Data Mining Beginner to Intermediate

Course Outline: Introduction to Machine Learning

* To introduce participants to the basic concepts and techniques of Machine Learning.
* To develop skills of using recent machine learning software for solving practical problems.
* To gain experience of doing independent study and research.

Course Description

According to Tom Mitchell “The field of Machine Learning is concerned with the question of how to construct computer programs that automatically improve with experience”. This course covers the basic concepts and techniques of Machine Learning from both theoretical and practical perspective. The material includes classical ML approaches such as Linear Regression and Decision Trees, more advanced approaches as Clustering and Association Rules as well as “hot” topics such as XGBoost.

Outline

Module 1 Introduction

* What is Machine Learning
* Use Cases
* Commonly used Terms
* Lifecycle of a ML Project
* Supervised Learning
* Unsupervised Learning
* Summary
* Quiz

Module 2 Data Exploration

* Data Acquisition
* Types of Data
* Data Types
* Exploratory Data Analysis
* Data Pre-processing
* Data Quality assessment
* Feature Scaling
* Descriptive Statistics
* Methods to impute missing values
* Outlier/Anomaly Detection
* Data Visualization
* Histogram
* Bar Graph
* Scatter Plot
* Pie Chart
* Box Plot
* Feature Selection
* Univariate Selection
* Feature Importance
* Correlation matrix and Heat map
* Underfitting vs Overfitting
* Bias-Variance Trade-off
* Summary
* Quiz

Module 3 Evaluation Metrics

* Introduction
* Hypothesis Testing
* Statistical Assumptions
* Null Hypothesis
* Alternate Hypothesis
* One sample Z-test
* Z-test in Python
* T-test
* T-test in Python
* Pearson’s Chi Squared Test
* Confusion Matrix
* Absolute Error
* Relative Error
* RMSE
* Precision, Accuracy
* Recall
* Specificity
* F-Score
* ROC/AUC
* Summary
* Quiz

Module 4 Linear Regression

* Introduction
* Cost Function
* Gradient Descent
* What is Regression
* Basic Idea
* Linear Regression Applications
* Linear Regression
* Types of Errors
* Better Regression Models
* Correlation is not Causation
* Polynomial Linear Regression
* Regularization
* Ridge Regression
* LASSO Regression
* Summary
* Quiz

Module 5 Classification

* Introduction
* Types of Classification Algorithms
* Applications of Classification Algorithms
* Logistic Function
* Logistic Regression
* Application of Logistic Regression
* Types of Logistic Regression
* Decision Trees
* Working of Decision Tree
* Attribute Selection measure
* Gini Index
* Information Gain
* Random Forests
* Working of RF
* Advantages and Disadvantages of RF algorithm
* Application of RF
* XGBoost
* Summary
* Quiz

Module 6 Unsupervised Machine Learning

* Why Unsupervised Learning
* Applications
* Clustering
* Types of Clustering
* Singular value Decomposition
* Independent Component Analysis
* Association Rules
* Summary
* Quiz