13M1WCI331: Machine Learning

Course Credit: 3
Semester: M.Tech, III

Introduction

The intent of this course is to present a broad introduction to Machine Learning, the study of computing systems that improve their performance with experience, including discussions of each of the major approaches. The primary focus of the course will be on understanding the underlying algorithms used in various learning systems. Class lectures will discuss general issues as well as present abstract algorithms in research context.
The format of the course will be lecture-discussions, assignments and project based. Students are strongly encouraged to participate actively in class discussions.

Course Objectives (Post-conditions)

Knowledge objectives:

On completion of the course students will be expected to:

- Have a good understanding of the fundamental issues and challenges of machine learning: data, model selection, model complexity, etc.
- Have an understanding of the strengths and weaknesses of many popular machine learning approaches.
- Appreciate the underlying mathematical relationships within and across Machine Learning algorithms and the paradigms of supervised and un-supervised learning.
- Be able to design and implement various machine learning algorithms in a range of real-world applications.
- Be able to read current research papers and understand the issues raised by current research.

Application objectives:

The homework portions of the course are intended to help you apply your understanding of machine learning algorithms and their use in data-driven knowledge discovery and program synthesis. You will design and implement several machine learning algorithms in Java. You will also be able to identify formulate and solve machine learning problems that arise in practical applications. You will have a knowledge of the strengths and weaknesses of different machine learning algorithms (relative to the characteristics of the application domain) and be able to adapt or combine some of the key elements of existing machine learning algorithms to design new algorithms as needed. You will have an understanding of the current state of the art in machine learning and be able to begin to conduct original research in machine learning.

Expected Student Background (Preconditions)

Machine Learning is a mathematical discipline and as such students will benefit from a familiarity with probability theory and linear algebra, as well as some basic programming skills.

Topics Outline:

S NO	Topics	Hrs
1	Introduction: Definition of learning systems. Goals and applications of machine learning. Aspects of developing a learning system: training data, concept representation, function approximation.	3
2	Data Preprocessing: Descriptive Data Summarization, Data Cleaning, Data Integration and Transformation, Data Reduction, Data Discretization and Concept Hierarchy Generation, From Data Warehousing to Machine Learning	3
3	Inductive Classification: The concept learning task. Concept learning as search through a hypothesis space. General-to-specific ordering of hypotheses. Finding maximally specific hypotheses. Version spaces and the candidate elimination algorithm. Learning conjunctive concepts. The importance of inductive bias.	6
4	Decision Tree Learning: Representing concepts as decision trees. Recursive induction of decision trees. Picking the best splitting attribute: entropy and information gain. Searching for simple trees and computational complexity. Occam's razor. Overfitting, noisy data, and pruning.	5
5	Ensemble Learning: Using committees of multiple hypotheses. Bagging, boosting, and DECORATE. Active learning with ensembles.	3
6	Experimental Evaluation of Learning Algorithms: Measuring the accuracy of learned hypotheses. Comparing learning algorithms: cross-validation, learning curves, and statistical hypothesis testing.	3
7	Rule Learning: Propositional and First-Order: Translating decision trees into rules. Heuristic rule induction using separate and conquer and information gain. First-order Horn-clause induction (Inductive Logic Programming) and Foil. Learning recursive rules. Inverse resolution, Golem, and Progol, Association Rule Mining	5
8	Artificial Neural Networks: Neurons and biological motivation. Linear threshold units. Perceptrons: representational limitation and gradient descent training. Multilayer networks and backpropagation. Hidden layers and constructing intermediate, distributed representations. Overfitting,	4

	learning network structure, recurrent networks.	
9	Support Vector Machines: Maximum margin linear separators. Quadractic programming solution to finding maximum margin separators. Kernels for learning non-linear functions.	5
10	Bayesian Learning: Probability theory and Bayes rule. Naive Bayes learning algorithm. Parameter smoothing. Generative vs. discriminative training. Logisitic regression. Bayes nets and Markov nets for representing dependencies.	5
11	Instance-Based Learning: Constructing explicit generalizations versus comparing to past specific examples. k-Nearest-neighbor algorithm. Case-based learning.	3
12	Clustering and Unsupervised Learning: Learning from unclassified data. Clustering. Hierarchical Aglomerative Clustering. k-means partitional clustering. Expectation maximization (EM) for soft clustering. Semi-supervised learning with EM using labeled and unlabled data.	5
	Total	50

References

1. "Machine Learning" by Tom Mitchell, McGraw Hill, 1997, ISBN 0070428077 will be used as the main text book; however the inputs will be supplemented with information from elsewhere wherever the same is required.
2. T. Hastie, R. Tibshirani, & J. H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer Verlag, 2001.
3. Ian H. Witten & Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, 1999.
4. S. M. Weiss & C. A. Kulikowski, Computer Systems that Learn, Morgan Kaufman Publishers, San Fancisco, CA, 1991

Evaluation Scheme:

S.No	Examination	Marks
1	T-1	15
2	T-2	25
3	T-3	35
4	*Internal Marks	25

*Internal Marks Breakdown:

Assignments	9 marks (3x3)
Quizzes	12 marks (3x4)
Regularity	4 Marks