Techno India NJR Institute of Technology

STATISTICAL APPLICATIONS IN DATA MINING

Training Module

Total Time: 1 Month

1. Fundamentals of Data Mining

- Mathematical representation of data
- Support count and confidence
- Seasonal and cyclic variations
- Data cascading
- Growth models
- Knowledge representation models
- Statistical curve fitting

2. Analysis

- Statistical hypothesis generation and testing
- Chi-Square test
- t-Test
- Analysis of variance
- Correlation analysis
- Maximum likelihood test

3. Evaluating what's been learned

- Basic issues
- Training and testing
- Estimating classifier accuracy (holdout, cross-validation, leave-one-out)
- Combining multiple models (bagging, boosting, stacking)
- Minimum Description Length Principle (MLD)

4. Clustering

- Basic issues in clustering

For Techno India NJR Institute of Technology
पंकज पोरवाल
Dr. Pankaj Kumar Porwal
(Principal)

- First conceptual clustering system: Cluster/2
- Partitioning methods: k-means, expectation maximization (EM)
- Hierarchical methods: distance-based agglomerative and divisible clustering
- Conceptual clustering: Cobweb

5.Classification and Regression

- Empirical Risk Minimization
- Nearest Neighbours
- Prototype Based Methods
- Classification and Regression
- Trees
- Linear Regression
- Linear Discriminant Analysis
- Quadratic DiscriminantAnalysis
- Naive Bayes
- Bayesian Methods
- Logistic Regression
- Neural Networks

6. Data mining algorithms: Prediction

- The prediction task
- Statistical (Bayesian) classification
- Bayesian networks
- Instance-based methods
- Linear models

For Techno India NJR Institute of Technology
पंकज पोरवाल
Dr. Pankaj Kumar Porwal
(Principal)