

B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester IV
BSDB32403T: Machine Learning

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 40%

Objectives

This course aims to provide sound foundation to fundamental concepts of machine learning and its application and prepare students for advanced research and real time problem solving in machine learning and related fields. This course will help students to understand the concepts related to regression, classification, clustering, concept clarity on deep learning, dimensionality reduction, model selection and boosting.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER:

1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A

Unit I: Introduction: Applications of Machine learning, machine learning as a future; Data Pre-processing: Importing the libraries, Importing the dataset, taking care of missing data, encoding categorical data, Splitting the dataset into training set and test set, Feature scaling.

Unit II: Regression: Simple linear regression, Multiple linear regression, Understanding the P-value, Polynomial regression.

Unit III: Classification: Logistic Regression, K-Nearest Neighbors, Support vector machine, Naïve Bayes, Decision tree classification, Random forest classification.

Unit IV: Clustering: k-means clustering, k means random initialization trap, selecting the number of clusters, Hierarchical clustering.

Section B

Unit V: ANN: Introduction to ANNs, Biological Neural Networks; Usefulness and Applications of ANNs; Architectures of ANNs: Single layer, Multi layer, Competitive layer; Learning: Supervised and Unsupervised; Activation functions; Linear and Non-linear Separability

Unit VI: Supervised Models: Hebb Net: introduction, algorithm, application for AND problem; **Perceptron:** architecture, algorithm, application for OR Problem; ADALINE: architecture, algorithm, application for XOR problem;

Unit VII: MADALINE: architecture, algorithm, application for XOR problem; Backpropagation Neural Network: architecture, parameters, algorithm, applications, different issues regarding convergence

Unit VIII Unsupervised Models: Kohonen Self –Organizing Maps: architecture, algorithm, application, Adaptive Resonance Theory: introduction, basic architecture, basic operation, ART1 and ART2

Suggested Readings

1. Andreas C. Müller, Introduction to Machine Learning with Python: A Guide for Data Scientists, Sarah Guido, 2016
2. E. Alpaydin, Introduction to Machine Learning, 3rd Edition, PHI Learning, 2015
3. K. P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012
4. <https://www.udemy.com/course/machinelearning/>