Meghnad Desai Academy of Economics

ELECTIVE COURSE

Course Title:
MACHINE LEARNING

Course Description:

This course provides a comprehensive introduction to the field of machine learning. It covers the fundamental concepts and techniques, including both supervised and unsupervised learning. Students will learn about various algorithms such as linear regression, logistic regression, decision trees, neural networks, clustering, and more. The course emphasizes understanding the theory behind these methods, as well as their practical application. Students will gain hands-on experience through programming assignments and projects, using popular machine learning libraries and tools.

Instructor Information:
Name: Dr. Neha Ansari
Contact Information: nehaansari20jun@gmail.com

Course Objectives:
- To understand the concepts and applications of machine learning
- To understand various types of learnings – supervised and unsupervised
- To understand and be able to implement regression and classification models
- To understand and be able to implement the unsupervised models
- To help them make sense of the present-day applications of AI and the prospects in future
- To prepare the students for higher studies and research in machine learning and related fields

Course Content:
1. Introduction to Machine Learning
   - Introduction to Supervised and Unsupervised learning
   - Regression vs Classification
   - Predictive vs. Descriptive Analysis
   - Machine Learning Terminologies

2. Linear Regression
   - What is regression analysis
   - Model & parameters in Simple Linear Regression & Multiple Linear Regression
   - Use Cases
   - OLS Technique based derivation of parameters
   - Assumptions
   - Evaluation metrics
   - Simple Linear Regression model building and evaluation (case studies)
   - Multiple Linear Regression model building and evaluation (case studies)

ELECTIVE COURSE

3. Logistic Regression
   - Intuitive understating of Logistic Regression
   - Logistic Regression Modelling
   - Sigmoidal Function Derivation
   - Logistic Regression for binary classification
   - Logistic Regression for multiclass classification
   - Evaluation metrics
   - Model Optimization
   - Use Cases on when should and shouldn't model be used

4. Data Preprocessing, Transformation & Normalization EDA (with & without visualization)
   - Data Scaling
   - Data Type Transformations
   - Handling outliers, missing values
   - EDA (with & without visualization)

5. Cross-Validation
   - Why do we need
   - KFold
   - HoldOut
   - Stratified KFold
   - Repeated Stratified KFold
   - Fine-Tuning hyperparameters

6. Handling overfitting with regularization
   - Ridge regression
   - Lasso regression
   - Selecting the Tuning Parameter
   - GridSearch CV to select the optimum hyperparameters
   - Use cases where they are and where they aren't needed

7. K Nearest Neighbours Algorithm
   - Intuitive understanding of KNN
   - Computation of Distance Matrix
   - The Optimum K-value
   - GridSearch CV to select the optimum hyperparameters
   - Evaluation of Model
   - Lazy Learning Notion
   - Advantages & Disadvantages of KNN Models

8. Naïve Bayes Algorithm

ELECTIVE COURSE

- Intuitive understanding of Naïve Bayes
  - Dependent and Independent Events
  - Bayesian Theorem
  - Probabilities - The Prior and Posterior Probabilities
- Model Building and Validation
- Evaluation of Model
- Classification Report
  - Precision,
  - Recall,
  - F1score
  - Support
- Derivation from Bayes Theorem
- NLP based case studies
- Advantages & Disadvantages of the model

9. Decision Tree
   - Intuitive Understanding to Trees Algorithms
   - Classification & Regression Trees
   - Splitting of the Tree - Gini Index and Entropy
   - Overfitting of the Trees
   - Model Building and Validation - Classification Tree
   - Types of Decision Tree algorithms, what are decision stumps
   - Evaluation of Model
   - Advantages & Disadvantages of the model

10. Random Forests
    - Understanding the Concept of Bootstrapping and Bagging
    - Model Building and Validation - Classification Case
    - Evaluation of Model
    - Practical Implementation for both classification and regression
    - Ensemble Techniques
      - Bagging Algorithms
      - Boosting Algorithms
    - Advantages & Disadvantages of the model
      - The Trade-Off between Prediction Accuracy and Model Interpretability.
      - Black box and white box

11. Support Vector Machines
    - Intuitive understanding of SVM
      - The Maximal Margin Classifier
      - Hyperplane
    - Support Vector Classifiers - Overview
    - Support Vector Machines

ELECTIVE COURSE

  - Concepts of Linearly separable vs non separable data
  - The Kernel Method in SVM
  - Understanding the hyperparameters- Cost and Gamma
- GridSearch CV - to select the optimum hyperparameters
- Model Building and Validation - Classification Case
- Evaluation of Model
- Advantages & Disadvantages of the model

12. Unsupervised Learning
- Understanding The Challenge of Unsupervised Learning
- Principal Components Analysis
  - What Are Principal Components?
  - Interpretation of Principal Components
  - Application of PCA
- Practical Implementation with visualization
- Definition of clustering
- Understanding working of K-means algorithm
- Cluster-size optimization - Elbow method
- Model Building and Validation
- Evaluation of Model
- Practical Issues in Clustering

13. Discussion on Application of Algorithms
- With respect to
  - Variable type
  - Accuracy vs Speed
- Prediction vs Approximation
- Prediction vs Inference

References:

- An Introduction to Statistical Learning, 2nd Edition - https://www.statlearning.com/

This textbook is available to you as reference texts. We won't necessarily be following it. We will not be following a textbook based learning approach. This is an application-based subject.