

DATA SCI 8010: Applied Machine Learning

Tozammel Hossain



Data Science & Analytics
University of Missouri

Learning Objectives

- **To illustrate the basic methodology of machine learning, beginning from a raw dataset to prediction or latent structure learning**
- **Operationalize ML Models**
- **Hands-on practices to help you identify recurring themes of data analytics in different problem domains**

Course structure

- **We aim to employ a “flipped classroom” model**
 - A module(video lectures, slides, coding practices) will be posted online **each sat at 6:00 am** and you are expected to digest this information before class (on-campus)/office hours (online)
 - Assignment due by the **following Sat 11:59 PM**
 - Class times will be used to reinforce these concepts and for exploratory discussions
- **Your active participation is important for the success of this model!**

Prerequisites

- **Data Sci 7020: Stat & Math Foundations for Data Science**
 - Or instructor's consent
- **Relevant Computing Skills**
 - Python Programming
 - Using Python package (sklearn) to learn machine learning
 - Familiarity with DSA computing environment + Git
 - Please check module 0
 - If you need help on Python, DSA environment and Git, please send me an email

Tools/Services

- **Canvas**
 - Announcements + grades
- **Jupyterhub**
 - Computing environment
- **Git**
 - Committing code
- **Communication**
 - Slack (online cohort)
 - Teams (on-campus cohort)
 - Office hours

Office Hours

- **Only for online cohort**
 - Wed, 8:30 PM – 9:30 PM
- **On-campus**
 - In-person meeting
 - Location: Naka Hall 116
 - Zoom option will also be available
 - Two sessions per week
 - Tue: 9:30 AM –10:45 AM
 - Thu: 9:30 AM –10:45 AM

Course Structure

- **8 modules**
 - delivered over 8 weeks
 - 6 modules for concepts + 2 modules for final project
- **No textbook is required**
- **Each module has**
 - Readings
 - Labs (examples done within the script)
 - Practices (solutions are provided separately)
 - Grades for completion (2 points)
 - Exercises
 - Graded for accuracy (18 points)
- **Score distribution**
 - Labs (10%) + Exercises (60%) + Project (30%)
 - Project problem is given
 - Need to develop individually

Schedule

Module	Release Date	Assign. Due	Comment
M1	Mar 13	Mar 20	
M2	Mar 20	Mar 03 (on-campus) Mar 10 (online)	M2 + M3
Break	Mar 28	Apr 02	
M3	Apr 03	Mar 10	M2 + M3
M4	Apr 10	Mar 17	
M5	Apr 17	Mar 24	
M6+M7	Mar 24	May 01	M6 due
		May 08	M7 due
M8	May 08	May 15	

Road Map

- **Module 1: Supervised Learning**
 - Classification: naïve Bayes
 - Model selection: train/test
- **Module 2: Supervised Learning**
 - Classification: tree methods
 - Regression: tree methods + linear method
- **Module 3: Unsupervised Learning**
 - Feature selection
 - Dimensionality reduction

Road Map

- **Module 4: Unsupervised Learning**
 - Clustering
 - Anomaly detection
- **Module 5: Developing ML Pipeline**
 - Pipeline
 - Classification: svm
- **Module 6: Project part I**
 - Using the concepts from M1-M5, solve a classification problem

Road Map

- **Module 7: Project part II**
 - Using the concepts from M1-M5, solve a classification problem
- **Module 8: Intro to Neural Network**
 - Basic concepts in neural net