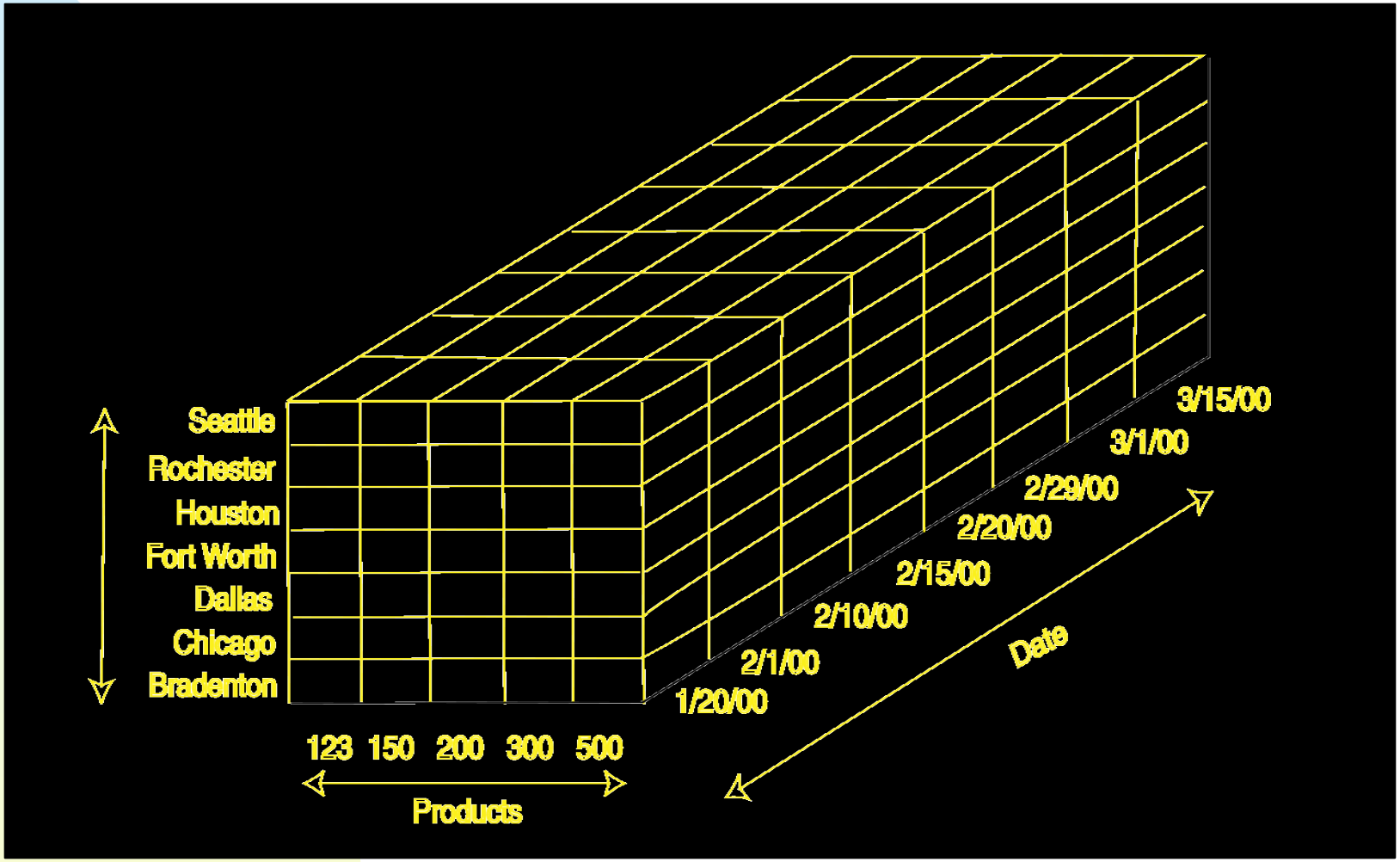
Relational View of Data

ProdID	LocID	Date	Quantity	UnitPrice
123	Dallas	022900	5	25
123	Houston	020100	10	20
150	Dallas	031500	1	100
150	Dallas	031500	5	95
150	Fort Worth	021000	5	80
150	Chicago	012000	20	75
200	Seattle	030100	5	50
300	Rochester	021500	200	5
500	Bradenton	022000	15	20
500	Chicago	012000	10	25

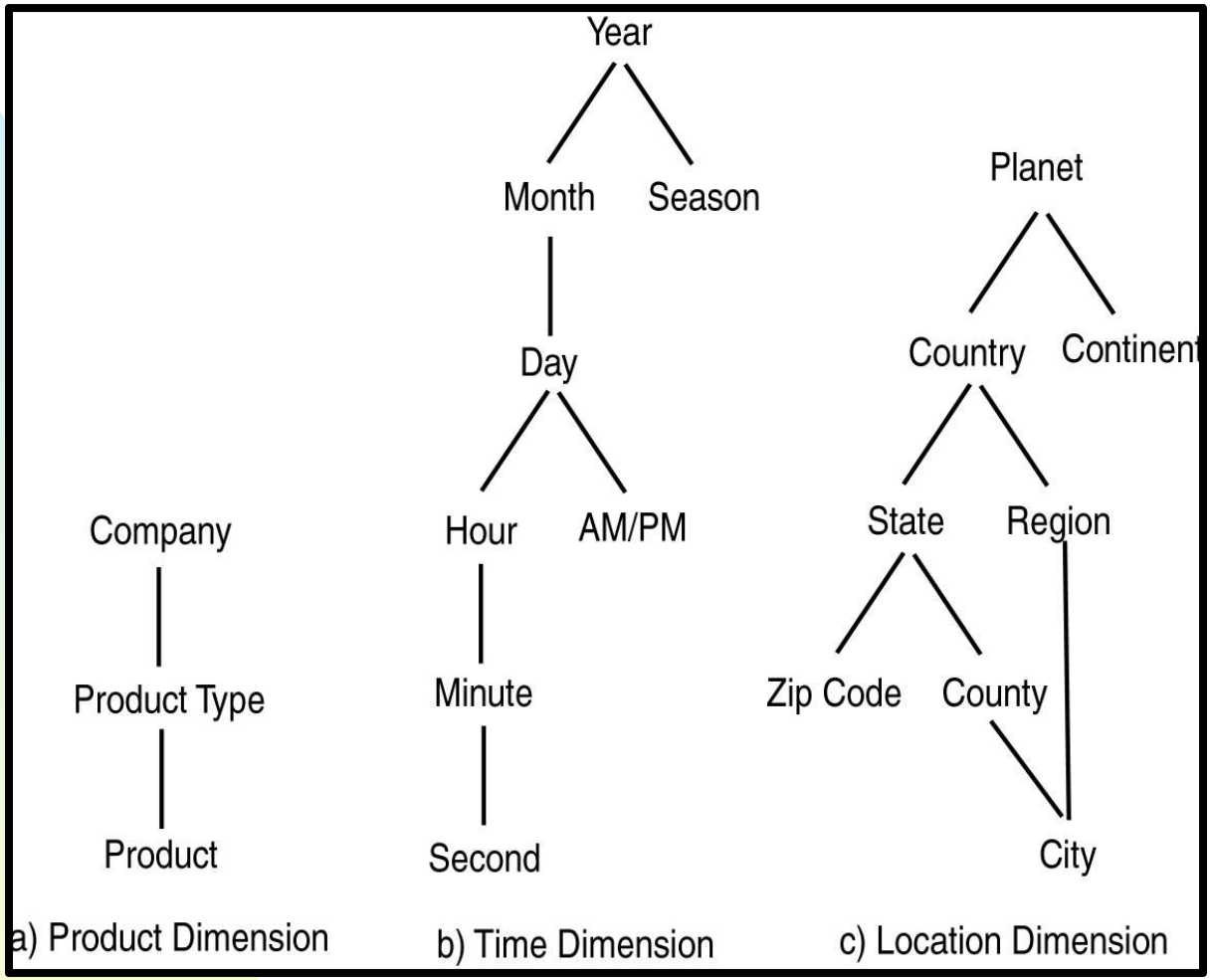
Dimensional Modeling Queries

- ***Roll Up:*** more general dimension
- ***Drill Down:*** more specific dimension
- Dimension (Aggregation) Hierarchy
- SQL uses aggregation
- ***Decision Support Systems (DSS):*** Computer systems and tools to assist managers in making decisions and solving problems.

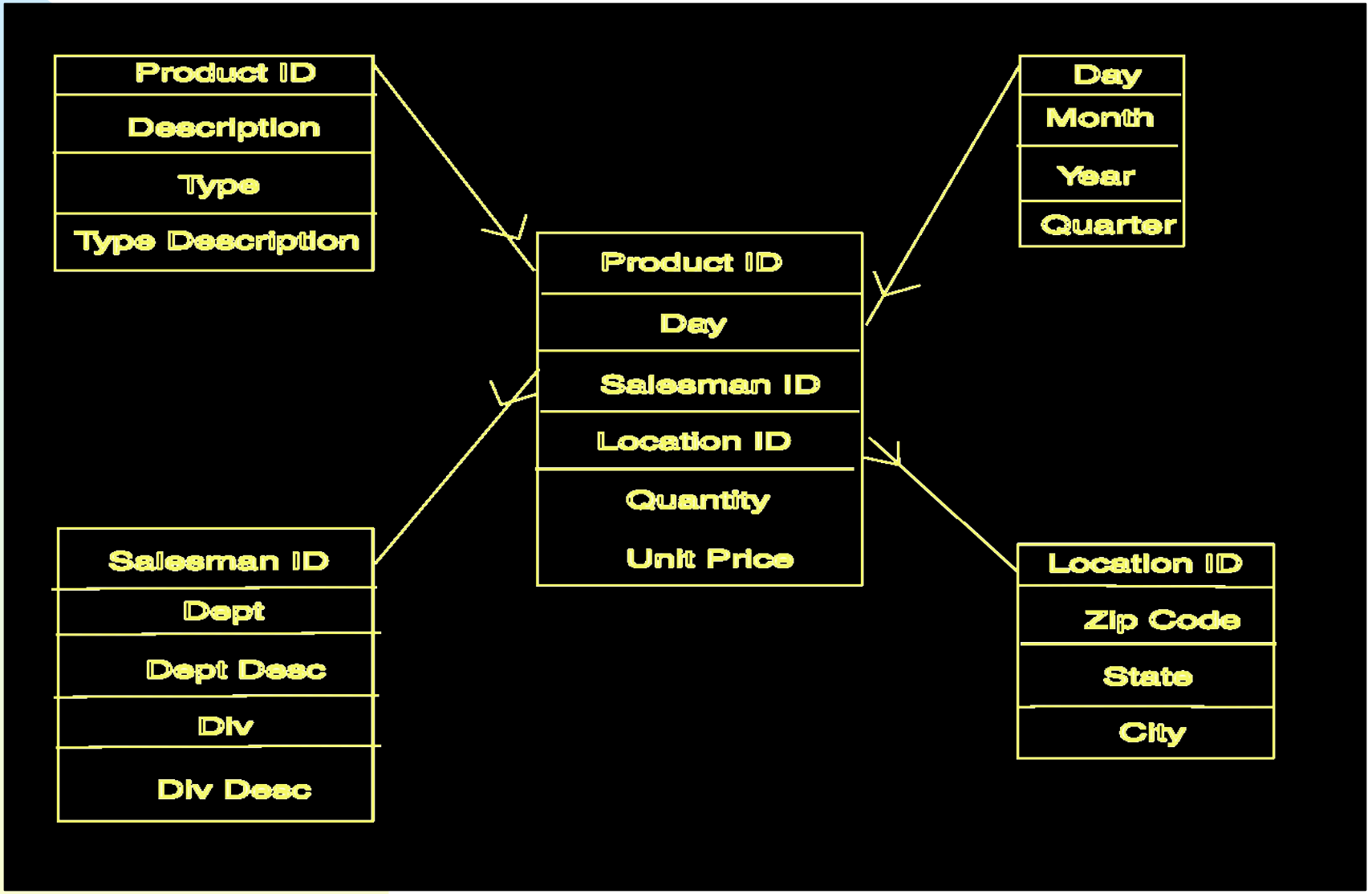
Cube view of Data



Aggregation Hierarchies



Star Schema


Product ID
Description
Type
Type Description
Day
Month
Year
Quarter
Product ID
Day
Salesman ID
Location ID
Quantity
Unit Price
Salesman ID
Dept
Dept Desc
Div
Div Desc
Location ID
Zip Code
State
City

Data Warehousing

- “Subject-oriented, integrated, time-variant, nonvolatile” William Inmon
- ***Operational Data:*** Data used in day to day needs of company.
- ***Informational Data:*** Supports other functions such as planning and forecasting.
- Data mining tools often access data warehouses rather than operational data.

DM: May access data in warehouse.

Operational vs. Informational

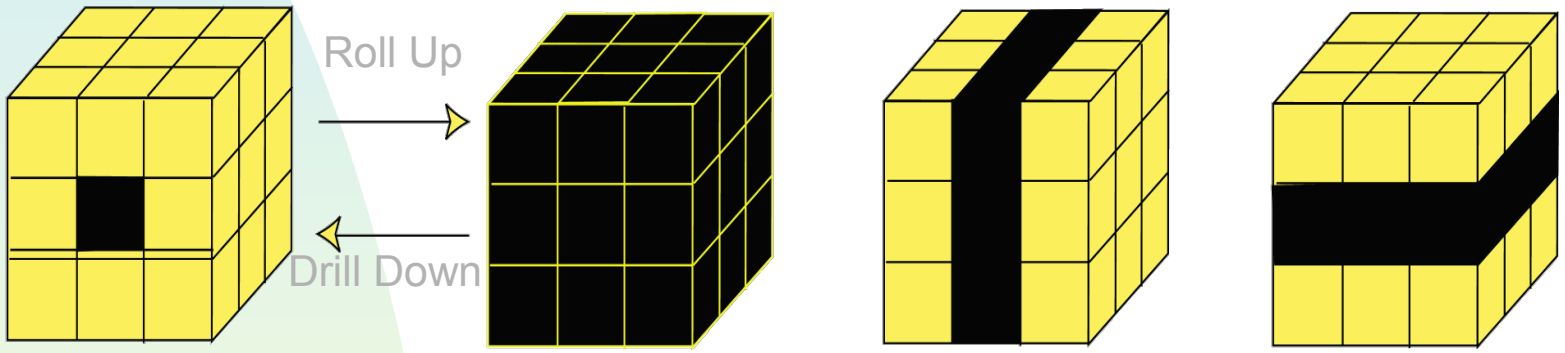
	Operational Data	Data Warehouse
Application	OLTP	OLAP
Use	Precise Queries	Ad Hoc
Temporal	Snapshot	Historical
Modification	Dynamic	Static
Orientation	Application	Business
Data	Operational Values	Integrated
Size	Gigabits	Terabits
Level	Detailed	Summarized
Access	Often	Less Often
Response	Few Seconds	Minutes
Data Schema	Relational	Star/Snowflake

OLAP

- ***Online Analytic Processing (OLAP):*** provides more complex queries than OLTP.
- ***OnLine Transaction Processing (OLTP):*** traditional database/transaction processing.
- Dimensional data; cube view
- Visualization of operations:
  - ***Slice:*** examine sub-cube.
  - ***Dice:*** rotate cube to look at another dimension.
  - ***Roll Up/Drill Down***

DM: May use OLAP queries.

OLAP Operations



Single Cell Multiple Cells Slice Dice

Statistics

- Simple descriptive models
- ***Statistical inference:*** generalizing a model created from a sample of the data to the entire dataset.
- ***Exploratory Data Analysis:***
  - Data can actually drive the creation of the model
  - Opposite of traditional statistical view.
- Data mining targeted to business user

DM: Many data mining methods come from statistical techniques.

Machine Learning

- ***Machine Learning:*** area of AI that examines how to write programs that can learn.
- Often used in classification and prediction
- ***Supervised Learning:*** learns by example.
- ***Unsupervised Learning:*** learns without knowledge of correct answers.
- Machine learning often deals with small static datasets.

DM: Uses many machine learning techniques.

Pattern Matching (Recognition)

- ***Pattern Matching:*** finds occurrences of a predefined pattern in the data.
- Applications include speech recognition, information retrieval, time series analysis.

DM: Type of classification.

DM vs. Related Topics

Area	Query	Data	Results	Output
DB/OLTP	Precise	Database	Precise	DB Objects or Aggregation
IR	Precise	Documents	Vague	Documents
OLAP	Analysis	Multidimensional	Precise	DB Objects or Aggregation
DM	Vague	Preprocessed	Vague	KDD Objects

Data Mining Techniques Outline

Goal: Provide an overview of basic data mining techniques

- Statistical
  - Point Estimation
  - Models Based on Summarization
  - Bayes Theorem
  - Hypothesis Testing
  - Regression and Correlation
- Similarity Measures
- Decision Trees
- Neural Networks
  - Activation Functions
- Genetic Algorithms

Point Estimation

- ***Point Estimate:*** estimate a population parameter.
- May be made by calculating the parameter for a sample.
- May be used to predict value for missing data.
- Ex:
  - R contains 100 employees
  - 99 have salary information
  - Mean salary of these is $50,000
  - Use $50,000 as value of remaining employee's salary.

Is this a good idea?

Estimation Error

- ***Bias:*** Difference between expected value and actual value.

$$Bias = E(\hat{\Theta}) - \Theta$$

- ***Mean Squared Error (MSE):*** expected value of the squared difference between the estimate and the actual value:

$$MSE(\hat{\Theta}) = E(\hat{\Theta} - \Theta)^2$$

- Why square?
- Root Mean Square Error (RMSE)

Jackknife Estimate

- ***Jackknife Estimate:*** estimate of parameter is obtained by omitting one value from the set of observed values.
- Ex: estimate of mean for $X=\{x_1, \ldots, x_n\}$

$$\hat{\mu}_{(i)} = \frac{\sum_{j=1}^{i-1} x_j + \sum_{j=i+1}^{n} x_j}{n-1}$$

$$\hat{\theta}_{(.)} = \frac{\sum_{j=1}^{n} \hat{\theta}_{(j)}}{n}$$

Maximum Likelihood Estimate (MLE)

- Obtain parameter estimates that maximize the probability that the sample data occurs for the specific model.
- Joint probability for observing the sample data by multiplying the individual probabilities. Likelihood function:

$$L(\Theta \mid x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i \mid \Theta)$$

- Maximize L.

MLE Example

- Coin toss five times: {H,H,H,H,T}
- Assuming a perfect coin with H and T equally likely, the likelihood of this sequence is:

$$L(p \mid 1,1,1,1,0) = \prod_{i=1}^{5} 0.5 = 0.03.$$

- However if the probability of a H is 0.8 then:

$$L(p \mid 1,1,1,1,0) = 0.8 \times 0.8 \times 0.8 \times 0.8 \times 0.2 = 0.08.$$

Expectation-Maximization (EM)

- Solves estimation with incomplete data.
- Obtain initial estimates for parameters.
- Iteratively use estimates for missing data and continue until convergence.

EM Example

$\{1, 5, 10, 4\}$; $n = 6$ $k = 4$; **Guess** $\hat{\mu}^0 = 3$.

$$\hat{\mu}^1 = \frac{\sum_{i=1}^{k} x_i}{n} + \frac{\sum_{i=k+1}^{n} x_i}{n} = 3.33 + \frac{3+3}{6} = 4.33$$

$$\hat{\mu}^2 = \frac{\sum_{i=1}^{k} x_i}{n} + \frac{\sum_{i=k+1}^{n} x_i}{n} = 3.33 + \frac{4.33+4.33}{6} = 4.77$$

$$\hat{\mu}^3 = \frac{\sum_{i=1}^{k} x_i}{n} + \frac{\sum_{i=k+1}^{n} x_i}{n} = 3.33 + \frac{4.77+4.77}{6} = 4.92$$

$$\hat{\mu}^4 = \frac{\sum_{i=1}^{k} x_i}{n} + \frac{\sum_{i=k+1}^{n} x_i}{n} = 3.33 + \frac{4.92+4.92}{6} = 4.97$$

EM Algorithm

Input:

$\Theta = \{\theta_1, ..., \theta_p\}$ //Parameters to be Estimated

$X_{obs} = \{x_1, ..., x_k\}$ //Input Database Values Observed

$X_{miss} = \{x_{k+1}, ..., x_n\}$ //Input Database Values Missing

Output:

$\hat{\Theta}$ //Estimates for Θ

EM Algorithm:

i := 0;

Obtain initial parameter MLE estimate,$\hat{\Theta}^i$;

repeat

Estimate missing data,$\hat{X}^i_{miss}$;

i++;

Obtain next parameter estimate,$\hat{\theta}^i$ to maximize data;

until estimate converges;

Bayes Theorem

- ***Posterior Probability:*** $P(h_1|x_i)$
- ***Prior Probability:*** $P(h_1)$
- ***Bayes Theorem:***

$$P(x_i) = \sum_{j=1}^{m} P(x_i \mid h_j)\ P(h_j).$$

$$P(h_1 \mid x_i) = \frac{P(x_i \mid h_1)\ P(h_1)}{P(x_i)}.$$

- Assign probabilities of hypotheses given a data value.

Bayes Theorem Example

- Credit authorizations (hypotheses): h_1=authorize purchase, h_2 = authorize after further identification, h_3=do not authorize, h_4= do not authorize but contact police
- Assign twelve data values for all combinations of credit and income:

	1	2	3	4
Excellent	x_1	x_2	x_3	x_4
Good	x_5	x_6	x_7	x_8
Bad	x_9	x_{10}	x_{11}	x_{12}

- From training data: $P(h_1) = 60\%$; $P(h_2)=20\%$; $P(h_3) =10\%$; $P(h_4)=10\%$.

Bayes Example(cont'd)

- Training Data:

ID	Income	Credit	Class	x_i
1	4	Excellent	h_1	x_4
2	3	Good	h_1	x_7
3	2	Excellent	h_1	x_2
4	3	Good	h_1	x_7
5	4	Good	h_1	x_8
6	2	Excellent	h_1	x_2
7	3	Bad	h_2	x_{11}
8	2	Bad	h_2	x_{10}
9	3	Bad	h_3	x_{11}
10	1	Bad	h_4	x_9

Bayes Example(cont'd)

- Calculate $P(x_i|h_j)$ and $P(x_i)$
- Ex: $P(x_7|h_1)=2/6$; $P(x_4|h_1)=1/6$; $P(x_2|h_1)=2/6$; $P(x_8|h_1)=1/6$; $P(x_i|h_1)=0$ for all other x_i.
- Predict the class for x_4:
  - Calculate $P(h_j|x_4)$ for all h_j.
  - Place x_4 in class with largest value.
  - Ex:
    - $P(h_1|x_4)=(P(x_4|h_1)(P(h_1))/P(x_4)$
      $=(1/6)(0.6)/0.1=1.$
    - x_4 in class h_1.

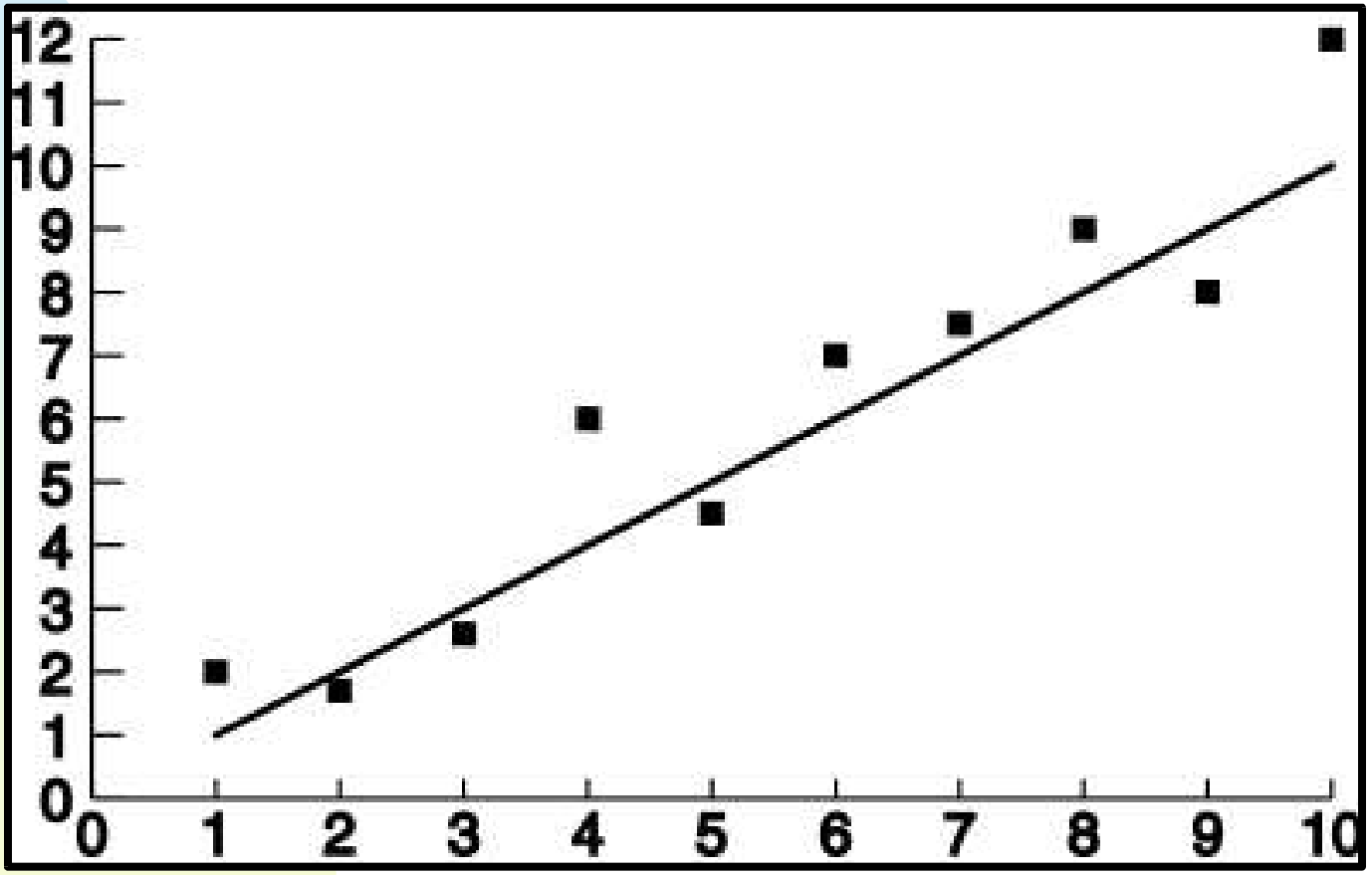
Regression

- Predict future values based on past values
- ***Linear Regression*** assumes linear relationship exists.

$$y = c_0 + c_1 x_1 + \ldots + c_n x_n$$

- Find values to best fit the data

Linear Regression



Correlation

- Examine the degree to which the values for two variables behave similarly.
- Correlation coefficient r:
  - 1 = perfect correlation
  - -1 = perfect but opposite correlation
  - 0 = no correlation

$$r = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum (x_i - \bar{X})^2 \sum (y_i - \bar{Y})^2}}$$

Similarity Measures

- Determine similarity between two objects.
- Similarity characteristics:
  - $\forall t_i \in D, sim(t_i, t_i) = 1$
  - $\forall t_i, t_j \in D, sim(t_i, t_j) = 0$ if t_i and t_j are not alike at all.
  - $\forall t_i, t_j, t_k \in D, sim(t_i, t_j) < sim(t_i, t_k)$ if t_i is more like t_k than it is like t_j.
- Alternatively, distance measure measure how unlike or dissimilar objects are.

Similarity Measures

Dice: $$sim(t_i, t_j) = \frac{2\sum_{h=1}^{k} t_{ih} t_{jh}}{\sum_{h=1}^{k} t_{ih}^2 + \sum_{h=1}^{k} t_{jh}^2}$$

Jaccard: $$sim(t_i, t_j) = \frac{\sum_{h=1}^{k} t_{ih} t_{jh}}{\sum_{h=1}^{k} t_{ih}^2 + \sum_{h=1}^{k} t_{jh}^2 - \sum_{h=1}^{k} t_{ih} t_{jh}}$$

Cosine: $$sim(t_i, t_j) = \frac{\sum_{h=1}^{k} t_{ih} t_{jh}}{\sqrt{\sum_{h=1}^{k} t_{ih}^2 \sum_{h=1}^{k} t_{jh}^2}}$$

Overlap: $$sim(t_i, t_j) = \frac{\sum_{h=1}^{k} t_{ih} t_{jh}}{min(\sum_{h=1}^{k} t_{ih}^2, \sum_{h=1}^{k} t_{jh}^2)}$$

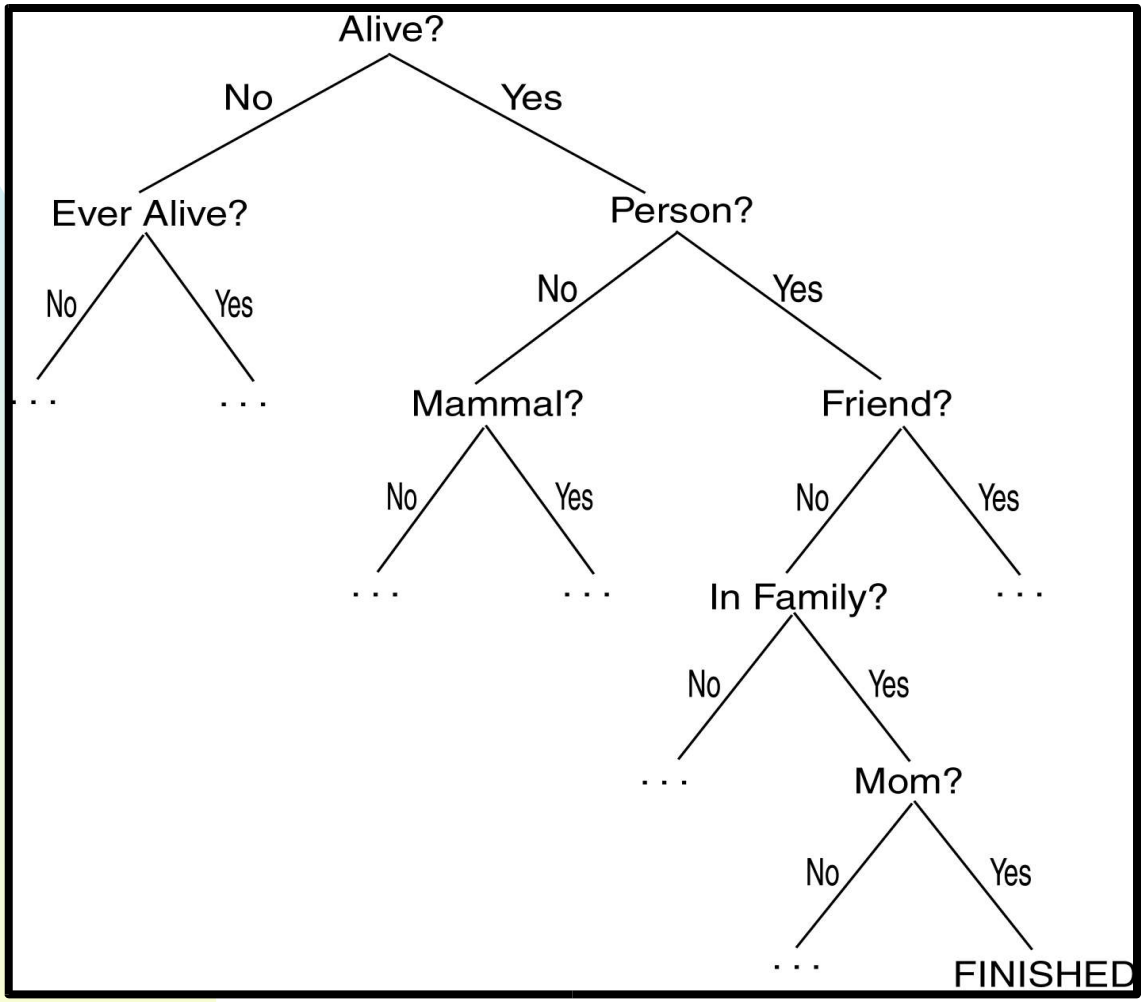
Distance Measures

- Measure dissimilarity between objects

Euclidean: $dis(t_i, t_j) = \sqrt{\sum_{h=1}^{k}(t_{ih} - t_{jh})^2}$

Manhattan: $dis(t_i, t_j) = \sum_{h=1}^{k} \mid (t_{ih} - t_{jh}) \mid$

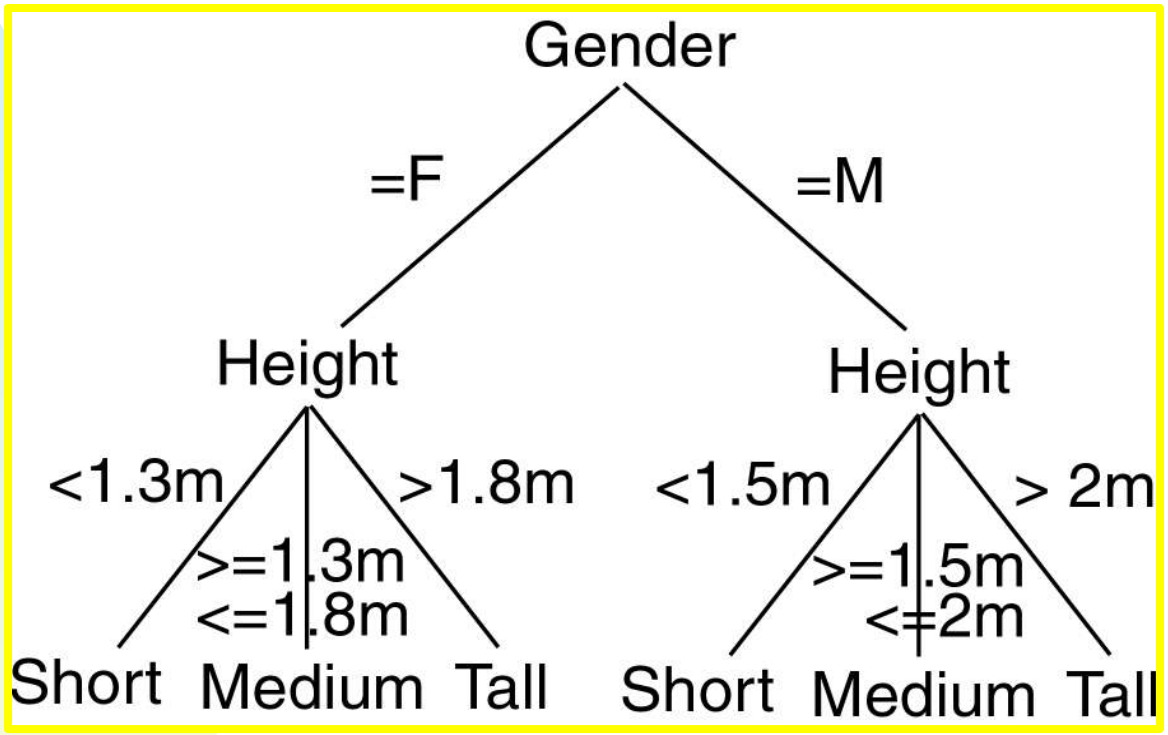
Twenty Questions Game



Decision Trees

- ***Decision Tree (DT):***
  - Tree where the root and each internal node is labeled with a question.
  - The arcs represent each possible answer to the associated question.
  - Each leaf node represents a prediction of a solution to the problem.
- Popular technique for classification; Leaf node indicates class to which the corresponding tuple belongs.

Decision Tree Example



Decision Trees

- A ***Decision Tree Model*** is a computational model consisting of three parts:
  - Decision Tree
  - Algorithm to create the tree
  - Algorithm that applies the tree to data
- Creation of the tree is the most difficult part.
- Processing is basically a search similar to that in a binary search tree (although DT may not be binary).

Decision Tree Algorithm

```
Input:
   T      //Decision Tree
   D      //Input Database
Output:
   M      //Model Prediction
DTProc Algorithm:
                //Illustrates Prediction Technique using DT
   for each t ∈ D do
      n = root node of T;
      while n not leaf node do
         Obtain answer to question on n applied t;
         Identify arc from t which contains correct answer;
         n = node at end of this arc;
      Make prediction for t based on labeling of n;
```

DT Advantages/Disadvantages

- Advantages:
  - Easy to understand.
  - Easy to generate rules
- Disadvantages:
  - May suffer from overfitting.
  - Classifies by rectangular partitioning.
  - Does not easily handle nonnumeric data.
  - Can be quite large – pruning is necessary.

Neural Networks

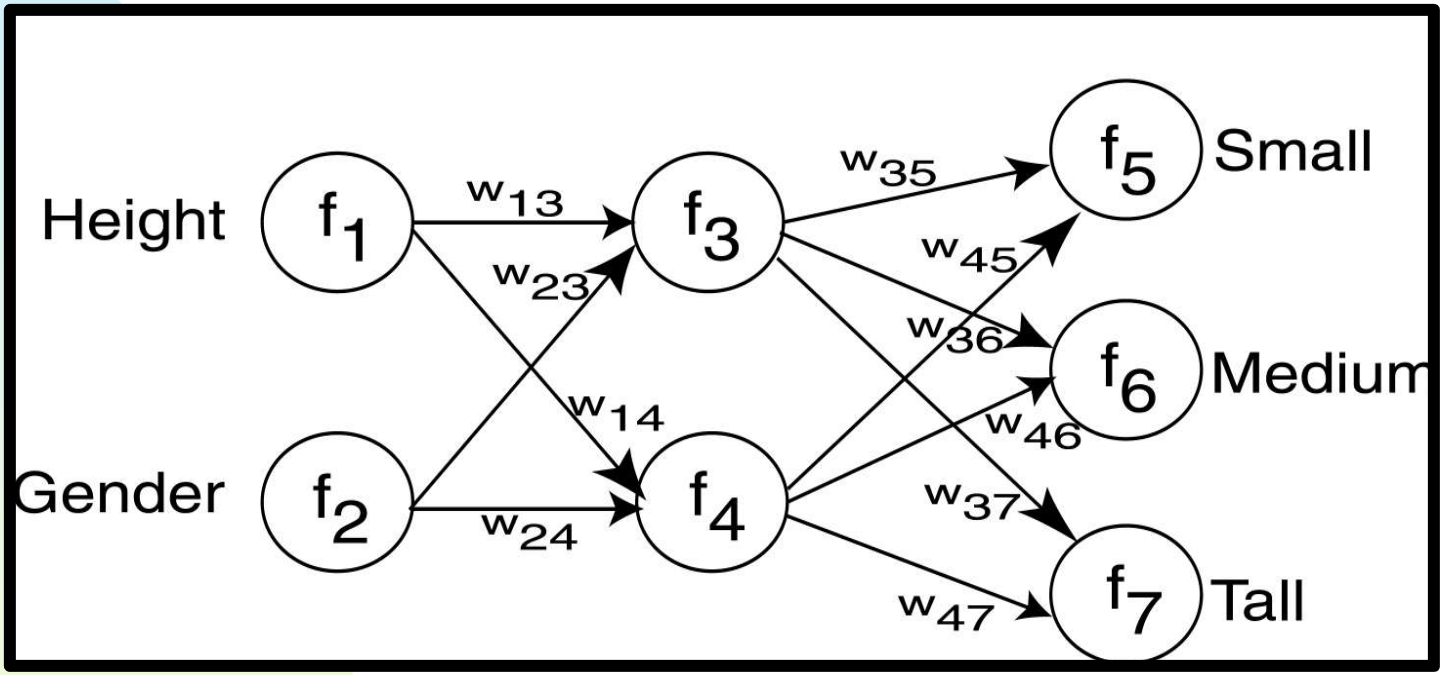
- Based on observed functioning of human brain.

- ***(Artificial Neural Networks (ANN)***
- Our view of neural networks is very simplistic.
- We view a neural network (NN) from a graphical viewpoint.
- Alternatively, a NN may be viewed from the perspective of matrices.
- Used in pattern recognition, speech recognition, computer vision, and classification.

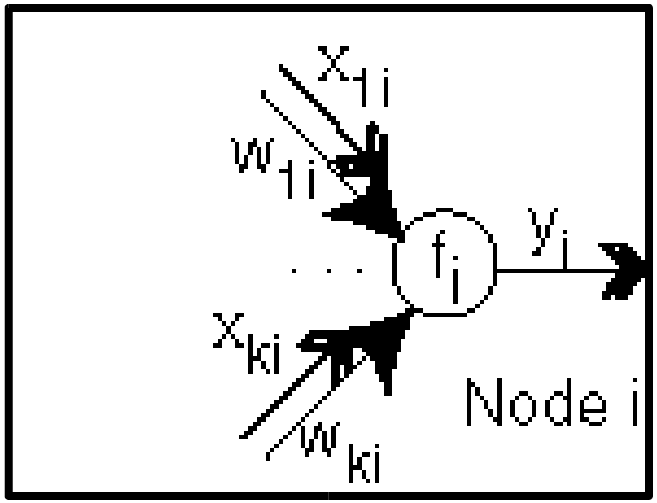
Neural Networks

- ***Neural Network (NN)*** is a directed graph F=<V,A> with vertices V={1,2,…,n} and arcs A={<i,j>| 1<=i,j<=n}, with the following restrictions:
  - V is partitioned into a set of input nodes, V_I, hidden nodes, V_H, and output nodes, V_O.
  - The vertices are also partitioned into layers
  - Any arc <i,j> must have node i in layer h-1 and node j in layer h.
  - Arc <i,j> is labeled with a numeric value w_{ij}.
  - Node i is labeled with a function f_i.

Neural Network Example



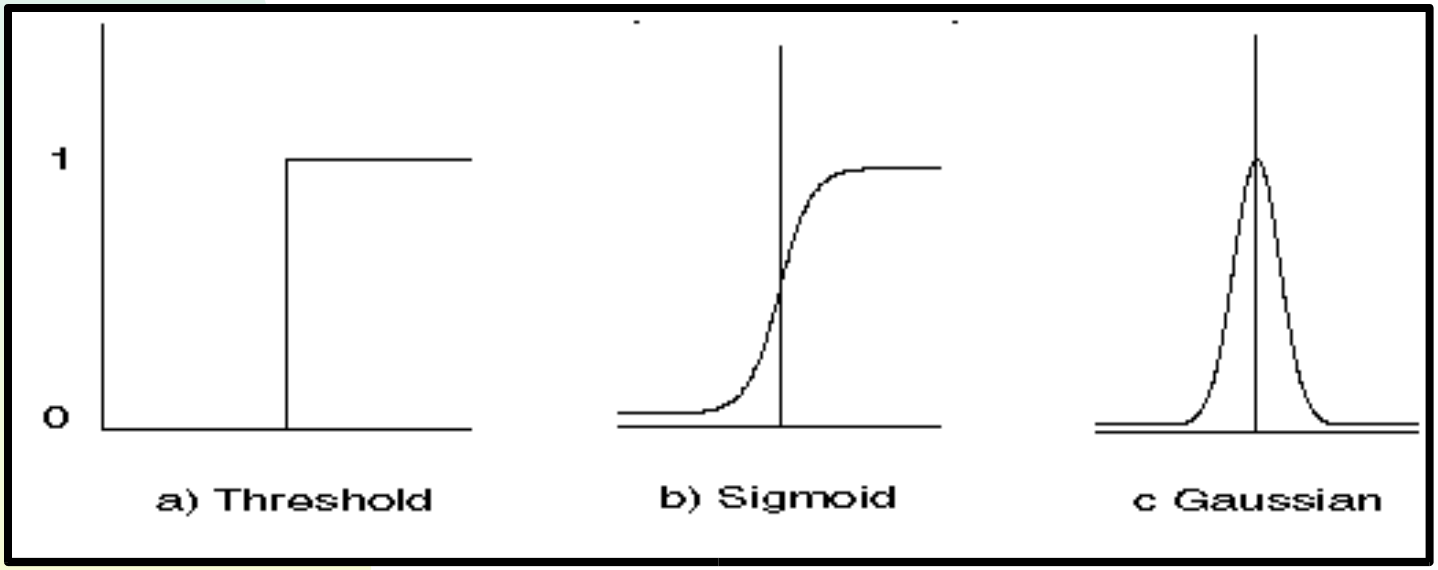
NN Node



$$y_i = f_i(\sum_{j=1}^{k} w_{ji}\ x_{ji}) = f_i([w_{1i}...w_{ki}] \begin{bmatrix} x_{1i} \\ ... \\ x_{ki} \end{bmatrix})$$

NN Activation Functions

- Functions associated with nodes in graph.
- Output may be in range [-1,1] or [0,1]



NN Activation Functions

Linear:

$$f_i(S) = c\,S$$

Threshold or Step:

$$f_i(S) = \left\{ \begin{array}{ll} 1 & if S > T \\ 0 & otherwise \end{array} \right\}$$

Ramp:

$$f_i(S) = \left\{ \begin{array}{ll} 1 & if S > T_2 \\ \frac{S-T_1}{T_2-T_1} & if T_1 \leq S \leq T_2 \\ 0 & if S < T_1 \end{array} \right\}$$

Sigmoid:

$$f_i(S) = \frac{1}{(1+e^{-c\,S})}$$

Hyperbolic Tangent:

$$f_i(S) = \frac{(1-e^{-S})}{(1+e^{-c\,S})}$$

Gaussian:

$$f_i(S) = e^{\frac{-S^2}{v}}$$

NN Learning

- Propagate input values through graph.
- Compare output to desired output.
- Adjust weights in graph accordingly.

Neural Networks

- A ***Neural Network Model*** is a computational model consisting of three parts:
  - Neural Network graph
  - Learning algorithm that indicates how learning takes place.
  - Recall techniques that determine how information is obtained from the network.
- We will look at propagation as the recall technique.

NN Advantages

- Learning
- Can continue learning even after training set has been applied.
- Easy parallelization
- Solves many problems

NN Disadvantages

- Difficult to understand
- May suffer from overfitting
- Structure of graph must be determined a priori.
- Input values must be numeric.
- Verification difficult.

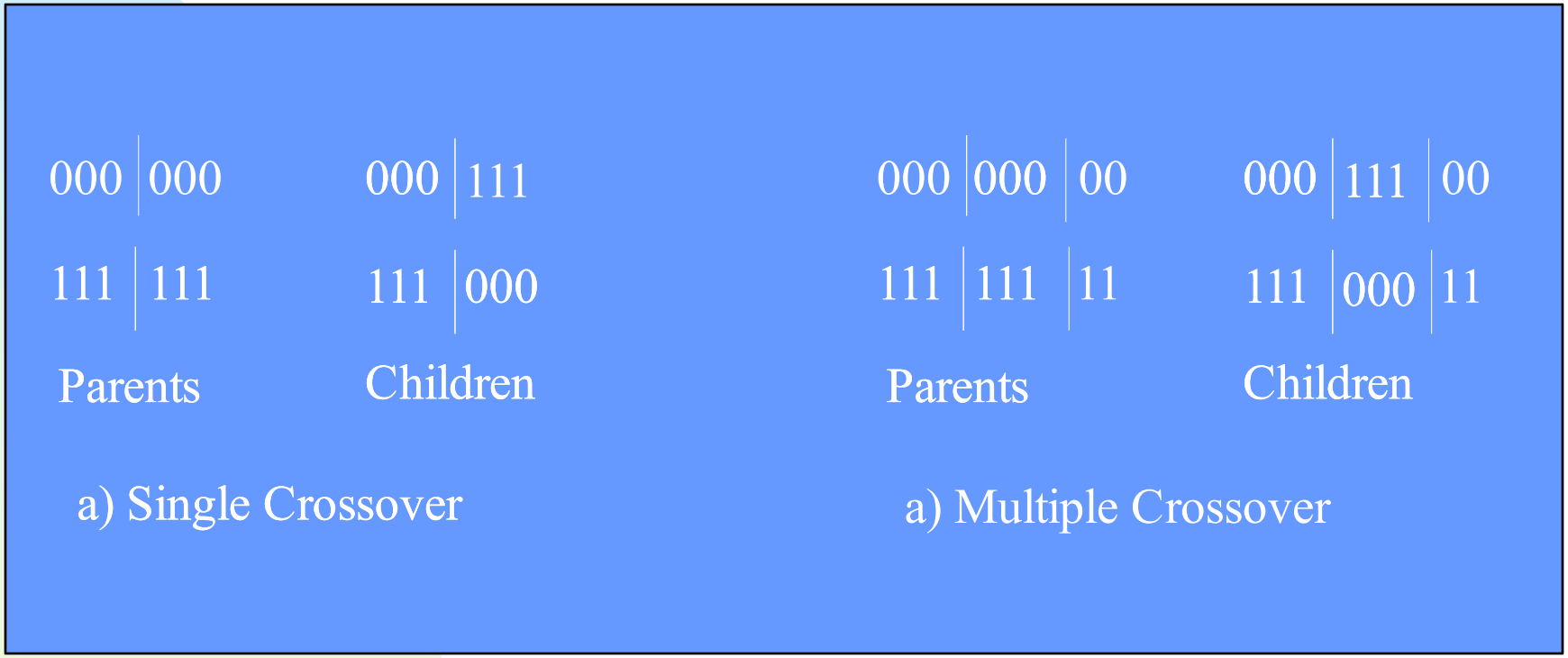
Genetic Algorithms

- Optimization search type algorithms.
- Creates an initial feasible solution and iteratively creates new “better” solutions.
- Based on human evolution and survival of the fitness.
- Must represent a solution as an individual.
- ***Individual:*** string $I=I_1,I_2,\ldots,I_n$ where I_j is in given alphabet A.
- Each character I_j is called a ***gene***.
- ***Population:*** set of individuals.

Genetic Algorithms

- A ***Genetic Algorithm (GA)*** is a computational model consisting of five parts:
  - A starting set of individuals, P.
  - ***Crossover***: technique to combine two parents to create offspring.
  - ***Mutation:*** randomly change an individual.
  - ***Fitness:*** determine the best individuals.
  - Algorithm which applies the crossover and mutation techniques to P iteratively using the fitness function to determine the best individuals in P to keep.

Crossover Examples



Genetic Algorithm

```
Input:
   P        //Initial Population
Output:
   P'       //Improved Population
Genetic Algorithm:
                //Illustrates Genetic Algorithm
   repeat
      N = | P |;
      P' = ∅;
      repeat
         i1, i2 = select(P);
         o1, o2 = cross(i1, i2);
         o1 = mutate(o1);
         o2 = mutate(o2);
         P' = P' ∪ {o1, o2};
      until | P' |= N;
      P = P';
   until termination criteria satisfied;
```

GA Advantages/Disadvantages

- Advantages
  - Easily parallelized
- Disadvantages
  - Difficult to understand and explain to end users.
  - Abstraction of the problem and method to represent individuals is quite difficult.
  - Determining fitness function is difficult.
  - Determining how to perform crossover and mutation is difficult.