MIE1516H: Structured Learning and Inference (Research Course)

Instructor:	Prof. Scott Sanner, BA8104, ssanner@mie.utoronto.ca
Office hours:	Immediately following lectures
Text:	Various readings (PDFs available)
Prerequisites:	MIE1513 (or machine learning experience, with instructor approval)
Lectures:	(3 hours/week) Tues, 13:10-14:00 in BA2159; Wed 13:10-15:00 in BA2159
Lab:	(2 hours/week) Thurs, 10:10-12:00 in RS303

Course description

This *Research Course* will provide students with the conceptual, theoretical, and implementational foundations of fundamental tools for structured learning and inference: probabilistic graphical models, probabilistic programming, and deep neural networks. The course will focus on the design and training of structured models for specific application use cases such as answering probabilistic queries over data, sequence tagging and classification, and image recognition through programming-intensive projects including a final independently proposed research project with report component.

Course structure

Lecture	Learn techniques for representing, learning, and performing inference with a variety of structured probabilistic and non-probabilistic formalisms.
Lab	Learn how to use Python-based tools to architect and evaluate specific applications of structured inference and learning.

Course goals

Become proficient with the principles and application of structured learning and inference tools:

- Understand graphical model representations of probability distributions.
- Understand message passing algorithms for graphical model inference and learning.
- Understand Markov Chain Monte Carlo algorithms for graphical model inference and learning.
- Understand the semantics of probabilistic programs and their reduction to graphical models.
- Understand basic principles of artificial neural networks.
- Understand advances that have enabled modern deep neural networks (regularization, optimization, autodifferentiation, GPU-based parallelization, novel architectural formalisms).
- Understand deep network architectures for sequential data processing (e.g., recurrent networks for speech or text) and multidimensional data processing (e.g, convolutional networks for images).
- Understand various methods for benchmarking and evaluation of predictive models as well as diagnosing and explaining predictions.
- Apply the above principles and tools such as STAN, RDDL, Infer.NET, Tensorflow, and Keras to a variety of data-driven engineering application use cases.

Grading

Assignment	Weight	Date
Course Warmup	10%	Week 2
Project 1 (Graphical Models)	10%	Week 4
Project 2 (Probabilistic Programs)	10%	Week 6
Midterm Exam	25%	Week 7
Project 3 (Deep Learning: Images)	10%	Week 8
Project 4 (Deep Learning: Sequences)	10%	Week 10
Independently Proposed Final Project	25%	Week 13

Projects are due on Sunday at 10:00pm of the week listed unless otherwise stated.

Project submissions:

- Project submissions will be online through github.
- Projects up to 24h late will be given a 30% penalty.

Academic honesty

In this course, we will make heavy use of online tutorial content that you will need to leverage in completing your assignments. In light of this, the following constitute plagiarism on project submissions:

- Copying any segment of code from online sources without citation
- Submitting code that you did not write yourself or adapt from cited online sources

The source code of all submitted projects will be analyzed for statistical similarities using Measure of Software Similarity (MOSS, http://theory.stanford.edu/~aiken/moss/). Students suspected of plagiarism on an assignment will be referred to the department for formal discipline for breaches of the Student Code of Conduct.

Student responsibilities

- Don't talk or otherwise cause distractions in lecture.
- Attend lab and tutorial: they are needed for projects and attendance will be taken.
- Ask questions in class, lab, office hours, or on blackboard.

Preliminary Schedule of topics

The schedule of topics below is subject to change. Following are explanations for abbreviations used in the schedule:

- PGM: Probabilistic Graphical Models
- PP: Probabilistic Programming
- DNN: Deep Neural Networks
- PGM+DNN: Advanced topics at the intersection of PGMs and DNNs

Week	Lecture	Lab Topic	Due
1	Course Overview	Microsoft Azure, IPython	
2	PGM (Representation, Message Passing)	Course Warmup	Course Warmup
3	PGM (Markov Chain Monte Carlo)	Project 1	
4	PP (Motivation, Reduction to PGM)	Project 1	Project 1
5	PP (Modeling, Use Cases)	Project 2	
6	DNN (Fully Connected Networks, Tensorflow)	Project 2	Project 2
7	DNN (Convolutional Deep Networks)	Project 3	
8	DNN (Embeddings and Sequential Classifiers)	Project 3	Project 3
9	DNN (Recurrent Neural Nets for Sequences)	Project 4	
10	DNN (Tagging and Transducer Models)	Project 4	Project 4
11	DNN+PGM (Autoregressive Deep Networks)	Final Project	
12	DNN+PGM (Variational Autoencoders and GANs)	Final Project	
13	Course Review	Final Project	Final Project