B.S.c II Year IV Semester (CBCS): Data Science

(With Mathematics Combination)
(Examination at the end of Semester - IV)

Paper – IV: Machine Learning

[4 HPW:: 4 Credits :: 100 Marks (External:80, Internal:20)]

Objectives: The main objective of this course is to teach the principles and foundations of machine learning algorithms

Outcomes:
At the end of the course the student will be able to understand

- Basics of Machine Learning and its limitations
- Machine Learning Algorithms: supervised, unsupervised, bio-inspired
- Probabilistic Modeling and Association Rule Mining

Unit-I
Introduction: What does it mean to learn, Some canonical Learning Problems, The Decision Tree Model of Learning, Formalizing the Learning Problem [Reference 1], ID3 Algorithm [[Reference 2]

Limits of Learning: Data Generating Distributions, Inductive Bias, Not Everything is learnable, Underfitting and Overfitting, Separation of training and test Data, Models, parameters and Hyperparameters, Real World Applications of Machine Learning [Reference 1]

Geometry and Nearest Neighbors: From Data to Feature Vectors, k-Nearest Neighbors, Decision Boundaries, k-means Clustering, High Dimensions [Reference 1]

Unit-II
The Perceptron: Bio-inspired Learning, The Perceptron Algorithm, Geometric

Interpretation, Interpreting Perceptron Weights, Perceptron Convergence and Linear

Separability, Improved Generalization, Limitations of the Perceptron [Reference 1]

Practical Issues: Importance of Good Features, Irrelevant and Redundant Features, Feature

Pruning and Normalization, Combinatorial Feature Explosion, Evaluating Model

Performance, Cross Validation, Hypothesis Testing and Statistical Significance, Debugging

Learning Algorithms, Bias Variance tradeoff [Reference 1]

Linear Models: The Optimization Framework for Linear Models, Convex Surrogate Loss

Functions, Weight Regularization, Optimization and Gradient Descent, Support Vector Machines [Reference 1]

Unit-III
Probabilistic Modeling: Classification by Density Estimation, Statistical Estimation, Naïve Bayes Models, Prediction [Reference 1]

Neural Networks: Bio-inspired Multi-Layer Networks, The Back-propagation Algorithm,

Initialization and Convergence of Neural Networks, Beyond two layers, Breadth vs Depth, Basis Functions [Reference 1]

Unit IV
Unsupervised Learning: Clustering Introduction, Similarity and Distance Measures,

Agglomerative Algorithms, Divisive Clustering, Minmum Spanning Tree [Reference 2] **Association Rules:** Introduction, large Itemsets, Apriori Algorithm [Reference 2]

References:
1. A Course in Machine Learning (CIML). Hal Daume III, 2017 (freely available online) http://ciml.info/
2. Data Mining: Introductory and Advanced Topics. Margaret H Dunham, Pearson Education, 2003

Suggested Reading:

1. Hands on Machine Learning with SciKit-Learn, Keras and Tensor Flow. AurélienGéron. O'Reily, 2019
2. Machine Learning with Python Cookbook. Chris Albo, O'Reily, 2018
3. Introduction to Machine Learning with Python: A guide. Andreas C Miller, Sarah Guido. O'Reily, 2017