DATA SCI 8010: Applied Machine Learning

Tozammel Hossain

Data Science & Analytics
University of Missouri

Basic Info

- **3 Credit Hours**
- **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

Learning Objectives

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

Course structure

- **We aim to employ a "flipped classroom"/ mostly asynchronous 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 is due by the **following Sat 11:59 PM**
  - Class times/Office hours will be used to reinforce these concepts and for exploratory discussions
- **Your active participation is important for the success of this model!**

Tools/Services

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

Office Hours

- **Only for online cohort**
  - Wed/Thu/Fri, 7:00 PM – 8:00 PM
- **On-campus**
  - Two in-person sessions per week
    - Location: Naka Hall 355
    - Time
      - Tue: 9:30 AM –10:45 AM
      - Thu: 9:30 AM –10:45 AM
  - Office hours
    - Wed/Fri: 7:00 PM – 8:00 PM

Module Layout

- **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**
  - Practices (10%) + Exercises (60%) + Project (30%)
  - Project: a problem will be given
    - Need to develop individually

Schedule: On-campus

Module	Release Date	Assign. Due	Comment
M1	Mar 12	Mar 19	
M2	Mar 19	Apr 02	
Spring Break: Mar 26 - Apr 02			
M3	Mar 26	Apr 09	
M4	Apr 09	Apr 16	
M5	Apr 16	Apr 23	
M6+M7 (FP)	Mar 23	Apr 30	M6 check-in
		May 07	M6+M7 due
M8	May 07	May 13	

Schedule: Online

Module	Release Date	Assign. Due	Comment
M1	Mar 12	Mar 19	
M2	Mar 19	Apr 09	
Spring Break: Mar 26 - Apr 02			
M3	Mar 26	Apr 09	M2+M3 due
M4	Apr 09	Apr 16	
M5	Apr 16	Apr 23	
M6+M7 (FP)	Mar 23	Apr 30	M6 check-in
		May 07	M6+M7 due
M8	May 07	May 13	

Road Map

- **Module 1: Supervised Learning**
  - Classification: decision tree
  - Model Evaluation: train/test
- **Module 2: Supervised Learning**
  - Classification: naïve Bayes
  - Regression: tree methods + linear method
  - Model Evaluation: cross validation
- **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