Data and Knowledge Modeling and Analysis

Course code: ECE657A **Term:** Winter 2021

Instructor: Prof. Mark Crowley (mcrowley@uwaterloo.ca) *Department: Department of Electrical and Computer Engineering, University of Waterloo, Faculty of Engineering*

Permanent Course Website: https://uwaterloo.ca/scholar/mcrowley/ece657a

Teaching Period: January 11 – April 14, 2021

Available Office hours: TBD first week or classes in discussion with students

Course Description

Engineers encounter data in many of their tasks, whether the sources of this data may be from experiments, databases, computer files or the Internet. There is a dire need for effective methods to model and analyze the data and extract useful knowledge from it and to know how to act on it. In this course you will learn the fundamental tools for assessing, preparing and analyzing data. You will learn to design a data and analysis pipeline to move from raw data to task solution. You will learn to implement a variety of analytical and machine learning algorithms to including supervised, unsupervised and other learning approaches. Students will gain practical experience with coding and analysis through assignments. Research students will have opportunity to connect course material to their research as a project instead of some of the assignments.

Recommended background: *Data Structures and Algorithms, Basic Programming Skills (Python especially) and Basic Knowledge of Probability and Statistics Theory*.

Major Topics:

These topics are an outline, and each year some subset of non-core topics will be skipped due to time constraints and in order to benefit students through deeper focus. When tests are planned to assess knowledge of the material the mandatory core topics will be highlighted to students so they know which will be tested.

1. Understanding and Preparing Data
    1. Data types, sources, nature, scales and distributions
    2. Data representations, transformation and normalization
    3. Experimental Methodology, statistical tests and validation metrics
2. Fundamentals of Estimation and Learning

    1. Parameter Estimation, statistical approaches, MLE, MAP, EM, density estimation
3. Representation Learning I
    1. Feature extraction : PCA, LDA, ISOMAP, LLE
    2. Dimensionality Reduction and Manifold Learning
    3. Vector Embeddings : tf-idf, Word2Vec, BERT
4. Supervised Learning
    1. Classification I : kNN, Decision Tree based, Ensemble Methods
    2. Classification II : kernels and latent models (SVM)
5. Semi-/Self-/Unsuprevised Learning
    1. Clustering: Partition, Hierarchical, Model and Density based.
    2. Anomaly Detection: Classification, Outlier, Density, and Isolation based
6. Deep Learning
    1. Fundamentals of Neural Networks
    2. Effective Deep Learning Training Methods: Attention, Regularization, Optimizers
    3. Types of Deep Learning : CNN, RNN,
    4. Classification III : Data, Image and Timeseries classification using Deep Learning
    5. Representation Learning II: Autoencoders
7. Additional Learning Topics (some of thes topics will be covered, depending on time)
    1. Active Learning
    2. Transfer Learning
    3. Incremental/Online Learning
    4. Reinforcement Learning

Timing of Assessments, Lectures and Interaction

- **Assessments:** No more than one Test or Assignment will be due on any given week.
- **Pre-Recorded Content:** Every week there will be pre-recorded lecture content available
- **Weekly Live Sessions:** There will be 2-3 hrs of weekly live sessions from course staff (instructor and/or TAs) which will include
    - lectures (recorded) on core, high level topics
    - help sessions (unrecorded) to discuss content, answer questions and build class community

Assessment:

The goal of this course is to help students learn how to **analyse and prepare** data, **describe and apply** theoretical concepts in Data Science and Machine Learning, **design** data processing pipelines and **implement** important machine learning algorithms on a range of datasets and tasks. This course will be a success if students can use these skills in their future endeavours, research and employment. To assess this, students will complete assignments that require all of these skills on real datasets and they will also complete small tests focussing on the theoretical and design aspects of

these skills.

Weighting of Assessments

- **65% Assignments** (four assignments, done in pairs or alone)
    - Assignments will arise from the major component topics of the course, some will buid on previous assignment outcomes.
    - Possibility to have later assignments as Kaggle-style competitive submissions (note: vast majority of grade will be based on performance and correcteness rather than based on competitive performance)
    - Assignments will involve multiple skills:
        - mathematical analysis of data and results
        - logical design and clear description of a expeirmental methodology
        - programming various algorithms for processing, training and analysis of data to achieve given tasks (Programming will be in Python using libraries such as sci-kit learn and tensor flow)
- **35% Tests** (3 tests on different parts of the course, online timed question banks or take-home formats submitted to Crowdmark)
- **?% Participation** (optional, will be discussed collectively at start of term and can be shifted from the other components)
    - This could include engagement in live sessions, creation of annotated writeups of course materials for other students to utilize, or literature reviews of papers for other students to read.

Getting Help:

- **Discussion board:** piazza.com/uwaterloo.ca/winter2021/ece657a *Access Code: DKMA*
    - *Piazza* will be the main place for detailed discussion and questions. Students can post anonymously (from students only), post a collaborative answer and course staff can confirm these, post their own or run Live Q&A events.
    - Go there there and sign up with your UWaterloo email now!
    - *Community:* we will strive to create an active community within the course using live sessions, social platforms and crosscutting study group assignments and fun challenges throughout the term. This will be an experiment, so it will be optional, but hopefully we can all build a Data Analysis and Machine Learning graduate student community in the deparment that will benenfit everyone.
- **Pre-recorded Video Lectures:** These will be made available on a youtube channel, and links from wihtin Learn
- **LEARN Website:** The main course content, announcements, grade stracking and materials will be made availble on Learn. All registered students should see this in their LEARN courses.

- **Email the Teaching Assistant and Instructor:** Office Hours will be arranged once term starts as needed.
- **AccessAbility Services :** http://uwaterloo.ca/accessability-services
    - If you need any accommodation, assistance with exams, learning environment, assignments, talk to this office and they can help you set it up as securely and anonymously as possible.

Resources:

There is no required textbook. But most of the course is based on the following books and will be useful to take a look at them.

1. Arxiv Tutorials - links will be provided to a number of detailed tutorials on some course topics.
2. K. Murphy. “Machine Learning: A Probabilistic Perspective”. MIT Press, 2012.
3. I. Goodfellow, Y. Bengio and A. Courville. “Deep Learning”. MIT Press, 2016.
    - Online for free at http://www.deeplearningbook.org*. The first half covers many of basics of this course, while the second half focusses on Deep Learning only.
4. R. O. Duda, P. E. Hart and D. G. Stork. Pattern Classification (2nd ed.), John Wiley and Sons, 2001.

Papers and electronic references will be made available on the course website which is on LEARN (go to http://learn.uwaterloo.ca to log in).

Recipe for success:

Ask questions. Connect with your classmates. Do the assignments. ***Ask questions.*** Most of all, *have fun*.

Policy and Rules

Academic Integrity:

In order to maintain a culture of academic integrity, members of the University of Waterloo community are expected to promote honesty, trust, fairness, respect and responsibility. [Check www.uwaterloo.ca/academicintegrity/ for more information.]

Grievance:

A student who believes that a decision affecting some aspect of his/her university life has been unfair or unreasonable may have grounds for initiating a grievance. Read Policy 70, Student Petitions and Grievances, Section 4, http://www.adm.uwaterloo.ca/infosec/Policies/policy70.htm. When in doubt please be certain to contact the department's administrative assistant who will provide further assistance.

Discipline:

A student is expected to know what constitutes academic integrity to avoid committing academic offenses and to take responsibility for his/her actions. A student who is unsure whether an action constitutes an offense, or who needs help in learning how to avoid offenses (e.g., plagiarism, cheating) or about "rules" for group work/collaboration should seek guidance from the course professor, academic advisor, or the undergraduate associate dean. For information on categories of offenses and types of penalties, students should refer to Policy 71, Student Discipline, http://www.adm.uwaterloo.ca/infosec/Policies/policy71.htm. For typical penalties check Guidelines for the Assessment of Penalties, http://www.adm.uwaterloo.ca/infosec/guidelines/penaltyguidelines.htm.

Plagiarism-detection software may be used on any submitted work.

Appeals:

A decision made or penalty imposed under Policy 70, Student Petitions and Grievances (other than a petition) or Policy 71, Student Discipline may be appealed if there is a ground. A student who believes he/she has a ground for an appeal should refer to Policy 72, Student Appeals, http://www.adm.uwaterloo.ca/infosec/Policies/policy72.htm.

Note for students with disabilities:

The Office for Persons with Disabilities (OPD), located in Needles Hall, Room 1132, collaborates with all academic departments to arrange appropriate accommodations for students with disabilities without compromising the academic integrity of the curriculum. If you require academic accommodations to lessen the impact of your disability, please register with the OPD at the beginning of each academic term.