|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Course Code:**  CSE3008 | **Course Title: Introduction to Machine Learning** | | **TPC** | | **3** | **2** | **4** |
| **Version No.** | **1.1** | | | | | | |
| **Course Pre-requisites/ Co-requisites** |  | | | | | | |
| **Anti-requisites (if any).** | None | | | | | | |
| **Objectives:** | 1. Introduce the fundamental problems of machine learning. 2. Provide understanding of techniques, mathematical concepts, and algorithms used in machine learning to facilitate further study in this area. 3. Provide understanding of the limitations of various machine learning algorithms and the way to evaluate performance of machine learning algorithms. 4. Practice software implementation of different concepts and algorithms covered in the course. | | | | | | |
| **Expected Outcome:** | On completion of the course, students will have the ability to   1. Understands basic machine learning techniques 2. How to apply Neural Networks 3. Apply the concepts of Bayesian and Computational Learning 4. Analyze different instance based learning models 5. Understand Hidden Markav Models and its applications to solve real time problems | | | | | | |
| **Module No. 1** | **Introduction** | **7 Hours** | | | | | |
| Learning problems, perspectives and issues, concept learning, version spaces and candidate eliminations, inductive bias, decision tree learning, representation, algorithm, Ensemble Learning and Random Forest | | | | | | | |
| **Module No. 2** | **Training Models** | **7 Hours** | | | | | |
| Linear Regression – The Normal equation, Computational complexity, Gradient Descent, Polynomial Regression, Learning Curves, Regularized Linear Models, Logistic Regression, Regularized Logistic Regression, Multi class classification using Logistic Regression,  K-Nearest neighbor learning, locally weighted regression, radial basis functions, case based learning. | | | | | | | |
| **Module No. 3** | **Support Vector Machine** | **7 Hours** | | | | | |
| functional and geometric margins, optimum margin classifier, constrained optimization, Lagrange multipliers, primal/dual problems, KKT conditions, dual of the optimum margin classifier, soft margins, kernels, SVM multi class classification, SVM Regression | | | | | | | |
| **Module No. 4** | **Bayesian and Computational Learning** | **8Hours** | | | | | |
| Bayes theorem , concept learning, maximum likelihood, minimum description length principle, Bayes optimal classifier, Gibbs Algorithm, Naïve Bayes Classifier, Bayesian belief network, EM algorithm, probability learning, sample complexity | | | | | | | |
| **Module No. 5** | **Artificial Neural Networks** | | **8 Hours** | | | | |
| ANN- Introduction, Neural Network Representations, Problems for Neural Network Learning, Perceptrons, Multilayer Networks and Backpropagation Algorithm, Introduction to deep learning architectures. | | | | | | | |
| **Module No. 6** | **Hidden Markov Models** | | | **8 Hours** | | | |
| Introduction, discrete Markov processes, hidden Markov models, three basic problems of HMMs evaluation problem, finding the state sequence, learning model parameters, continuous observations, the HMM with input, model selection in HMM. | | | | | | | |
| **Text Books**   1. Tom M. Mitchell, Machine Learning, McGraw Hill , 2017. 2. EthemAlpaydin, Introduction to Machine Learning (Adaptive Computation and Machine Learning), The MIT Press, 2015 | | | | | | | |
| **References**   1. AurelienGeron, Hands-On Machine Learning WithScikit-Learn and Tensorflow, O’Really publication 2019 2. Shai Shalev-Shwartz and Shai Ben-David, U[nderstanding Machine Learning](http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/), Cambridge University Press. 2017 3. T. Hastie, R. Tibshirani, J. H. Friedman, Introduction to Statistical Machine Learning 1/e, Springer, 2017.   4. M NarasimhaMurty, Introduction to Pattern Recognition and Machine Learning, World Scientific Publishing Company, 2015 | | | | | | | |
| Lab Exercises  In machine learning lab session student will work selected set of problems and case studies on the following topics, tools and techniques.   1. Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file 2. For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples 3. Write a program to demonstrate the working of the decision tree based ID3 algorithm consider weather dataset. 4. Consider the Titanic dataset, summarized according to economic status (class), sex, age and survival. Using Random Forest asked to predict whether a passenger on the titanic would have been survived or not. 5. Implement Linear regression on IRIS dataset 6. Implement Logistic Regression for digit recognition 7. Implement Support Vector Machine for digit recognition and compare the accuracy with logistic regression 8. SVM Multiclass classification categorizing news article to sports, politics, economics, or social 9. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. 10. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets. K-NN and Weighted- KNN classifiers  |  | | --- | | 1. Apply EM algorithm to cluster a set of data stored in a .CSV file |  1. Implement Hidden Markov Model for malware classification | | | | | | | |
| **Mode of Evaluation** | **Practice Tests-20%, Continuous Assessment Tests-60%, Practical Assesment-20%**  Practice Tests - Cumulative for 16 Weeks 20%  Continuous Assessment Test-1 20%  Continuous Assessment Test-2 20%  Continuous Assessment Test-3 20%  Practical Assessment (Mini Project) 20% | | | | | | |
| **Recommended by the Board of Studies on** | 07.01.2021 | | | | | | |
| **Date of Approval by the Academic Council** |  | | | | | | |