**APPLIED MACHINE LEARNING**

**Course overview**

Machine Learning (ML) is the interdisciplinary field at the intersection of mathematics, statistics, and computer science and engineering that is a subfield of Artificial Intelligence (AI). Currently on the daily basis in the world, the enormous amount of data being produced from the various sources. So the aim of the ML is to make use of different data and draw meaning conclusions. ML methods achieve that by using statistical models and interweaves them with computer algorithms. It underpins many modern technologies, such as speech recognition, Internet search, bioinformatics and computer vision, recommender system like Amazon’s, Google’s driverless car and the most recent imaging systems for cancer diagnosis are all based on Machine Learning and Artificial Intelligence technologies. The course will provide a solid background to Applied Machine Learning and its core models and algorithms. The aim of the course is to provide students with detailed knowledge of how Machine Learning methods work and how statistical models can be brought to bear in computer systems not only to analyze large data sets, but to let computers perform tasks that traditional methods of computer science are unable to address. Examples range from speech recognition and text analysis through bioinformatics and medical diagnosis. This course provides hands on concepts on real data sets for performing classifications, predictive analytics, clustering or analyzing associations. Famous and solid frameworks like scikit-learn, tensorflow and keras are going to be used during the course.

*Subject to change and improve*

**Pre-requisites**

1. Including the list of topics that student must know to understand the course

|  |  |  |
| --- | --- | --- |
| No | Course Code | Course Title |
| 1 |  | Algorithms and Programming – I/II |
| 2 |  | Data Structures and Algorithms |
| 3 |  | Calculus – I/II |
| 4 |  | Linear Algebra |
| 5 |  | Probability and Statistics |
| 6 |  | Programming Languages |

**Curriculum Plan (or Course Content or Weekly Distribution of Themes)**

|  |  |  |
| --- | --- | --- |
| **WEEK** | **Lecture’s title** | **Practice works** |
| 1 | * Overview of the course and regulations * Intro to ML (What is Machine Learning [ML] and Artificial Intelligence [AI]). Concept of Big-data. Rule-based methods and drawbacks * Applications and Types (supervised, unsupervised, reinforcement, deep learning?) of ML. * Four main problems of ML (Classification, prediction (regression), clustering and Association) with examples * Introduction to the tools used (Python, scikit-learn, tensorflow, keras, etc.,) | Implementation of rule-based model for the weather dataset by using Python PL |
| 2 | * Supervised Learning * Features engineering (data wrangling, pre-processing and features selections) * Correlation and covariance * Linear methods for classification and regression * Regression analysis (multivariate regression analysis). Gradient descent algorithm * Building regression model with scikit-learn framework * MS Excel Data Toolbox and Regression Analysis | Implementation of regression model over Boston house prices dataset by using scikit-learn  Cater size prediction for heart diseases patients with MS Excel Data Toolbox |
| 3 | * KNN ml algorithm and distance measures (Euclidean, Manhattan etc.,) * Decision trees algorithm (Information gain, Entropy, Gini index) * Overfitting and underfitting (pruning) | Implementation of classification models over Diabetes patient dataset by using scikit-learn |
| 4 | * Naïve-base algorithm (likelihood estimation) * Bayes theorem (prior-posterior probabilities [breast cancer problem and solution] ) | Under development |
| 5 | * Logistic regression (single perceptron) * Artificial Neural Networks (practical example with logical gates) * Activation functions | Implementation of classification models over Breast cancer dataset by using scikit-learn and keras |
| 6 | * Data Transformations * Model evaluation techniques (cross-validation [hold-out, k-fold, random sampling, bootstrap method]) * Model Accuracy Evaluation (confusion matrix table, accuracy, sensitivity (recall), specificity, precision, F-measure, G-measure, AUC (roc curve)) * Error metrics (MSE-RMSE, MAE etc.,) * Bias-Variance tradeoff. Regularization and model/feature selection | Under development |
| 7 | * Support Vector Machines * Kernels (Gaussian RBF, Sigmoidal, Polynomial) * Model Selection and Boosting (GridSearch) | Under development |
| 8 | * Ensemble Learner Models * Bootstrap aggregating (Bagging and Boosting methods) * Ada-Boost algorithm * Random-Forest algorithm | Under development |
| 9 | * Unsupervised Learning (clustering, dimensionality reduction, association rules) * K-means clustering (k-means plus). Elbow method (WCSS) * Hierarchical clustering (agglomerative and divisible) * Applications of clustering algorithms (bioinformatics) | Under development |
| 10 | * Association rules * Apriori algorithm (support, confidence, lift) * Eclat algorithm * Applications of association rules * Recommendation systems | Under development |
| 11 | * Factor Analysis * Principal Component Analysis * Linear Discriminant Analysis | Under development |
| 12 | * Introduction to Deep Learning (Convolutional Neural Networks - CNN) * Natural Language Processing and Sentimental Analysis * Time series analysis and ARIMA prediction model | Under development |

**Course Materials**

***1. Machine Learning by Tom M Mitchell***

This textbook provides a single source introduction to the primary approaches to machine learning. It is intended for advanced undergraduate and graduate students, as well as for developers and researchers in the field. No prior background in artificial intelligence or statistics is assumed. Several key algorithms, example date sets and project- oriented home work assignments discussed in the book are accessible through the Internet. Book contains more theoretical knowledge.

***2. Learning from Data by Yaser Abu Mostafa etal.,***

This book provides a perfect introduction to machine learning and various famous algorithms. This book prepares you to understand complex areas of machine learning. It has been prepared according to the famous ML and AI course taught at Caltech.

***3. Machine Learning for Absolute Beginners by Oliver Theobald***

This book is a simplified introduction to machine learning for beginners, it gives learner a fair idea of what, why and how of machine learning and ends with a practical examples in python. Provides various examples using python to build and test particular ML model in action. The latest version of the book covers topics like cross validation, ensemble modeling, grid search, feature engineering etc.

***4. Hands On Machine Learning with Scikit-Learn and TensorFlow(O’reiley)***

The goal of this book is to give a reader the concepts, the intuitions, and the tools needed to actually implement programs capable of learning from data.

It covers a large number of techniques, from the simplest and most commonly used (such as linear regression) to some of the Deep Learning techniques that regularly win competitions. Rather than implementing our own toy versions of each algorithm, we will be using actual production-ready Python frameworks like:

*Scikit-Learn*

Scikit-Learn is very easy to use, yet it implements many Machine Learning algorithms efficiently, so it makes for a great entry point to learn Machine Learning.

*TensorFlow*

TensorFlow is a more complex library for distributed numerical computation using data flow graphs. It makes it possible to train and run very large neural networks efficiently by distributing the computations across potentially thousands of multi-GPU servers. TensorFlow was created at Google and supports many of their large-scale Machine Learning applications. It was open-sourced in November 2015.

***5. Python Machine Learning By Example by Yuxi (Hayden) Liu***

This book starts with an introduction to machine learning and the python language and shows how to complete the setup. The readers will learn all the important concepts such as, exploratory data analysis, data pre-processing, feature extraction, data visualization and clustering, classification, regression and model performance evaluation.

**Lab work details**

Under development

**Capstone Project**

Under development