CIS600 Data Science and Causality

L.C. Smith College of Engineering and Computer Science Syracuse University
Spring 2019

Catalog Description
Causal inference is an integral part of data science where experimental datasets are not available. With the correct assumptions, it is possible to discover the underlying causal structure without intervention. This course covers and discusses causality and the causal discovery algorithms, including most recent research in the field.

Prerequisites
CIS 321 - Introduction to Probability and Statistics or equivalent understanding and knowledge of probability and statistics (check with lecturer)

Instructor Information
Name: Dr. Martin Harrison
Email: mharriso@syr.edu
Office: 4-??? CST
Office Hours: TBA

Name: Jakob Zeitler
Email: jkzeitle@syr.edu
Office: 4-124 CST
Office Hours: TBA

Meeting Time and Location
Lecture: Twice a week

Course Web Address
The course will be hosted on http://causal-inference.org . We will use Blackboard (http://blackboard.syr.edu) for administration such as Quizzes and marking.

Textbook
1. Required: Elements of Causal Inference, Peters et al., 2017 (link: https://www.dropbox.com/s/o4345krw428kyld/11283.pdf?dl=1)
2. Recommended: Causal Inference in Statistics: A Primer, Judea Pearl
3. Research papers provided on a need basis (link: https://docs.google.com/spreadsheets/d/1bylwl_hI-kVtxyTsNcNd6SQRQXJ3CO4xZszDIgFMtIM/)

Software and Computer Systems
The majority of lectures and assignments will be available as Jupyter Notebooks, a python based data science platform. Occasionally, examples might be presented in C, Matlab or R.

Acquired Knowledge (Course Outcomes)

Upon successful completion of this course, students are expected to have the ability to

- Outcomes: describe and explain the identification of causality in experimental and observational data.
  Assessment: Students will take final exams, homework and weekly quizzes.
- Outcomes: describe and explain the fundamental issues of causality.
  Assessment: Students will take final exams, homework and weekly quizzes.
- Outcomes: define, restate, discuss, and explain the different types of causal discovery models such ANM and IGCA
- Assessment: Students will take final exams, homework and weekly quizzes.
- Outcomes: design and construct the following causal discovery algorithms: independence test for ANM and Kolmogorov complexity for IGCI
  Assessment: Students will design and implement the above methods in Jupyter Notebooks with python kernels
- Outcomes: illustrate, construct, compose and design data science projects
  Assessment: Students will design and implement the above methods in Jupyter Notebooks with python kernels
- Outcomes: measure, evaluate, and compare causal inference methods
  Assessment: Students will run experiments on their own implemented ANM/ICGI models and compare the statistical results
- Outcomes: discuss with fellow students about designing new components of OS.
  Assessment: There will be one hour of discussion session in the semester to discuss this matter. The session is to provoke alternative design ideas for OS components. The exams and homework will assess the outcomes.

Course Outline
Section 1: Introduction
Section 2: Correlation and Causality
Section 3: Review: Probability and Statistics
Section 4: Reichenbach Principle
Section 5: Structural Causal Models
Section 6: Additive Noise Models
Section 7: Information Geometric Causal Inference (Kolmogorov complexity)
Section 8: Benchmarking and comparison of causal discovery models

Course Outcomes
Implement causal discovery methods from start to finish to solve a data science problem
Understand the assumptions for causal inference
Program in Python and R
Use Jupyter Notebook or an equivalent platform to develop and present data science projects
Explain concepts of causal graphs, causal discovery, independence, probability, and other fundamental ideas in causality
Explain differences between computer science and statistics
Present history of statistical inference and the pioneers in causal inference

Explain various sub-disciplines in causal inference and how these sub-disciplines are related to each other and related to other academic disciplines outside of computer science

Outcome Measurement

Late penalty: 2^days points

Quiz	15%
Homework	15%
Project and Presentation (1-2 students)	30%
Final	30%

Academic Integrity

All members of the Syracuse University community — faculty, staff, and students — are expected to exhibit academic integrity in all situations. As a member of this community, you should also be familiar with the University's academic-integrity policy, which is available at http://academicintegrity.syr.edu. In addition, you must be familiar with the course specific policies on academic integrity.

Accommodations

Students who are in need of disability-related academic accommodations must register with the Office of Disability Services (ODS), 804 University Avenue, Room 309, 315-443- 4498. Students with authorized disability-related accommodations should provide a current Accommodation Authorization Letter from ODS to the instructor and review those accommodations with the instructor. Accommodations, such as exam administration, are not provided retroactively; therefore, planning for accommodations as early as possible is necessary. For further information, see the ODS website http://disabilityservices.syr.edu.

Religious Observances

SUs religious observances policy recognizes the diversity of faiths represented among the campus community and protects the rights of students, faculty, and staff to observe religious holy days according to their tradition. Under the policy, students are provided an opportunity to make up any examination, study, or work requirements that may be missed due to a religious observance provided they notify their instructors before the end of the second week of classes. For fall and spring semesters, an online notification process is available through MySlice/Student Ser- vices/Enrollment/My Religious Observances from the first day of class until the end of the second week of class. SUs religious observances policy can be found at http://supolicies.syr.edu/emp_ben/religious observance.htm.

	Week of	Quiz	Homework	Project	Exam	Content	Lecture 1	Lecture 2
1	January 14	1	Intro to Jupyter Notebooks			Section 1	Introduction	Correlation
2	January 21	2					Causality	Causality
3	January 28	3				Section 2	Probability	Probability
4	February 4	4				Section 3	Statistics	Statistics
5	February 11	5		Proposal out		Section 4	Reichenbach Pri	Reichenbach Pri
6	February 18	6		Proposal due		Section 5	SCM 1	SCM 1
7	February 25	7				Section 6	ANM	ANM
8	March 4	8	ANM out				ANM	ANM
9	March 11	9	Break	Break	Break	Break	Break	Break
10	March 18	10	ANM due	Review out		Section 7	ICGA	ICGA
11	March 25	11		Review due			ICGA	ICGA
12	April 1	12	IGCI out			Section 8	Comparison	Comparison
13	April 8	13					Comparison	Comparison
14	April 15	14	IGCI due				Open-ended	Open-ended
15	April 22	15		Presentation		Project	Presentation	Presentation
16	April 29	16		Presentation			Presentation	Presentation
17	May 6				Final			