AMITY SCHOOL OF ENGINEERING & TECHNOLOGY

Machine Learning

Total teaching weeks: 14 **Semester – VIII**

Total Lectures: 40 **Credit – 3**

	TOPICS TO BE COVERED	LECTURE
1	**UNIT I** Definition of learning system, Goal and applications of Machine learning	2
2	Aspects of developing a learning system: training data, concept representation, function approximation.	2
3	Supervised learning, unsupervised learning, overview of classification.	2
4	Setup, training, test, validation test, over fitting, classification families	2
5	Linear discriminative, non-linear discriminative, decision trees, probabilistic (conditional and generative), nearest neighbor	2
6	**UNIT II** Logistic regression, perceptron, Exponential family	2
7	Generative learning algorithms, Gaussian discriminant analysis, Naïve Bayes,	3
8	Support vector machine: optimal hyper planes, kernels, Model selection and feature selection	2
9	Combining classifier: Bagging, boosting (The Ada Boost algorithm), Evaluating and debugging learning algorithms. Classification errors	3
	End of 1st Term	
10	**UNIT III** Clustering K means, EM Algorithm, Mixture of Gaussians	3
11	Factor Analysis, Principal Component Analysis, Independent Component Analysis	3
12	Latent Semantic Indexing, Spectral clustering, Markov Model, Hidden Markov Model	4
13	**UNIT IV**	
14	Reinforcement Learning, MDP, Bellman equations, Value iteration and policy iteration,	4
15	Linear quadratic regularization (LQR), LQG Q-learning	3
16	Value function approximation, Policy search, Reinforce, POMDPs	3
17	**End of 2nd Term**	

Text Books:

[T1] TOM M Mitchell, "Machine Learning, McGraw Hill Education
[T2] Bishop, C (2006) Pattern Recognition and Machine Learning, Berlin Springer Verlag
[T3] Duda, Richard, Peter Hart, and David stock. Pattern Classification. 2nd ed. New York, NY: Wiley-Interscience, 2000, ISBN:9780471056690.
[T4] Tom M.Mitchell, Machine Learning. ISBN-9781259096952, McGraw Hill series, Edition First.

Reference Books:

[R1] Bishop, Christopher. Neural Networks for Pattern Recognition. New York, NY:Oxford, University Press, 1995.
[R2] Introduction to Machine Learning – Ethem Alpaydin, MIT Press, PHI.

MACHINE LEARNING

Paper Code: ETCS-402	**L**	**T/P**	**C**
Paper: Machine Learning	**3**	**0**	**3**

INSTRUCTIONS TO PAPER SETTERS: **MAXIMUM MARKS: 75**

1. Question No. 1 should be compulsory and cover the entire syllabus. This question should have objective or short answer type questions. It should be of 25 marks.
2. Apart from Question No. 1, rest of the paper shall consist of four units as per the syllabus. Every unit should have two questions. However, student may be asked to attempt only 1 question from each unit. Each question should be of 12.5 marks.

Objective: To introduce the students about the knowledge of basic concepts of machine learning systems, types of learning etc.

UNIT-I

Introduction:

Basic concepts: Definition of learning systems, Goals and applications of machine learning. Aspects of developing a learning system: training data, concept representation, function approximation.

Types of Learning: Supervised learning and unsupervised learning. Overview of classification: setup, training, test, validation dataset, over fitting.

Classification Families: linear discriminative, non-linear discriminative, decision trees, probabilistic (conditional and generative), nearest neighbor.

[T1, T2][No. of Hrs: 12]

UNIT-II

Logistic regression, Perceptron, Exponential family, Generative learning algorithms, Gaussian discriminant analysis, Naive Bayes, Support vector machines: Optimal hyper plane, Kernels. Model selection and feature selection. Combining classifiers: Bagging, boosting (The Ada boost algorithm), Evaluating and debugging learning algorithms, Classification errors.

[T1, T2][No. of Hrs: 11]

UNIT-III

Unsupervised learning: Clustering. K-means. EM Algorithm. Mixture of Gaussians. Factor analysis. PCA (Principal components analysis), ICA (Independent components analysis), latent semantic indexing. Spectral clustering, Markov models Hidden Markov models (HMMs).

[T1, T2][No. of Hrs: 11]

UNIT-IV

Reinforcement Learning and Control: MDPs. Bellman equations, Value iteration and policy iteration, Linear quadratic regulation (LQR). LQG. Q-learning. Value function approximation, Policy search. Reinforce. POMDPs.

[T1, T2][No. of Hrs: 10]

Text Books:

[T1] Tom M Mitchell, Machine Learning, McGraw Hill Education

[T2] Bishop, C. (2006). Pattern Recognition and Machine Learning. Berlin: Springer-Verlag.

[T3] Duda, Richard, Peter Hart, and David Stork. Pattern Classification. 2nd ed. New York, NY: Wiley-Interscience, 2000. ISBN: 9780471056690.

[T4] Tom M. Mitchell, Machine Learning .ISBN – 9781259096952, McGraw-Hill Series, Edition – First

Reference Books:

[R1] Bishop, Christopher. Neural Networks for Pattern Recognition. New York, NY: Oxford University Press, 1995. ISBN: 9780198538646.

[R2] Introduction to Machine Learning - Ethem Alpaydin, MIT Press, Prentice hall of India.

AMITY SCHOOL OF ENGINEERING & TECHNOLOGY
Machine Learning (ETCS-402)

Assignment:

1. Write Matlab code for McCulloch-Pitts neuron for implementation of logical NOT, AND and OR gates.
2. Write a Matlab program to implement Back Propagation Method.
3. Write a Matlab program to calculate union, intersection, complement and difference of two fuzzy sets.
4. Write a Matlab program to calculate the Demorgan's Law.
5. Write a Matlab program to find whether the given matrix is (a) reflexive (b) tolerance and (c) transitivity matrix or not.
6. Write a Matlab program to find whether the given matrix is symmetry or not.
7. Using max–product and max-min method by a Matlab program to find the fuzzy relation between two vectors.
8. Using Matlab program find the crisp lambda cut set relations.
9. Use Matlab's Fuzzy Logic Toolbox to model the tip given after a dinner for two, where the food can be disgusting, not good, bland, satisfying, good, or delightful, and the service can be poor, average, or good. To get started, you type fuzzy in a Matlab window. Then use the fuzzy inference system and membership function editors to define and tune your rules.
10. Write a program in Matlab to implement Roulette wheel and ranking selection method.

Dr. A K Yadav

MACHINE LEARNING LAB

Paper Code: ETCS-454	L	T/P	C
Paper: Machine Learning Lab	0	2	1

List of Experiments:

1. Study and Implement the Naive Bayes learner using WEKA. (The datasets taken can be: Breast Cancer data file or Reuters data set).
2. Study and Implement the Decision Tree learners using WEKA. (The datasets taken can be: Breast Cancer data file or Reuter's data set).
3. Estimate the accuracy of decision classifier on breast cancer dataset using 5-fold cross-validation. (You need to choose the appropriate options for missing values).
4. Estimate the precision, recall, accuracy, and F-measure of the decision tree classifier on the text classification task for each of the 10 categories using 10-fold cross-validation.
5. Develop a machine learning method to classifying your incoming mail.
6. Develop a machine learning method to Predict stock prices based on past price variation.
7. Develop a machine learning method to predict how people would rate movies, books, etc.
8. Develop a machine learning method to Cluster gene expression data, how to modify existing methods to solve the problem better
9. Select two datasets. Each dataset should contain examples from multiple classes. For training purposes assume that the class label of each example is unknown (if it is known, ignore it). Implement the K-means algorithm and apply it to the data you selected. Evaluate performance by measuring the sum of Euclidean distance of each example from its class center. Test the performance of the algorithm as a function of the parameter k.
10. Implement the EM algorithm assuming a Gaussian mixture. Apply the algorithm to your datasets and report the parameters you obtain. Evaluate performance by measuring the sum of Mahalanobis distance of each example from its class center. Test performance as a function of the number of clusters.
11. Suggest and test a method for automatically determining the number of clusters.
12. Using a dataset with known class labels compare the labeling error of the K-means and EM algorithms. Measure the error by assigning a class label to each example. Assume that the number of clusters is known.

NOTE:- At least 8 Experiments out of the list must be done in the semester.

GURU GOBIND SINGH
INDRAPRASTHA
UNIVERSITY