DATA SCI 8010: Data Analytics from Applied Machine Learning

Synopsis

This course leverages the foundations in statistics and modeling to teach applied concepts in machine learning (ML). Participants will learn various classes of machine learning and modeling techniques, and gain an in-depth understanding how to select appropriate techniques for various data science tasks. Topics cover a spectrum from simple Bayesian modeling to more advanced algorithms such as support vector machines, decision trees/ forests, and neural networks. Students learn to incorporate machine learning workflows into data-intensive analytical processes.

Credit hours: 3
Num of modules: 8
Duration: 8 weeks

Course Prerequisites

DATA_SCI 7020: Statistical and Mathematical Foundations for Data Analytics or instructor's consent

Course Instructor

Tozammel Hossain (hossaink@missouri.edu) Office Hours: TBA

Textbook

No textbook is required. Students are expected to have access to personal computing resources capable of accessing web resources and Jupyter Notebook environments. Course contents are primarily delivered using Jupyter notebooks.

Meeting Schedule

There are two sections for this course: online and on-campus. For both sections the flow of this course uses asynchronous learning activities.This course will have regular weekly readings, labs, discussions, and assignments. Course contents will be released each Sat morning and the assignments are due by the following Sat.

For the on-campus sections, there are two weekly synchronous sessions conducted on Tue and Thu from 9:30 AM to 10:45 PM. For the online sections, there are weekly office hours conducted each Wed evening. Although this course follows an asynchronous learning strategy, students will have access to instructor and TA support via Slack or Microsoft Teams.

Evaluation

Practices: 10%
Exercises: 60%
Project: 30%

Course Schedule

Module 1: Introduction to Machine Learning & Supervised Learning

- Introduction to ML
- ML training and validation
- Supervised ML: Decision Tree

Module 2: Supervised Learning: Classification & Regression

- Cross Validation
- Classificaiton: Naive Bayes & Logistic Regression
- Regression: Linear Regression

Module 3: Unsupervised Learning: Feature Selection & Dimensionality Reduction

- Feature Selection
    - Forward Selection
    - Backward Elimination
- Dimensionality Reduction
    - PCA
    - Factor Analysis

Module 4: Unsupervised Learning: Clustering & Anomaly Detection

- Clustering
    - K-means
    - Hierarchical Clustering
    - DBSCAN Clustering
- Anomaly Detection

Module 6: Operationalizing a Machine Learning Pipeline - I

- Mini-project part I
- Data Exploration
- Sampling
- Pipelines & Hyperparameter Search Comparative Analysis

Module 7: Operationalizing a Machine Learning Pipeline - II

- Mini-project part II

- Full Data Training
- Evaluation
- Business Case Analysis

Module 8: Neural Networks

- Perceptron/Neuron
- Feed Forward
- Back Propagation