BCA (HONS) 5th SEMESTER

DISCIPLINE SPECIFIC ELECTIVE COURSE (DSE)

OPTION - I

BCA520D2A: DATA MINING

CREDITS: THEORY: 4; PRACTICAL: 2
MAX. MARKS: THEORY: 60; PRACTICAL: 30
MIN. MARKS: THEORY: 24; PRACTICAL: 12

THEORY: 60 LECTURES

UNIT-I

Predictive and descriptive data mining techniques, supervised and unsupervised learning techniques, process of knowledge discovery in databases, pre-processing methods. **(15 Lectures)**

UNIT-II

Association Rule Mining, Association Analysis: Basic concepts, Algorithm, Advanced concepts. **(15 Lectures)**

UNIT-III

Classification and Regression Techniques, Basic Concepts, Decision Trees, Model Evaluation. **(15 Lectures)**

UNIT-IV

Cluster Analysis: Basic Concepts and Algorithm, Scalability and data management issues. **(15 Lectures)**

BOOKS RECOMMENDED:

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson Education.2005.
2. Richard Roiger, Michael Geatz, Data Mining: A Tutorial Based Primer, Pearson Education 2003.
3. G.K. Gupta, Introduction to Data Mining with Case Studies, PHI,2006.
4. Soman K P, DiwakarShyam, Ajay V Insight in to Data Mining: Theory And Practice, PHI, 2006

LAB: BCA519D2A: DATA MINING **(CREDITS: 2; LECTURES:60)**

PRACTICAL EXERCISES BASED ON CONCEPTS LISTED IN THEORY COURSE

BCA (HONS) 5th SEMESTER

DISCIPLINE SPECIFIC ELECTIVE COURSE (DSE)

OPTION - II

BCA520D2B: MACHINE LEARNING

CREDITS: THEORY: 4; PRACTICAL: 2
MAX. MARKS: THEORY: 60; PRACTICAL: 30
MIN. MARKS: THEORY: 24; PRACTICAL: 12

THEORY: 60 LECTURES

UNIT I

Introduction: **(8 Lectures)**

Concept of Machine Learning, Applications of Machine Learning, Key elements of Machine Learning, Supervised vs. Unsupervised Learning, Statistical Learning: Bayesian Method, The Naive Bayes Classifier

Softwares for Machine Learning and Linear Algebra Overview: **(7 Lectures)**

Plotting of Data, Vectorization, Matrices and Vectors: Addition, Multiplication, Transpose and Inverse using available tool such as MATLAB.

UNIT II

Linear Regression: **(15 Lectures)**

Prediction using Linear Regression, Gradient Descent, Linear Regression with one variable, Linear Regression with multiple variables, Polynomial Regression, Feature Scaling/Selection.

UNIT III

Logistic Regression: **(8 Lectures)**

Classification using Logistic Regression, Logistic Regression vs. Linear Regression, Logistic Regression with one variable and with multiple variables.

Regularization: **(7 Lectures)**

Regularization and its utility: The problem of Overfitting Application of Regularization in Linear and Logistic Regression, Regularization and Bias/Variance.

UNIT IV

Neural Networks: **(15 Lectures)**

Introduction, Model Representation, Gradient Descent vs. Perceptron Training, Stochastic Gradient Descent, Multilayer Perceptrons, Multiclass Representation, Backpropagation Algorithm.

Suggested Books:

1. Ethem Alpaydin, "Introduction to Machine Learning" 2nd Edition, The MIT Press, 2009.
2. Tom M. Mitchell, "Machine Learning", First Edition by Tata McGraw-Hill Education, 2013.
3. Christopher M. Bishop, "Pattern Recognition and Machine Learning" by Springer, 2007.
4. Mevin P. Murphy, "Machine Learning: A Probabilistic Perspective" by The MIT Press, 2012.

Lab: Machine Learning (Credits: 4; Theory: 60 Lectures)

For practical Labs for Machine Learning, students may use softwares like MABLAB/Octave or Python. For later exercises, students can create/use their own datasets or utilize datasets from online repositories like UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/).

1. Perform elementary mathematical operations in Octave/MATLAB like addition, multiplication, division and exponentiation.

2. Perform elementary logical operations in Octave/MATLAB (like OR, AND, Checking for Equality, NOT, XOR).

3. Create, initialize and display simple variables and simple strings and use simple formatting for variable.

4. Create/Define single dimension / multi-dimension arrays, and arrays with specific values like array of all ones, all zeros, array with random values within a range, or a diagonal matrix.

5. Use command to compute the size of a matrix, size/length of a particular row/column, load data from a text file, store matrix data to a text file, finding out variables and their features in the current scope.

6. Perform basic operations on matrices (like addition, subtraction, multiplication) and display specific rows or columns of the matrix.

7. Perform other matrix operations like converting matrix data to absolute values, taking the negative of matrix values, adding/removing rows/columns from a matrix, finding the maximum or minimum values in a matrix or in a row/column, and finding the sum of some/all elements in a matrix.

8. Create various type of plots/charts like histograms, plot based on sine/cosine function based on data from a matrix. Further label different axes in a plot and data in a plot.

9. Generate different subplots from a given plot and color plot data.

10. Use conditional statements and different type of loops based on simple example/s.

11. Perform vectorized implementation of simple matrix operation like finding the transpose of a matrix, adding, subtracting or multiplying two matrices.

12. Implement Linear Regression problem. For example, based on a dataset comprising of existing set of prices and area/size of the houses, predict the estimated price of a given house.

13. Based on multiple features/variables perform Linear Regression. For example, based on a number of additional features like number of bedrooms, servant room, number of balconies, number of houses of years a house has been built - predict the price of a house.

14. Implement a classification/ logistic regression problem. For example, based on different features of student's data, classify, whether a student is suitable for a particular activity. Based on the available dataset, a student can also implement another classification problem like checking whether an email is spam or not.

15. Use some function for regularization of dataset based on problem 14.

Use some function for neural networks, like Stochastic Gradient Descent or backpropagation - algorithm to predict the value of a variable based on the dataset of problem 14.