CSE602: Machine Learning - I

Credit hours structure	(3,0,3) = (Theory, Lab, Total)
Prerequisites	See the note below
Co-requisites	None
Barred combinations	N/A
Course Lead / E-mail	Dr. Sajjad Haider (sahaider@iba.edu.pk)
Last Modified	January 15, 2023

Course Description

This course provides a comprehensive introduction to the field of machine learning, with a focus on supervised learning techniques. Students will learn about the mathematical foundations of key algorithms, as well as how to evaluate and select the most appropriate method for a given problem. The course covers a range of techniques for addressing classification, regression, and time-series problems, including both parametric and non-parametric methods. Topics like Bias-Variance tradeoff, over-fitting vs underfitting, feature reduction, loss functions, and training optimization are discussed in sufficient detail. Students are also introduced to automated machine learning, deep learning and interpretable machine learning. The course emphasizes the practical application of these methods, with students working on real-world problems using the Python programming language.

Learning Outcomes

By the end of this course, students are expected to:

1) understand and apply mathematical concepts and algorithms used in supervised machine learning, including classification, regression, and time-series problems
2) evaluate and select the most appropriate machine learning method for a given problem, taking into account factors such as the bias-variance tradeoff, overfitting and underfitting, feature reduction, and loss functions
3) participate in data analytics competitions held at Kaggle.com and other sites and have proficiency in doing machine learning tasks using the Python language
4) understand the complete life cycle of building machine learning systems that include data pre-preparation, experimental design, and model selection, evaluation and interpretation

Prerequisites:

Students are expected to be proficient in Python programming. In addition, a decent understanding and recollection of undergraduate-level statistics, linear algebra and algorithm concepts is assumed. It is expected that MS(Data Science) students coming from non-CS background have cleared the courses of “Algorithm” and “Python Programming”.

(Tentative) Weekly Plan

Weeks	Contents
1	Course Overview, Applications of Machine Learning, Introduction to Regression Analysis, Assessing Model Accuracy
2	Multiple Linear Regression, Polynomial and Interaction Models, Variable Selection, k-Nearest Neighbor Regression
3	Data and Feature Engineering, , Cross-Validation, Feature Selection (Forward, Backward and Step-wise), Model Assessment (AIC, BIC and Adjusted R2)
4	Regularization: Ridge and Lasso, Grid Search, Case Studies
5	Elastic Net, Regression Tree, Variance-Bias Tradeoff
6	Ensemble Models: Bagging, Boosting, Random Forest,
7	Huber Loss Regression, Feature Importance using Ensemble, Meta-Heuristics: Simulated Annealing and Evolutionary Algorithms
8	Classification:k-Nearest Neighbor, Bayesian Methods, Model Evaluation Metrics
9	Decision Tree (Gini Index and Gain Ratio), ROC Curve, Case Study
10	Boosting (Adaboost and Gradient Boosting), Categorical and Gaussian Naïve Bayes
11	Logistic Regression, Time-Series Handling, Regression-based Forecasting, Smoothing
12	Introduction to Netural Networks (Architecture, Weight Updating, Activation Functions, Case vs Batch Processing, Dropout and Regularization, Loss Functions)
13	Decision Boundaries Learned by Neural Networks, Interpretable and Explainable Machine Learning, AutoML, Course Recap
14	Project Presentation

Reference Books

An Introduction to Statistical Learning (2nd Edition) by James et al. (2020)
Introduction to Machine Learning with Python by Muller and Guido (2017)
Data Mining for Business Analytics in Python by Shmueli et al. (2020)
Interpretable Machine Learning by Molnar (2019)

Marks Distribution

Midterms - 25%
Final - 40%
Project - 15%
Assignments - 14%
Quizzes - 6%