RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Data Science/Data Science, V semester
CD502-Machine Learning

COURSE OBJECTIVES:

The objective of this course is to impart necessary knowledge of different machine learning techniques and develop programming skills required to build machine learning based applications.

COURSE OUTCOMES:

After completing the course student should be able to:

1. Describe in-depth about theories, methods, and algorithms in machine learning.
2. Find and analyze the optimal hyper parameters of the machine learning algorithms.
3. Examine the nature of a problem at hand and determine whether machine learning can solve it efficiently.
4. Solve and implement real world problems using machine learning.

COURSE CONTENTS:

UNIT-I

Introduction to machine learning, Machine learning life cycle, Types of Machine Learning System (supervised and unsupervised learning, Batch and online learning, Instance-Based and Model based Learning), scope and limitations, Challenges of Machine learning, data visualization, hypothesis function and testing, data pre-processing, data augmentation, normalizing data sets, , Bias-Variance tradeoff, Relation between AI (Artificial Intelligence), ML (Machine Learning), DL (Deep Learning) and DS (Data Science).

UNIT-II

Clustering in Machine Learning: Types of Clustering Method: Partitioning Clustering, Distribution Model-Based Clustering, Hierarchical Clustering, Fuzzy Clustering. Birch Algorithm, CURE Algorithm. Gaussian Mixture Models and Expectation Maximization. Parameters estimations – MLE, MAP. Applications of Clustering.

UNIT-III

Classification algorithm: - Logistic Regression, Decision Tree Classification, Neural Network, K-Nearest Neighbors (K-NN), Support Vector Machine, Naive Bayes (Gaussian, Multinomial, Bernoulli). Performance Measures: Confusion Matrix, Classification Accuracy, Classification Report: Precisions, Recall, F1 score and Support.

UNIT-IV

Ensemble Learning and Random Forest**:** Introduction to Ensemble Learning, Basic Ensemble Techniques (Max Voting, Averaging, Weighted Average), Voting Classifiers, Bagging and Pasting, Out-of-Bag Evaluation, Random Patches and Random Subspaces, Random Forests (Extra-Trees, Feature Importance), Boosting (AdaBoost, Gradient Boosting), Stacking.

UNIT-V

Dimensionality Reduction:The Curse of Dimensionality, Main Approaches for Dimensionality Reduction (Projection, Manifold Learning) PCA: Preserving the Variance, Principal Components, Projecting Down to d Dimensions, Explained Variance Ratio, Choosing the Right Number of Dimensions, PCA for Compression, Randomized PCA, Incremental PCA. Kernel PCA: Selecting a Kernel and Tuning Hyper parameters. Learning Theory: PAC and VC model.

REFERENCE BOOKS:

1. Tom M. Mitchell, “Machine Learning”, McGraw Hill Education, First edition, 2017.
2. Aurelien Geon, “Hands-On Machine Learning with Scikit-Learn and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems”, Shroff/O'Reilly; First edition (2017).
3. Andreas Muller, "Introduction to Machine Learning with Python: A Guide for Data Scientists", Shroff/O'Reilly; First edition (2016).
4. Leonard Kaufman and P. J. Rousseau. Finding groups in data: An introduction to cluster analysis, Wiley, 2005
5. NelloCristianini and John Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge University Press, 2000.

PRACTICAL:

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. Such assignments are to be framed for ten to twelve lab sessions.