ENEL 645 – Data Mining & Machine Learning

Overview of the course

Roberto Souza
Assistant Professor
Electrical and Computer Engineering
Schulich School of Engineering

09 January 2023

ADVANCED IMAGING & ARTIFICIAL INTELLIGENCE LAB

@lab_ai2

@CalgaryAI

UNIVERSITY OF CALGARY

Instructor and TAs

Instructor:

- Roberto Souza – roberto.medeirosdeso@ucalgary.ca

TAs:

- Mike Lasby – mklasby@ucalgary.ca
- Abbas Omidi – abbas.omidi@ucalgary.ca

- Please avoid contacting the instructor and TAs directly by email unless it is an issue specific to a grade produced by the TAs.

Course Delivery

- Synchronous and in-person
  - MWF – 9:00 am to 9:50 am
  - Room SA 106

- Office hours between 10 am and 10:50 am after M and W classes
  - Starts on the 2nd week of class
  - My office ICT 352C

- It is a large class, but **don't be a stranger! Come to the office hours with your questions.**

Course Syllabus

Types of data mining: **classification, clustering, association, prediction. Processes: data preparation, model building**. Techniques: decision tree, **neural network**, evolutionary computing, Bayesian network. Applications: multi-media, text and web mining.

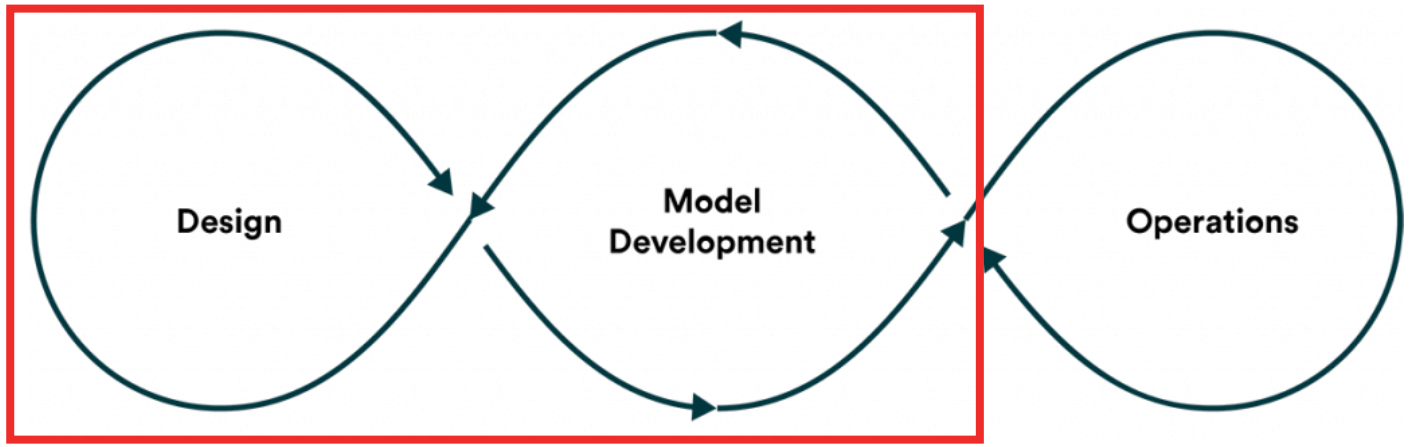
Course Syllabus (main topics)

- Python Bootcamp and machine learning concepts
- Data preparation and pre-processing
- Regularization techniques
- Traditional machine learning models (Decision Trees, Random Forests, …)
- Neural Networks
- Transfer Learning and Domain Adaptation
- Generative models
- Self-supervised learning
- Automated Machine Learning

Learning Outcomes

1. Design and develop data mining and machine learning solutions for relevant problems
2. Select appropriate experimental setups and metrics for evaluating machine learning models
3. Select appropriate machine learning models for different types of problems
4. Have a comprehensive overview of current trends in machine learning
5. Acquire hands-on experience with machine learning programming frameworks

What this course is about?



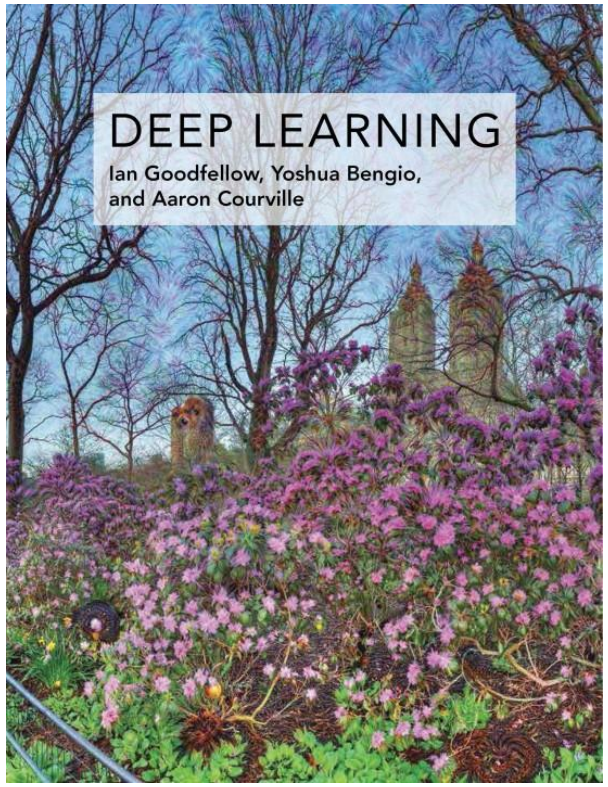
- Model deployment
- CI/CD pipelines
- Monitoring & triggering

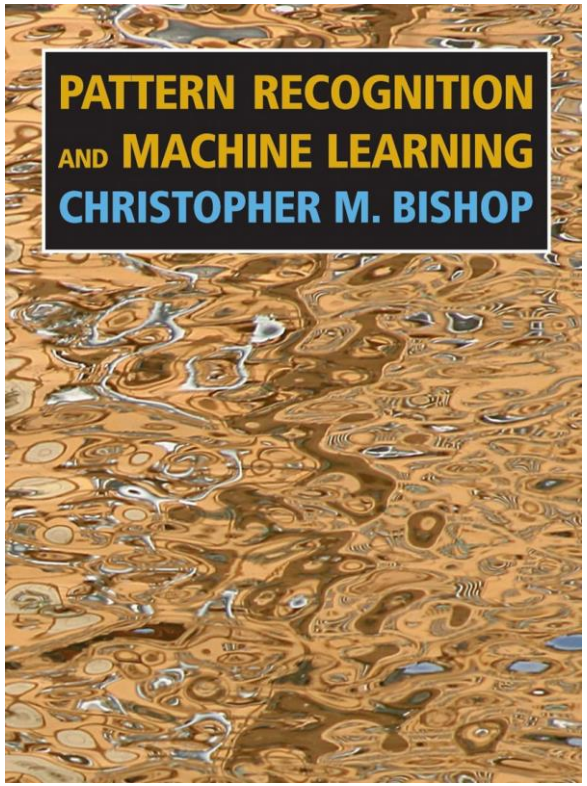
This course is about designing and developing machine learning models to achieve the best quantitative metrics* to the problems being modelled.

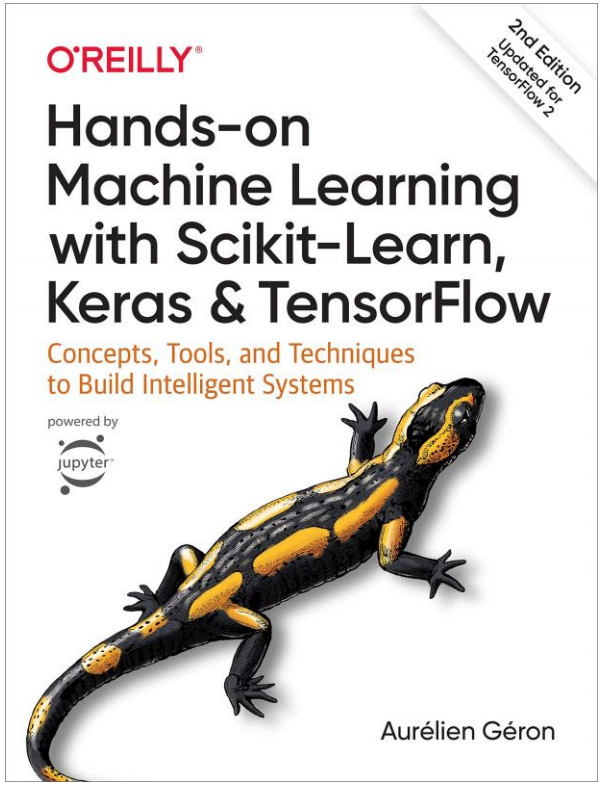
*Quantitative results are determined on a problem-by-problem basis.

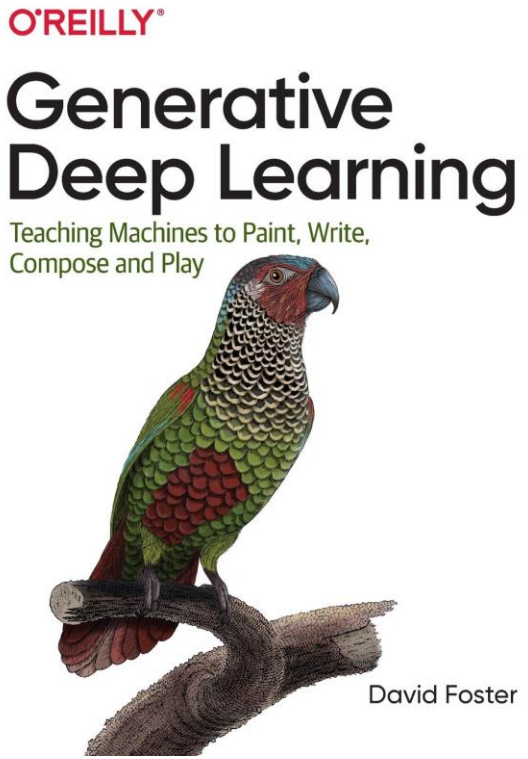
Textbook

- No mandatory textbook for this course









Class GitHub repo: https://github.com/rmsouza01/ENEL645

Course Assessment

Component	Learning Outcomes	Weight
Participation	1, 2, 3, 4	5%
Assignments (2)	1, 2, 3, 4, 5	20%
Quizzes (3)	1, 2, 3, 4	30%
Final Project	1, 2, 3, 4, 5	45%

- The lowest quiz grade will be dropped

Participation (5%)

- Student participation will account for 5% of the final grade. How participation will be measured?
  - Students questions and answers during class
  - Students participation on the D2L discussion board
  - Students helping each other during class

Assignments (20%)

- Team-based – 6 people per team

- **Assignment 01** (10%)**:**
  - Proposing a garbage classification system based on images and natural language
  - **Due:** January 30th (midnight) | **Delivery method:** D2L dropbox

- **Assignment 02** (10%):
  - Create a machine learning hands-on tutorial
  - **Due:** March 6th (midnight) **| Delivery method:** D2L dropbox

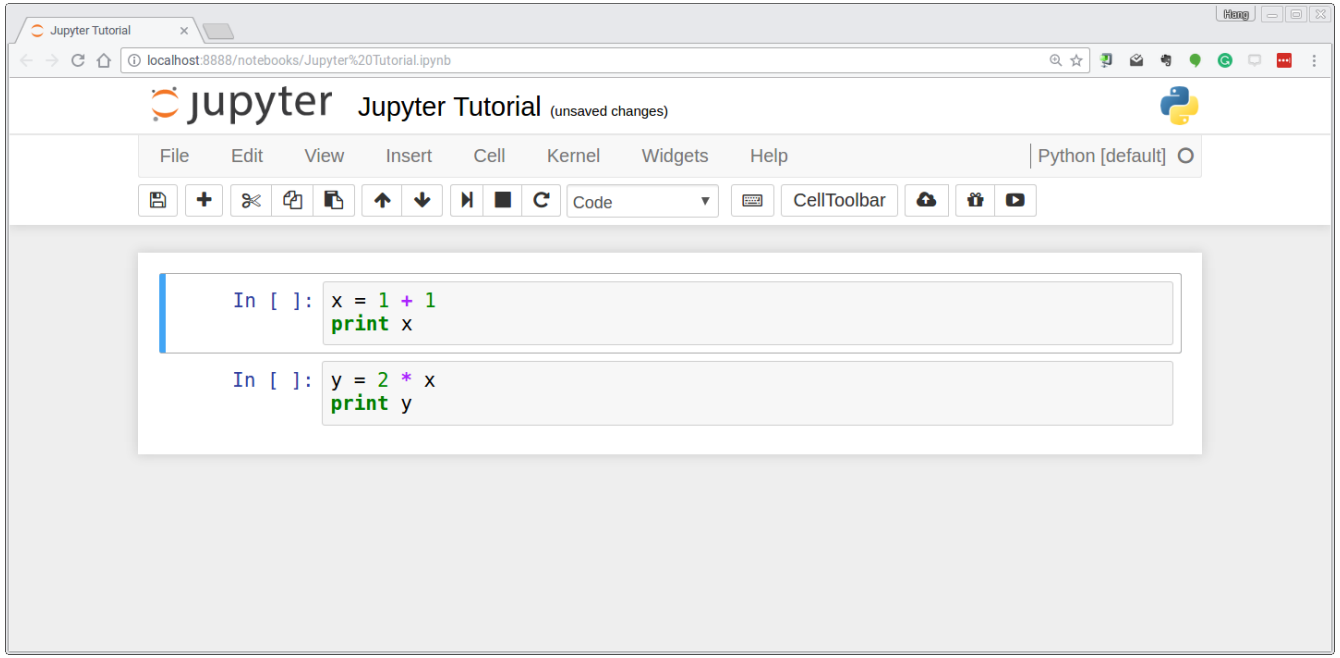
Assignment 1

Clean cardbox -> blue trash bin

Greasy cardbox -> green trash bin

“Another potential issue is the class overlap for certain items. For example, a clean pizza box can be disposed of in the blue bin, however if it is greasy/covered in food toppings, it should be disposed of in the green bin. Depending on how the photo is taken, it may be impossible to know for sure which class this item belongs to.”

Assignment 02



- Create a machine learning tutorial using techniques seen in class

UNIVERSITY OF CALGARY

Quizzes (30%)

- 3 quizzes – 2 highest grades are kept
- Quizzes are individual
- Content: all topics covered until the day of the quiz
- **Quiz 01: January 27th**
- **Quiz 02: February 17th**
- **Quiz 03: February 27th**
- The dropped quiz is to accommodate potential student absence

Final Project (45%)

- 6-page report + 1 additional page only with references (if necessary)

- Report template
  - Overleaf - please make a copy for your team.
  - Microsoft word

- Report and 10-minute presentation recording due date: **March 27 at 9 am**

- Final project presentations: **March 27th to April 5th**
  - 5-minute presentation + 5-minutes for questions
  - Send slides one day before your presentation

Grades

Letter Grade	Total Mark (T)
A+	T≥95%
A	90% ≤ T < 95%
A-	85% ≤ T < 90%
B+	80% ≤ T < 85%
B	75% ≤ T < 80%
B-	70% ≤ T < 75%
C+	65% ≤ T < 70%
C	60% ≤ T < 65%
C-	55% ≤ T < 60%
D+	50% ≤ T < 55%
D	45% ≤ T < 50%
F	T < 45%

The Programming Environment (Part 1)

https://colab.research.google.com/

https://jupyter.org/

https://github.com/rmsouza01/ENEL645

https://www.overleaf.com/project

UNIVERSITY OF CALGARY
Research Computing Services me_Page

The Programming Environment (Part 2)

- Python 3
- Python libraries:
  - NumPy
  - Matplotlib
  - Pandas
  - Scikit-learn
  - Tensorflow (version ≥ 2.0)
  - PyTorch
  - Weight and Bias

- Please have your programming environment in your computer or on Google Colab set up asap

Deep Learning Frameworks

I hope you enjoy the class ☺

Questions?