Stat-424 Data Mining and Simulation

100 Marks: 03 Credits

Number of Class: 35-40

Overview: Concept of data mining and its advantages and disadvantages, knowledge Discovery, Basices, Data Mining Tasks , Classification, Regression, Time Series Analysis Prediction Chistening, Summarization Association, Rules Sequesnce Discovery, Development of Data Mining, Data Mining Issue and Metrics, Social Implications of Data Mining.

Fuzzy Sets: Introduction Classical Set, Set operation, Boolean Logic, Basic Concepts of Fuzzy Sets, Other Representations of Fuzzy Sets , Determination of Membership Functions, Fuzzy Sets Properties, operations on Fuzzy Sets , Logic Operations Algebric operations on Fuzzy Sets.

Fuzzy Relation: Classical Relations, Classical Reasoning, Fundamentals of Fuzzy relations, Operations on Binary Fuzzy Relations Types of Fuzzy Relations, Fuzzy Reasoning, Concluding Remarks, Bibliography, Web Resources.

Database/OLTP Systems, Logic, Information Retrival, Decision Support Systems, DimensionalModeling Multidimensional Schems, Indexing Data Warehousing, OLAP, Web Search Engines Statistics, Machine

Data Mining Techniques: Statistical Perspective on Data mining , Point Estimation, Models based on Summarization, Bayes Theorem,, Hypothesis Testing , Regression and correlation, Similarity Measures, Decision Tree Genetic Algorithms.

Neural Network: Background, Learning, Basic Neuron Model, Perception, Multiplayer Perception, Recurrent Network, Hopfield Network, Boltzanann Machine Network,kohonen Self Organizing network, Background, Description, Determining the Winning Neuron, Learning Algorithm.

Classification: Issues in Classification , Statistical- based algorithm , regression , Bayesian Classification, Distance- based Algorithm, Propagation NN, Supervised learning, Radial basis Function network, Perceptions, Rule Based Algorithms, Generating Rules form DT, Generating Rules for Neural Net.

Association Rules: Meaning of Association, Large Item Sets, Basic Algorithms, Apriori algorithm, Sampling Algorithm, Partitioning, Parallel and Distributed Algorithms, Data Parallelism, Task Parallelism, Advanced Association Rule / Technique, Quantitative Association Rules, Correlation Rules, Measuring quality of rules.

Simulation: Concept of Simulation, General Techniques for Simulating Continuous Random Variables, Inverse Transformation Method, Rejection Method, Hazard Rate Method, Special Techniques for Simulating Continuous Random Variables, Simulating from Discrete Distribution, Alias Method, Variance Reduction Techniques.

Text

1. Dunham, N. H. (2003): *Data Mining*, Pearson Education, Asia.

References

1. Ibrahim, A.M. (2004): *Fuzzy Logic for Embedded Systems Applications*, Elsevier Science, USA.
2. Larose, D.T. (2006): *Data Mining Methods and Models*, Wiley-Interscience, India.
3. Schalkoff, R. (2005): *Pattern Recognition Statistical Structural and Neural Approaches*, John Wiley and Sons, New York.