THE UNIVERSITY OF CHICAGO

Master of Science in Analytics

Course Syllabus

Machine Learning and Predictive Analytics

MSCA 31009
Wed 6 – 9pm CST
Spring 2021
Mar 29 to Jun 12, 2021
Location: remote

Arnab Bose, Ph.D.
630-258-0761
abose@uchicago.edu

COURSE DESCRIPTION

The objective of this course is three-folds – first, to understand machine learning concepts and methodologies. Second, to develop the ability to apply those concepts and methodologies to diverse practical applications, evaluate the results and recommend next best action. Third, to outline recent machine learning and deep learning research and development and their applications.

This course focuses on different types of learning and methodologies using Python. It also includes different data characteristics and how they affect model learning. The course covers industry applications via case studies, assignments and a class project.

PREREQUISITES

Required: Students need to have taken Data Mining and Advanced Linear Algebra for Machine Learning. Students need to be familiar and comfortable with Python software for data manipulation and model development.

COURSE MATERIALS

Most of the course materials are class notes derived from different Machine Learning books and articles. Textbook chapters are recommended for reading along with class notes.

Class textbooks are:

[**DL**] Goodfellow I., Bengio Y. and Courville A., *Deep Learning*, MIT Press, 2016.

[**HML**] Geron A., *Hands-On Machine Learning with Scikit-Learn & TensorFlow*, 2nd Edition, O'Reilly Media, 2019.

[**ISL**] James G., Witten D., Hastie T. and Tibshirani R., *An Introduction to Statistical Learning*, Springer, 2013.

Optional reference books are:

[**DLP**] Francois Chollet, *Deep Learning with Python*, Manning Publication, 2018.

[**NNDL**] Michael Nielsen, *Neural Networks and Deep Learning*, Online book

(http://neuralnetworksanddeeplearning.com/), 2016.

[**PRM**] Bishop C.M., *Pattern Recognition and Machine Learning*, Springer, 2006.

Additionally, students are given supplemental papers and case studies. These cover topics such as applied research and applications not covered in the class. The purpose is to illustrate practical applications of different machine learning techniques and concepts.

Additional Reading Materials

1. Nassim N. Taleb, *Fooled by Randomness*, Penguin Books, 2004.
2. Nassim N. Taleb, *Antifragile*, Random House, 2012.
3. Garry Kasparov, *Deep Thinking*, John Murray, 2017.
4. Max Tegmark, *Life 3.0*, Penguin, 2017.
5. Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, 2015.
6. Nick Bostrom, *Superintelligence: Path, Dangers, Strategies*, Oxford University Press, 2014.
7. Ray Kurzweil, *How to Create a Mind: The Secret of Human Though Revealed*, Viking, 2012.
8. David Epstein, *Range: Why Generalists Triumph in a Specialized World*, Riverhead Books, 2019.
9. Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, 2018.
10. Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Broadway Books, 2017.

SOFTWARE

The statistical software package Python version 3.6 or higher is to be used for this course. It is freely available from http://www.python.org/.
Recommend Jupyter Notebook (http://jupyter.org/) for Python code development.

LEARNING OBJECTIVES

After completing this course, students are be able to:

- exhibit a strong fundamental understanding of machine learning and deep learning
- comfortably manage industry analytics projects using machine learning techniques
- separate media noise on AI from research & development and their practical applications
- understand the difference between general intelligence, human intelligence and machine intelligence

EVALUATION

Students are evaluated as follows. There are 8 assignments, each posted on canvas with a due date. The assignment problems are aimed at testing the students' understanding of the materials presented in the class.

There are 8 quizzes held during class hours. The best 7 out of 8 count towards grade. Quizzes are multiple choice questions based on the materials covered in class till that point.

There is a class project that is done by groups of students. Depending on the class size, each group consists of 3 – 4 students depending on class size. I will make an announcement on team size after the 2nd week of class. The purpose of the project is to test the students' ability to apply the machine learning techniques and tools presented in the class to diverse practical problems and interpret the results.

The class project covers:

- Choose a real-world machine learning problem with an objective
- Procure and analyze the data
- Develop at a minimum 3 (if group = 3 students) or 4 (if group = 4 students) machine learning models
- Outline data / model assumptions & limitations / hypotheses
- Evaluate model metrics such as overfitting / underfitting
- Select final model based on selection criteria
- Determine how the model can be improved
- Effectively communicate analytical work

For the class project, students are graded on 8 criteria outlined below. Each criterion is 5 points for a total of 40 points. The criteria are:

1. Presentation

2. Problem Statement
3. Assumptions/Hypotheses about data and/or modeling
4. Exploratory Data Analysis
5. Feature Engineering & Transformations
6. Proposed Approaches (Models) with checks for overfitting/underfitting
7. Proposed Solution (Model Selection) with regularization, if needed
8. Results (Accuracy), and Future Work

I am available to work with each group on their project problem formulation and objectives. Each group member is expected to work at least on one model in addition to data analysis and other parts of the project and needs to explicitly state her/his contribution to the project.

To determine that a student has met the objectives of this course, the grading includes homework assignments and class project weighted as follows:

Assignments	40%
Quizzes	15%
Class Project	40%
Class participation	5%

GRADING SCALE

A = 93%–100%
A- = 90%–92%
B+ = 87%–89%
B = 83%–86%
B- = 80%–82%
C+ = 77%–79%
C = 73%–76%
C- = 70%–72%
F = 0%–69%

COURSE SCHEDULE

Academic quarters consist of 9 weeks of instruction, with the 10th week for assessment or course rescheduling. Refer to the university's academic calendar at www.uchicago.edu/academics/calendar/ for quarterly start and end dates.

Course book chapters corresponding to each week's topics are in parentheses.

Important Note: Changes may occur to the syllabus at the instructor's discretion. When changes are made, students will be notified via email and in-class announcement.

Week 1 (DL: 2 – 4, 5, HML: 1, 4, ISL: 2 – 3, DLP: 4)

Introduction to Machine Learning
Types of Learning
ML Terminologies
Universal ML workflow
Capacity, Overfitting and Underfitting
Nonparametric models
Data Mining with Ordinary Least Squares (OLS)

Week 2 (DL: 5, 7, 16, HML: 4, 8, ISL: 4, 6, DLP: 3)

Linear Algebra and Gradient Descent
Bayes Error
Classification
Bayesian Networks
Bias-variance tradeoff
Regularization
Ridge and Lasso regressions
Dimensionality reduction

Week 3 (DL: 5, HML: 2, 6 – 7, ISL: 5, 8, DLP: 4)

Types of Binning
Historical Data Challenges – selection bias
Feature Engineering
Validation and Cross-validation in action
Loss Functions – Jacobian and Hessian

Week 4 (DL: 5, HML: 5, 7, ISL: 8, 9)

Bagging and Boosting
Random Forests
Ensemble learning
Support Vector Machines
Kernel models

Week 5 (DL: 6, HML: 5, 7, 10, DLP: 2 – 3)

Recommenders
Collaborative filtering
Artificial Neural Network
Back propagation

Week 6 (DL: 6, 7, 9, 12, HML: 10 – 11, 13, DLP: 1 – 3, 5)

ANN Activation Functions
ANN Loss Functions
Deep Learning
Convolution Neural Network (CNN)

Week 7 (DL: 8, 10, HML: 14, 16, DLP: 6 – 7)

Batch Normalization
Recurrent Neural Network (RNN)
Vanishing Gradient
LSTM and GRU

Week 8 (DL: 11, 12, 14, HML: 10, 11, 14, 15, DLP: 4, 8)

Active Learning – learning with limited label datasets

Reinforcement Learning

Hyperparameter Tuning

Computer Vision and Transfer Learning

Double descent

Week 9 (DL: 12, 20, HML: 11, 12 – 13, 16, DLP: 7, 9)

Natural Language Processing (NLP) and Transformers

Autoencoders and Variational Autoencoders (VAE)

Generative Adversarial Networks (GAN)

Model Calibration and Fairness

Safe ML Models – Assurance using Interpretability

Model development to production deployment

Model Interoperability with ONNX

Week 10 (DLP: 9)

Final Project Presentations

Graph Neural Network (GNN)

Limitations of Deep Learning

Course Review

ATTENDANCE

This course meets Wed 6 – 9pm CST. Your attendance is required and paramount to your success in this class. You are allowed to miss at most two sessions, provided that you make arrangements with the instructor in advance.

In order to allow students to follow quarantine guidelines, instructors should be prepared to offer students the ability to complete their coursework remotely while they self-isolate. All students that require accommodations for COVID-related absences should be referred to their current area Dean of Students, dosgraham@uchicago.edu.

Students who have been exposed to or who are experiencing symptoms of COVID-19 should contact UChicago Student Wellness immediately to be tested, and reach out to their area Dean of Students to request accommodations for classes

- At least 10 days have passed since symptoms first appeared and;
- At least 3 days (72 hours) have passed since recovery- defined as resolution of fever without the use of fever-reducing medications and improvement in respiratory symptoms (e.g., cough, shortness of breath).

LATE WORK

All assignments must be submitted to this course's Canvas site on the due date. A student is allowed one assignment due date extension (no more than 3 days from due date) with no penalty and two delayed submissions (no more than 2 weeks from due date) with a 20% penalty. Beyond these 3 allowed extensions/delays, no credit is given for other late assignments. Please do not expect any extension beyond 3 days for the last assignment since evaluation has to be completed on time for grades submission.

REQUESTING REASONABLE ACCOMODATIONS

The University of Chicago is committed to ensuring equitable access to our academic programs and services. Students with disabilities who have been approved for the use of academic accommodations by Student Disability Services (SDS) and need a reasonable accommodation(s) to participate fully in this course should follow the procedures established by SDS for using accommodations. Timely notifications are required in order to ensure that your accommodations can be implemented. Please meet with me to discuss your access needs in this class after you have completed the SDS procedures for requesting accommodations.

Phone: (773) 702-6000
Email: disabilities@uchicago.edu

Please follow accommodation implementation instructions provided by the disability liaison in the division after you have completed the SDS procedures for requesting accommodations.

You may want to begin by reading through the information published on this website https://disabilities.uchicago.edu/.

ACADEMIC HONESTY & PLAGIARISM

It is contrary to justice, academic integrity, and to the spirit of intellectual inquiry to submit another's statements or ideas of work as one's own. To do so is plagiarism or cheating, offenses punishable under the University's disciplinary system. Because these offenses undercut the distinctive moral and intellectual character of the University, we take them very seriously.
Proper acknowledgment of another's ideas, whether by direct quotation or paraphrase, is expected. In particular, if any written or electronic source is consulted and material is used from that source, directly or indirectly, the source should be identified by author, title, and page number, or by website and date accessed. Any doubts about what constitutes "use" should be addressed to the instructor.

Academic Honesty and Plagiarism sites:
https://studentmanual.uchicago.edu/academic-policies/academic-honesty-plagiarism/
https://internationalaffairs.uchicago.edu/page/honest-work-and-academic-integrity-plagiarism

Copyright site: http://www.lib.uchicago.edu/copyrightinfo/

STUDENT HEALTH PACT

All students on campus are required to adhere to the guidelines in the UChicago Health Pact in order to create a safe environment in the classroom.

- Secure face coverings must be worn appropriately at all times while in University buildings
- Maintain a distance of 6 feet from others
- Do not attend an in-person class if you feel unwell or are experiencing COVID-19 related symptoms

The complete text of the UChicago Health Pact along with additional information about COVID-19 protocols can be found here.

REPORTING COVID-19 RELATED CONCERNS

Any concerns over inappropriate PPE usage, physical distancing, cleaning/disinfection, or other COVID-19 related public health concerns should be directed to UCAIR. If there is an emergency, call 773-702-8181 or dial 123 on any campus phone.

REPORTING COVID-19 EXPOSURE OR A CONFIRMED CASE

If you were potentially exposed to COVID-19 or your COVID-19 test results come back positive, reach out immediately to C19HealthReport@uchicago.edu.

RECORDING AND DELETION POLICIES FOR ACADEMIC YEAR 2020-2021

The Recording and Deletion Policies for the current academic year can be found in the Student Manual under Petitions, Audio & Video Recording on Campus.

- Do not record, share, or disseminate any course sessions, videos, transcripts, audio, or chats.
- Do not share links for the course to those not currently enrolled.
- Any Zoom cloud recordings will be automatically deleted 90 days after the completion of the recording.