AL405 Machine Learning

Course Objectives: This course provides a broad introduction to machine learning. It offers some of the most cost-effective approaches to automated knowledge acquisition in emerging data-rich disciplines and focuses on the theoretical understanding of these methods, as well as their computational implications. To provide an understanding of the theoretical concepts of machine learning and prepare students for research or industry application of machine learning techniques

Unit I :Introduction to machine learning, scope and limitations, machine learning models, Supervised Learning, Unsupervised Learning, hypothesis space and inductive bias, evaluation, cross-validation, Dimensionality Reduction: Subset Selection, Shrinkage Methods, Principle Components Analysis, Partial Least Squares.

Unit II :Neural Networks: From Biology to Simulation, Neural network representation, Neural Networks as a paradigm for parallel processing Perceptron Learning, Training a perceptron, Multilayer perceptron, back propagation Algorithm, Training & Validation, Activation functions, Vanishing and Exploding Gradients.

Unit III :Supervised Learning Techniques:- Decision Trees, Naive Bayes, Classification, Support vector machines for classification problems, Random forest for classification and regression problems, Linear regression for regression problems, Ordinary Least Squares Regression, Logistic Regression.

Unit IV :Unsupervised Learning, Clustering: k-means, adaptive hierarchical clustering, Gaussian mixture model, Optimization Using Evolutionary Techniques, Number of Clusters, Advanced discussion on clustering, Expectation Maximization.

Unit V :Design and Analysis of Machine Learning Experiments:Factors, response and strategy of experimentation, Guidelines for machine learning experiments, cross-validation and resampling methods, Measuring classifier performance, Hypothesis testing, comparing multiple algorithms, comparison over multiple datasets

Books and references :
1. Machine Learning. Tom Mitchell. First Edition, McGraw- Hill, 1997.
2. Introduction to Machine Learning Edition 2, by EthemAlpaydin
3. Introduction to Machine learning, Nils J.Nilsson
4. Machine learning for dummies, IBM Limited ed, by Judith Hurwitz and Daniel Kirsch
5. Introduction to Machine Learning with Python A guide for data scientists, Andreas, C. Muller & Sarah Guido, O'Reilly

List of Experiments:

Different problems to be framed to enable students to understand the concept learnt and get hands-on on various tools and software related to the subject

1. How to calculate important numbers based on data sets, how to use various Python modules and how to make functions that are able to predict the outcome based on what we have learned (Small Dataset)
2. How can we get Big Data Sets, Learn: Data Distribution, Normal data distribution, Random Data Distribution, Scatter Plot.
3. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.
4. The probability that it is Friday and that a student is absentis 3%. Since there are 5 school days in a week,the probability that it is Friday is 20%. What is the probability that a student is absent given that today is Friday? Apply Baye's rule in python to get the result.
5. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set.
6. Implement and demonstrate simple Linear Regression Algorithm based on a given set of training data samples. Read the training data from a .CSV file.
7. Implement and demonstrate Logistic Regression (Binomial) based on a given set of training data samples. Read the training data from a .CSV file.
8. Implement and demonstrate 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.
9. 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.
10. Given the following data, which specify classifications for nine combinations of VAR1 and VAR2 predict a classification for a case where VAR1=0.906 and VAR2=0.606, using the result of k-means clustering with 3 means (i.e., 3 centroids)

VAR1	VAR2	CLASS
1.713	1.586	0
0.180	1.786	1
0.353	1.240	1
0.940	1.566	0
1.486	0.759	1
1.266	1.106	0
1.540	0.419	1
0.459	1.799	1
0.773	0.186	1

MACHINE LEARNING

Machine Learning

Supervised Learning   Unsupervised Learning
        ↓   ↓
labelled data   Unlabelled date
      ↙   ↘   ↓
Regression   Classification   clustering / Grouping
(Continuous   (Discrete.   (IInd Unit)
output)   Output)
(Ist Unit)   (IIIrd & IVth Unit)