| CSE4020 | MACHINE LEARNING | L | T | P | J | C |
|---|---|---|---|---|---|---|
| | | 2 | 0 | 2 | 4 | 4 |
| **Pre-requisite** | **Nil** | **Syllabus version** | | | | |
| | | v1.0 | | | | |

**Course Objectives:**

1. Ability to comprehend the concept of supervised and unsupervised learning techniques
2. Differentiate regression, classification and clustering techniques and to implement their algorithms.
3. To analyze the performance of various machine learning techniques and to select appropriate features for training machine learning algorithms.

**Expected Course Outcome:**

1. Recognize the characteristics of machine learning that makes it useful to solve real-world problems.
2. Provide solution for classification and regression approaches in real-world applications.
3. Gain knowledge to combine machine learning models to achieve better results.
4. Choose an appropriate clustering technique to solve real world problems.
5. Realize methods to reduce the dimension of the dataset used in machine learning algorithms.
6. Choose a suitable machine learning model, implement and examine the performance of the chosen model for a given real world problems.
7. Understand cutting edge technologies related to machine learning applications.

| **Module:1** | **Introduction to Machine Learning** | **3 hours** |
|---|---|---|

What is Machine Learning, Examples of Various Learning Paradigms, Perspectives and Issues, Version Spaces, Finite and Infinite Hypothesis Spaces, PAC Learning

| **Module:2** | **Supervised Learning - I** | **4 hours** |
|---|---|---|

Learning a Class from Examples, Linear, Non-linear, Multi-class and Multi-label classification, Generalization error bounds: VC Dimension, Decision Trees: ID3, Classification and Regression Trees, Regression: Linear Regression, Multiple Linear Regression, Logistic Regression.

| **Module:3** | **Supervised Learning - II** | **5 hours** |
|---|---|---|

Neural Networks: Introduction, Perceptron, Multilayer Perceptron, Support vector machines: Linear and Non-Linear, Kernel Functions, K-Nearest Neighbors

| **Module:4** | **Ensemble Learning** | **3 hours** |
|---|---|---|

Ensemble Learning Model Combination Schemes, Voting, Error-Correcting Output Codes, Bagging: Random Forest Trees, Boosting: Adaboost, Stacking

| **Module:5** | **Unsupervised Learning - I** | **7 hours** |
|---|---|---|

Introduction to clustering, Hierarchical: AGNES, DIANA, Partitional : K-means clustering, K-Mode Clustering, Self-Organizing Map, Expectation Maximization, Gaussian Mixture Models

| **Module:6** | **Unsupervised Learning - II** | **3 hours** |
|---|---|---|

Principal components analysis (PCA), Locally Linear Embedding (LLE), Factor Analysis

| **Module:7** | **Machine Learning in Practice** | **3 hours** |
|---|---|---|

Machine Learning in Practice Design, Analysis and Evaluation of Machine Learning Experiments, Feature selection Mechanisms, Other Issues: Imbalanced data, Missing Values, Outliers

| **Module:8** | **Recent Trends** | **2 hours** |
|---|---|---|

Industry Expert talk

| | | | |
|---|---|---|---|
| | **Total Lecture hours:** | **30 hours** | |

| | |
|---|---|
| **Text Book(s)** | |
| 1. | Ethem Alpaydin, Introduction to Machine Learning , MIT Press, Prentice Hall of India, Third Edition 2014 |
| **Reference Books** | |
| 1. | Sergios Theodoridis, Konstantinos Koutroumbas, Pattern Recognition, Academic Press, 4th edition, 2008, ISBN:9781597492720. |
| 2. | Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar ”Foundations of Machine Learning, MIT Press, 2012 |
| 3. | Tom Mitchell, Machine Learning, McGraw Hill, 3rd Edition,1997. |
| 4 | Charu C. Aggarwal, Data Classification Algorithms and Applications , CRC Press, 2014 |
| 5 | Charu C. Aggarwal, DATA CLUSTERING Algorithms and Applications, CRC Press, 2014 |
| 6 | Kevin P. Murphy ”Machine Learning: A Probabilistic Perspective”, The MIT Press, 2012 |
| Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar | |

| **List of Challenging Experiments (Indicative)** | | |
|---|---|---|
| 1. | Implement Decision Tree learning. | 2 hours |
| 2. | Implement Logistic Regression. | 2 hours |
| 3. | Implement classification using Multilayer perceptron. | 2 hours |
| 4. | Implement classification using SVM | 2 hours |
| 5. | Implement Adaboost | 2 hours |
| 6. | Implement Bagging using Random Forests | 2 hours |
| 7. | Implement K-means Clustering to Find Natural Patterns in Data. | 2 hours |
| 8. | Implement Hierarchical clustering. | 2 hours |
| 9. | Implement K-mode clustering | 2 hours |
| 10. | Implement Principle Component Analysis for Dimensionality Reduction. | 2 hours |
| 11. | Implement Multiple Correspondence Analysis for Dimensionality Reduction. | 2 hours |
| 12. | Implement Gaussian Mixture Model Using the Expectation Maximization. | 2 hours |
| 13. | Evaluating ML algorithm with balanced and unbalanced datasets. | 2 hours |
| 14. | Comparison of Machine Learning algorithms. | 2 hours |
| 15. | Implement k-nearest neighbors algorithm | 2 hours |
| | Total Laboratory Hours | 30 hours |

| | | | |
|---|---|---|---|
| Mode of assessment: Project/Activity | | | |
| Recommended by Board of Studies | 04-04-2014 | | |
| Approved by Academic Council | No. 37 | Date | 16-06-2015 |