



**RAMA
UNIVERSITY**

www.ramauniversity.ac.in

FACULTY OF ENGINEERING

**DATA MINING & WAREHOUSEING
LECTURE-37**

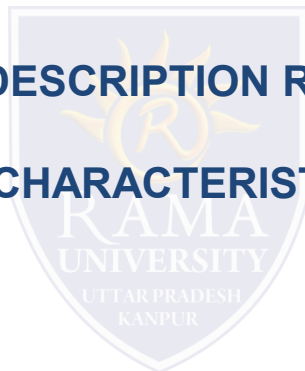
MR. DHIRENDRA

ASSISTANT PROFESSOR

RAMA UNIVERSITY

OUTLINE

- ❖ **QUANTITATIVE DISCRIMINANT RULES**
- ❖ **EXAMPLE: QUANTITATIVE DISCRIMINANT RULE**
- ❖ **CLASS DESCRIPTION**
- ❖ **EXAMPLE: QUANTITATIVE DESCRIPTION RULE**
- ❖ **MINING DATA DISPERSION CHARACTERISTICS**
- ❖ **MCQ**
- ❖ **REFERENCES**



Quantitative Discriminant Rules

- C_j = target class
- q_a = a generalized tuple covers some tuples of class
 - but can also cover some tuples of contrasting class
- d-weight
 - range: [0, 1]

$$d\text{-weight} = \frac{\text{count}(q_a \in C_j)}{\sum_{i=1}^m \text{count}(q_a \in C_i)}$$

- quantitative discriminant rule form

$$\forall X, \text{target_class}(X) \Leftarrow \text{condition}(X) \quad [d : d_weight]$$

Example: Quantitative Discriminant Rule

Status	Birth_country	Age_range	Gpa	Count
Graduate	Canada	25-30	Good	90
Undergraduate	Canada	25-30	Good	210

Count distribution between graduate and undergraduate students for a generalized tuple

- Quantitative discriminant rule

$\forall X, \text{graduate_student}(X) \Leftarrow$

$\text{birth_country}(X) = \text{"Canada"} \wedge \text{age_range}(X) = \text{"25-30"} \wedge \text{gpa}(X) = \text{"good"} \quad [d:30\%]$

– where $90/(90+210) = 30\%$

- Quantitative characteristic rule

$\forall X, target_class(X) \Rightarrow condition(X) [t : t_weight]$
– necessary

- Quantitative discriminant rule

$\forall X, target_class(X) \Leftarrow condition(X) [d : d_weight]$
– sufficient

- Quantitative description rule

$\forall X, target_class(X) \Leftrightarrow$
 $condition_1(X) [t : w_1, d : w'_1] \vee \dots \vee condition_n(X) [t : w_n, d : w'_n]$
– necessary and sufficient

Example: Quantitative Description Rule

Location/item	TV			Computer			Both_items		
	Count	t-wt	d-wt	Count	t-wt	d-wt	Count	t-wt	d-wt
Europe	80	25%	40%	240	75%	30%	320	100%	32%
N_Am	120	17.65%	60%	560	82.35%	70%	680	100%	68%
Both_regions	200	20%	100%	800	80%	100%	1000	100%	100%

Crosstab showing associated t-weight, d-weight values and total number (in thousands) of TVs and computers sold at AllElectronics in 1998

- Quantitative description rule for target class *Europe*

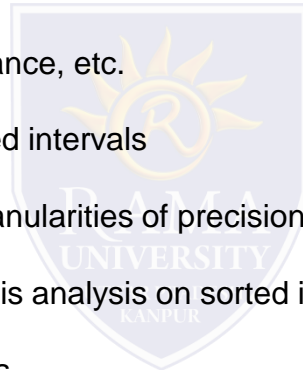
$\forall X, Europe(X) \Leftrightarrow$

$(item(X) = "TV") [t : 25\%, d : 40\%] \vee (item(X) = "computer") [t : 75\%, d : 30\%]$

Mining Data Dispersion Characteristics

Motivation

- To better understand the data: central tendency, variation and spread
- Data dispersion characteristics
 - median, max, min, quantiles, outliers, variance, etc.
- Numerical dimensions -correspond to sorted intervals
 - Data dispersion: analyzed with multiple granularities of precision
 - Boxplot or quantile analysis on sorted intervals
- Dispersion analysis on computed measures
 - Folding measures into numerical dimensions
 - Boxplot or quantile analysis on the transformed cube



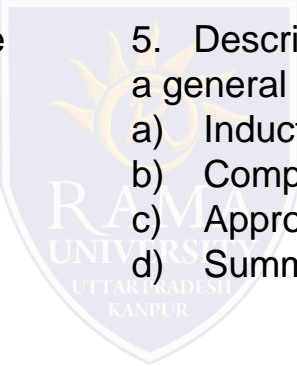
Multiple Choice Question

1. Various visualization techniques are used in _____ step of KDD.
 - a) selection
 - b) transformaion
 - c) data mining.
 - d) interpretation.

 2. Extreme values that occur infrequently are called as _____.
 - a) outliers
 - b) rare values.
 - c) dimensionality reduction.
 - d) All of the above.

 3. Box plot and scatter diagram techniques are _____.
 - a) Graphical
 - b) Geometric
 - c) Icon-based.
 - d) Pixel-based.

 4. _____ is used to proceed from very specific knowledge to more general information.
 - a) Induction
 - b) Compression.
 - c) Approximation.
 - d) Substitution.

 5. Describing some characteristics of a set of data by a general model is viewed as _____.
 - a) Induction
 - b) Compression
 - c) Approximation
 - d) Summarization
- 

REFERENCES

- https://www.tutorialspoint.com/dwh/dwh_overview.htm
- <https://www.geeksforgeeks.org/>
- <http://myweb.sabanciuniv.edu/rdekharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf> DATA MINING BOOK WRITTEN BY Micheline Kamber
- <https://www.javatpoint.com/three-tier-data-warehouse-architecture>
- M.H. Dunham, “ Data Mining: Introductory & Advanced Topics” Pearson Education
- Jiawei Han, Micheline Kamber, “ Data Mining Concepts & Techniques” Elsevier
- Sam Anahory, Dennis Murray,” data warehousing in the Real World: A Practical Guide for Building Decision Support Systems, “ Pearson Education
- Mallach,” Data Warehousing System”, TMH
- R. Agrawal, A. Gupta, and S. Sarawagi. Modeling multidimensional databases. ICDE’97 S. Chaudhuri and U. Dayal. An overview of data warehousing and OLAP technology. ACM SIGMOD Record, 26:65-74, 1997
- S. Agarwal, R. Agrawal, P. M. Deshpande, A. Gupta, J. F. Naughton, R. Ramakrishnan, and S. Sarawagi. On the computation of multidimensional aggregates. VLDB’96 D. Agrawal, A. E. Abbadi, A. Singh, and T. Yurek. Efficient view maintenance in data warehouses. SIGMOD’97
- E. F. Codd, S. B. Codd, and C. T. Salley. Beyond decision support. Computer World, 27, July 1993.
- J. Gray, et al. Data cube: A relational aggregation operator generalizing group-by, cross-tab and sub-totals. Data Mining and Knowledge Discovery, 1:29-54, 1997.