Savitribai Phule Pune University
Fourth Year of Artificial Intelligence and Data Science (2020 Course)
417521: Machine Learning

Teaching Scheme: TH: 03 Hours/Week	Credit 03	Examination Scheme: In-Sem (Paper): 30 Marks End-Sem (Paper): 70 Marks

Prerequisites Courses: Data Science (317529), Artificial Neural Network (317531)

Course Objectives:

- Explain the learning paradigms, and models of machine learning
- Apply different regression techniques for making predictions in different applications
- Apply the classification algorithms to classify the data with appropriate labels
- Apply the clustering algorithms to divide the unlabeled data into the similar groups
- Introduce and integrate models in the form of advanced ensembles
- Explain reinforcement learning and its algorithms

Course Outcomes:

After completion of the course, learners should be able to-

CO1: Describe and compare different models of machine learning

CO2: Design ML models to make predictions by using linear, non-linear and logistic regression techniques

CO3: Implement classification models for two class problems and multiclass problems

CO4: Implement clustering models for unlabeled data

CO5: Integrate multiple machine learning algorithms in the form of ensemble learning

CO6: Apply reinforcement learning and its algorithms for different applications

Course Contents

Unit I	Introduction to Machine Learning	06 Hours

Introduction: What is Machine Learning, Definitions and Real-life applications, Comparison of Machine learning with traditional programming, ML vs AI vs Data Science.
Learning Paradigms: Learning Tasks- Descriptive and Predictive Tasks, Supervised, Unsupervised, Semi-supervised and Reinforcement Learnings.
Models of Machine learning: Geometric model, Probabilistic Models, Logical Models, Grouping and grading models, Parametric and non-parametric models.
Feature Transformation: Dimensionality reduction techniques- PCA and LDA

#Exemplar/Case Studies	**Explore the machine learning paradigms with its application:** This case study is about exploring three different machine learning paradigms that help to solve different problem categories in plain language and from a technical standpoint. **Reference URL:** https://www.analyticsvidhya.com/blog/2022/07/machine-learning-paradigms-with-example/
***Mapping of Course Outcomes for Unit I**	CO1

Unit II	Regression	06 Hours

Introduction- Regression, Need of Regression, Difference between Regression and Correlation, Types of Regression: Univariate vs. Multivariate, Linear vs. Nonlinear, Simple Linear vs. Multiple Linear, Bias-Variance tradeoff, Overfitting and Underfitting. **Regression Techniques -** Polynomial Regression, Stepwise Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression, Ridge Regression, Lasso Regression, ElasticNet Regression, Bayesian Linear Regression. **Evaluation Metrics:** Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE),R-squared ,Adjusted R-squared.	
#Exemplar/Case Studies	**Comparison of different regression models:** Build and compare the Lasso, Ridge, and Elastic Net regression models, consider the big market sales to predict sales depending on features selected. This case study discusses regression models and how they can be used to solve prediction problems. **Reference URL:** https://www.analyticsvidhya.com/blog/2017/06/a-comprehensive-guide-for-linear-ridge-and-lasso-regression/
***Mapping of Course Outcomes for Unit II**	CO2

Unit III	**Classification**	**06 Hours**

Introduction: Need of Classification, Types of Classification (Binary and Multiclass), Binary-vs-Multiclass Classification, Balanced and Imbalanced Classification Problems. **Binary Classification:** Linear Classification model, Performance Evaluation- Confusion Matrix, Accuracy, Precision, Recall, F measures. **Multiclass Classification:** One-vs-One and One-vs-All classification techniques, Performance Evaluation- Confusion Matrix, Per Class Precision, Per Class Recall **Classification Algorithms:** K Nearest Neighbor, Linear Support Vector Machines (SVM) – Introduction, Soft Margin SVM, Kernel functions– Radial Basis Kernel, Gaussian, Polynomial, Sigmoid.	
#Exemplar/Case Studies	**Explore Multiclass Classification with imbalanced dataset:** This case study uses a "20 Newsgroups" data set that is converted into an imbalanced form. A multiclass classification algorithm is applied on an imbalanced dataset and its performance is compared with the model after applying undersampling/oversampling techniques. **Reference URL:** https://builtin.com/machine-learning/multiclass-classification
***Mapping of Course Outcomes for Unit III**	CO3

Unit IV	**Clustering**	**06 Hours**

Introduction: What is clustering, Need of Clustering, Types of Clustering **Hierarchical clustering algorithms /connectivity-based clustering):** Agglomerative Hierarchical Clustering (AHC) algorithm, Divisive Hierarchical Clustering (DHC) algorithm. **Centroid-based clustering algorithms / Partitioning clustering algorithms:** K-Means clustering algorithm, Advantages and disadvantages of K-Means clustering algorithm, Elbow method, The Silhouette method, K-Medoids, K-Prototype. **Density-based clustering algorithms:** DBSCAN algorithm, how it works, Advantages and disadvantages of DBSCAN. **Distribution-based clustering algorithms:** Gaussian mixture model. **Application of Clustering Technique:** Market Segmentation, Statistical data analysis, Social network analysis, Image segmentation, Anomaly detection.

#Exemplar/Case Studies	**Customer segmentation using clustering algorithms:** This case study demonstrates the concept of segmentation of a customer data set from an e-commerce site using k-means clustering in python. The data set contains the annual income of ~300 customers and their annual spend on an e-commerce site. The k-means clustering algorithm is applied to derive the optimum number of clusters and understand the underlying customer segments based on the data provided. **Reference URL:** https://towardsdatascience.com/clustering-algorithms-for-customer-segmentation-af637c6830ac
***Mapping of Course Outcomes for Unit IV**	CO4

Unit V	**Ensemble Learning**	**06 Hours**

Ensemble Learning: Introduction to Ensemble Learning, Need of Ensemble Learning, Homogeneous and Heterogeneous ensemble methods, Advantages and Limitations of Ensemble methods, Applications of Ensemble Learning.

Basic Ensemble Learning Techniques: Voting Ensemble, Types of Voting: Max Voting, Averaging, Weighted Average.

Advanced Ensemble Learning Techniques:

Bagging: Bootstrapping, Aggregation.

Boosting: Adaptive Boosting (AdaBoost), Gradient Boosting, XGBoost .

Stacking: Variance Reduction, Blending, Random Forest Ensemble, Advantages of Random Forest.

#Exemplar/Case Studies	**Apply ensemble learning techniques:** This case study uses ensemble learning techniques on the Heart Attack dataset. It indicates that ensemble techniques, such as bagging and boosting, are effective in improving the prediction accuracy of weak classifiers and exhibit satisfactory performance in identifying risk of heart disease. **Reference URL:** https://www.sciencedirect.com/science/article/pii/S235291481830217X?via%3Dihub
***Mapping of Course Outcomes for Unit V**	CO5

Unit VI	**Reinforcement Learning**	**06 Hours**

Reinforcement learning: What is Reinforcement Learning? Need for Reinforcement Learning, Supervised vs Unsupervised vs Reinforcement Learning, Types of Reinforcement, Elements of Reinforcement Learning, Real time applications of Reinforcement learning.

Markov's Decision Process: Markov property, Markov chain/process, Markov reward process (MRP), Markov decision process (MDP), Return, Policy, Value functions, Bellman equation

Q Learning: Introduction of Q-Learning, Important terms in Q learning, Q table, Q functions, Q learning algorithm.

#Exemplar/Case Studies	**Implement Tic Tac Toe Game using reinforcement Learning:** The case study explores the implementation of reinforcement learning techniques to create an agent capable of playing Tic-Tac-Toe. It discusses the use of Q-learning and the construction of a reward system to train the agent, resulting in a player that can learn and improve its gameplay over time. **Reference URL:**https://towardsdatascience.com/reinforcement-learning-implement-tictactoe-189582bea542

*Mapping of Course Outcomes for Unit VI	CO6

Learning Resources

Text Books:

1. Ethem Alpaydin, "Introduction to Machine Learning", Publisher: The MIT Press,2014
2. Peter Flach: "Machine Learning: The Art and Science of Algorithms that Make Sense of Data", Cambridge University Press, Edition 2012

Reference Books:

1. Ian H Witten, Eibe Frank, Mark A Hall, "Data Mining, Practical Machine Learning Tools and Techniques", Elsevier, 3rd Edition
2. Jiawei Han, Micheline Kamber, and Jian Pie, "Data Mining: Concepts and Techniques", Elsevier Publishers Third Edition, ISBN: 9780123814791, 9780123814807
3. Shalev-Shwartz, Shai, and Shai Ben-David, "Understanding machine learning: From theory to algorithms", Cambridge university press, 2014
4. McKinney, "Python for Data Analysis O' Reilly media, ISBN : 978-1-449- 31979-3

e-Resources:

1. https://timeseriesreasoning.com/
2. Reinforcement Learning: https://www.cs.toronto.edu/~urtasun/courses/CSC411_Fall16/19_rl.pdf
3. A brief introduction to machine learning for Engineers: https://arxiv.org/pdf/1709.02840.pdf
4. Introductory Machine Learning Nodes: http://lcsl.mit.edu/courses/ml/1718/MLNotes.pdf

MOOC Courses:

1. Introduction to Machine Learning(IIT kharagpur) : https://nptel.ac.in/courses/106105152
2. Introduction to Machine Learning (IIT Madras): https://onlinecourses.nptel.ac.in/noc22_cs29/prevew
3. Machine Learning A-Z™: AI, Python & R + ChatGPT Bonus [2023] https://www.udemy.com/course/machinelearning/
4. Machine Learning and Deep Learning A-Z: Hands-On Python https://www.udemy.com/course/machine-learning-and-deep-learning-a-z-hands-on-python/

The CO-PO Mapping Matrix

CO/ PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	1	-	-	-	-	-	-	1	-	-	1
CO2	3	3	3	2	3	-	-	-	1	-	-	1
CO3	3	3	3	2	3	-	-	-	1	-	-	1
CO4	3	3	3	2	3	-	-	-	1	-	-	1
CO5	3	3	3	2	3	-	-	-	1	-	-	1
CO6	3	3	3	2	3	-	-	-	1	-	-	1