FOUNDATIONS OF MACHINE LEARNING LABORATORY

IV Semester: CSE(AI & ML)

Course Code	Category	Hours / Week			Credits	Maximum Marks		
		L	T	P	C	CIA	SEE	Total
ACAC04	Core	0	0	2	1	30	70	100
Contact Classes: Nil	Tutorial Classes: Nil	Practical Classes: 24				Total Classes:24		

Prerequisite: Python Programming

I. COURSE OVERVIEW:

This course introduces the fundamental concepts and methods of machine learning, including the description and analysis of several modern algorithms, their theoretical basis, and the illustration of their applications. Machine learning as a field is now incredibly pervasive, with applications spanning from business intelligence to text and speech processing, bioinformatics, and other areas in real-world products and services. This will familiarize students with a broad cross-section of models and algorithms for machine learning, and prepare students for research or industry application of machine learning techniques.

II. COURSE OBJECTIVES:

The students will try to learn:

I. The underlying mathematical principles from probability, linear algebra and optimization.
II. The knowledge of using machine learning to make predictions in a scientific computing environment.
III. The underlying mathematical relationships within and across machine learning algorithms and the paradigms of supervised and un-supervised learning.
IV. The advanced topics such as robotics, machine learning, deep learning, pattern recognition, computer vision, cognitive computing, human-computer interaction etc.

III. COURSE SYLLABUS:

Week – 1: FIND-S ALGORITHM

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.

Week – 2: CANDIDATE ELIMINATION ALGORITHM

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.

Week – 3: ID3 ALGORITHM

3. Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.

Week – 4: ARTIFICIAL NEURAL NETWORK

4. Build an Artificial Neural Network by implementing the Back propagation algorithm and test the same using appropriate data sets.

Week – 5: NAÏVE BAYESIAN CLASSIFIER

5. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.

Week – 6: NAÏVE BAYESIAN CLASSIFIER

6. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Calculate the accuracy, precision, and recall for your data set.

Week – 7: HEART DISEASE PREDICTION USING BAYESIAN NETWORK

7. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set.

Week – 8: CLUSTERING USING EM AND k-MEANS ALGORITHM

8. Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering.

Week – 9: BUILT-IN FUNCTIONS

9. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions.

Week – 10: REGRESSION

10. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.

Week – 11: CLUSTERING

11. Write a program to for automatically determining the number of clusters.

Week – 12: k-MEANS ALGORITHM

12. 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.

IV. REFERENCE BOOKS:

1. Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, "Foundations of Machine Learning", MIT Press, 2nd Edition, 2018.
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Springer, 2nd Edition, 2009.
3. Avrim Blum, John Hopcroft, Ravindran Kannan, "Foundations of Data Science", Cambridge University Press, 2020.
4. Tom M. Mitchell, "Machine Learning", Mc Graw Hill, Indian Edition, 2017.
5. Gareth James, Daniela Witten, Trevor Hastie and Rob Tibshirani, "An Introduction to Statistical Learning: with applications in R", Springer Texts in Statistics, 2017.

V. WEB REFERENCES:

1. https://onlinecourses.nptel.ac.in/noc19_cs52/preview
2. https://ece.iisc.ac.in/~parimal/2019/ml.html
3. https://www.springer.com/gp/book/9780387848570
4. https://www.cse.iitb.ac.in/~sunita/cs725/calendar.html
5. https://www.analyticsvidhya.com/blog/2018/12/guide-convolutional-neural-network-cnn/
6. https://cs.nyu.edu/~mohri/mlu11/