PRINCIPIA COLLEGE

CSCI 390 - Topics in CSCI: Data Science
Fall 2022

August 29, 2022 - December 14, 2022

Instructor Information

Instructor: Clinton Staley
Title: Professor, CSCI
Email: Clint.Staley@Principia.edu
Office Location: SC-137, or online at: https://principia-edu.zoom.us/j/2822437109?pwd=VHBwVktqYWtPVnpxblhmRjQweEg0UT09
Office Hours: TWF 4:15-5p, MR 10-11p (CSCI lab)
Cell Phone: 805-835-5024 (prefer texts)

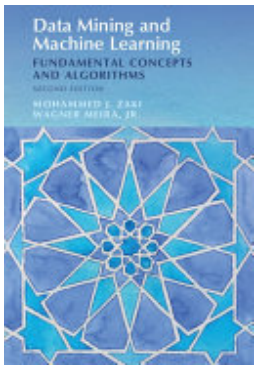
Basic Information

Semester Hours: 4

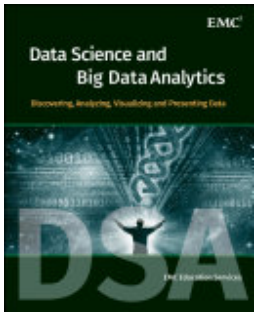
Prerequisite: See catalog description below.
Meeting Times/Locations: M W F 0220 - 0410 pm , OWL 104
Catalog Description: Topics will vary from year to year based on the needs and interests of students and instructor. All offerings will involve in-depth study of a specific topic, software design, the construction of working software, and self-instruction in addition to standard instruction. May be taken more than once if topics are different. Prerequisite: CSCI 240.

Course Texts

Data Mining and Machine Learning

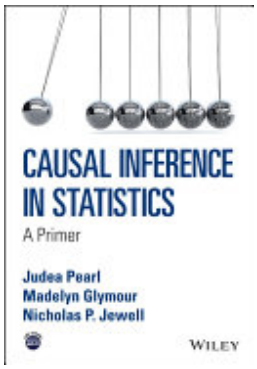


The fundamental algorithms in data mining and machine learning form the basis of data science, utilizing automated methods to analyze patterns and models for all kinds of data in applications ranging from scientific discovery to business analytics. This textbook for senior undergraduate and graduate courses provides a comprehensive, in-depth overview of data mining, machine learning and statistics, offering solid guidance for students, researchers, and practitioners. The book lays the foundations of data analysis, pattern mining, clustering, classification and regression, with a focus on the algorithms and the underlying algebraic, geometric, and probabilistic concepts. New to this second edition is an entire part devoted to regression methods, including neural networks and deep learning.
9781108658690
Mohammed J. Zaki, Wagner Meira, Jr
Cambridge University Press
2019-12-31



Data Science and Big Data Analytics
Data Science and Big Data Analytics is about harnessing the power of data for new insights. The book covers the breadth of activities and methods and tools that Data Scientists use. The content focuses on concepts, principles and practical applications that are applicable to any industry and technology environment, and the learning is supported and explained with examples that you can replicate using open-source software. This book will help you: Become a contributor on a data science team Deploy a structured lifecycle approach to data analytics problems Apply appropriate analytic techniques and tools to analyzing big data Learn how to tell a compelling story with data to drive business action Prepare for EMC Proven Professional Data Science Certification Corresponding data sets are available from the book's page at Wiley which you can find on the Wiley site by searching for the ISBN 9781118876138. Get started discovering, analyzing, visualizing, and presenting data in a meaningful way today!
9781118876053
EMC Education Services
John Wiley & Sons
2015-01-05

Causal Inference in Statistics



Many of the concepts and terminology surrounding modern causal inference can be quite intimidating to the novice. Judea Pearl presents a book ideal for beginners in statistics, providing a comprehensive introduction to the field of causality. Examples from classical statistics are presented throughout to demonstrate the need for causality in resolving decision-making dilemmas posed by data. Causal methods are also compared to traditional statistical methods, whilst questions are provided at the end of each section to aid student learning.

9781119186847
Judea Pearl, Madelyn Glymour, Nicholas P. Jewell
John Wiley & Sons
2016-03-07

Additional Materials

Note that Zaki et al is available online without charge at dataminingbook.info. Also note that Pearl et al is expected only for an extra credit assignment, and can be delayed till later in the term.

Course Description

We'll cover theoretical and practical aspects of the analysis of very large datasets, including classic statistical methods like correlation and PCA, unsupervised learning methods like K-means and other clustering, supervised methods like decision trees and SVMs. An extra credit project may also involve using recent mathematical tools for causality analysis, e.g. causal graphs.

Course Outcomes

1. Write code to analyze large datasets for significant patterns, using a variety of algorithms and methods.

2. Understand intuitively the mathematics underlying algorithms for pattern recognition in large data sets.

3. Understand the organizational and logistical problems associated with high volume data, and the methods for solving them.

4. (Optional) Understand and apply recent mathematical developments regarding causality analysis, e.g. causality graphs.

Course Assignments

This will be adjusted as the semester progresses.

Grading Practices

(Tentative -- may be adjusted)

Work	**Points**	**Due**
Lab work analyzing data sets	35	During term
Personal project of choice	20	End of course
Class Participation and Quizzes	15	Each class session
Midterms	30	Weeks 8 and 14
(Extra) Project involving causality	20	End of course

Course Policies

Class Attendance
This course includes in-class discussion of concepts and lecture content not included in the text. Every occurrence of absence or tardiness after the first three may result in a 5% reduction in the final grade for the course. Your full attention on course work is expected during class periods.

Academic Honesty
Programming, like other creative disciplines such as art or writing, requires individual practice and excellence. Work on individual projects is to be done entirely on your own, or with assistance from approved resources, e.g. the instructor and LAs. Work on team projects is to be done only with team members.

If you're not sure a resource is approved, check with the instructor. If you would like to volunteer as a formally approved assistant to other students, contact the instructor for such approval. Your help is welcome if your class standing is good and if you understand proper tutoring practices.

The course materials online include a non-collaboration agreement, which you must read and sign during the first week of class.

College Policies

Syllabus Changes
The faculty member reserves the right to make changes to this published syllabus if it is in the best interest of the educational development of this class. Any such changes will be announced as soon as possible and, insofar as practical, after consultation with the whole class.

Inclement Weather Policy
During the winter season, Principia may decide to cancel all classes due to inclement weather. In that event, faculty will communicate through Canvas the course expectations during the cancellation. If there are any in-person graded activities scheduled for that day (e.g. presentation, exam, quiz, etc.), faculty will let students know how it will be rescheduled. Remote courses will also be cancelled. Students are expected to keep on top of their homework during a "snow day." Deadlines for assignments will still be observed.

Principia's Academic Integrity Policy and **Academic Misconduct Policy** can be found in the College Catalog, and are reprinted below. The Principia Student Community Commitment can be found here: www.principiacollege.edu/pcc Academic Misconduct Policies and Procedures can be found on the Registrar's website at: prinweb.principia.edu/internal/registrar/academic-misconduct

> **Academic Integrity Policy**
> A Principian is expected to pursue a life of integrity. See the Principia Student Community Commitment. Therefore academic honesty is essential to a Principia College education.
>
> **Academic Misconduct Policy**
> Students are expected to refrain from all forms of academic misconduct, including but not limited to lying for academic gain, cheating, plagiarizing, unauthorized sharing of work, and any form of misrepresenting another's work as one's own. Students are also expected to refrain from helping others with any of these or other forms of academic misconduct.

Final Exam Schedule Fall 2022

Regardless of the expectations of this class, students are reminded that they are **required to attend all finals scheduled for the classes they are taking.** Students must petition the Scholastic Committee **before the last class day of week 11** if they wish to request a change to a scheduled exam time. ***Students should NOT purchase air fare for a flight at the end of a term that conflicts with an exam time, unless approval has already been granted for an***

exam time change by the Scholastic Committee.

Class Meeting Days/Times: M W F 0220 - 0410 pm

Final exam days: Monday May 9 - Wednesday May 11

Time Slot	Meeting Days	Meeting Time	Exam Day	Exam Time
M1	Mon, Wed, Fri	8:00-8:50 a.m.	Monday	8:00-10:00 a.m.
M2	Mon, Wed, Fri	9:00-9:50 a.m.	Monday	10:30 a.m.-12:30 p.m.
M3	Mon, Wed, Fri	10:00-10:50 a.m.	Tuesday	8:00-10:00 a.m.
M4	Mon	11:55 a.m.-12:45 p.m.	Tuesday	6:00-8:00 p.m.
MW	Mon, Wed	12:55-2:10 p.m.	Tuesday	10:30 a.m.-12:30 p.m.
M5	Mon, Wed, Fri	2:20-3:10 p.m.	Wednesday	8:00-10:00 a.m.
M6	Mon, Wed, Fri	3:20-4:10 p.m.	Wednesday	10:30 a.m.-12:30 p.m.
M7	Mon, Wed	4:20-6:30 p.m.	Monday	6:00-8:00 p.m.
M8	Mon	7:30-10:20 p.m.	Tuesday	3:30-5:30 p.m.
T1	Tues, Thur	8:00-9:15 a.m.	Monday	1:00-3:00 p.m.
T2	Tues, Thur	9:25-10:40 a.m.	Monday	3:30-5:30 p.m.
T3	Tues, Thur	12:05-1:20 p.m.	Tuesday	1:00-3:00 p.m.
T4	Tues, Thur	1:30-2:45 p.m.	Wednesday	1:00-3:00 p.m.
T5	Tues, Thur	2:55-4:10 p.m.	Wednesday	3:30-5:30 p.m.
T7	Tues, Thur	4:20-6:30 p.m.	Tuesday	6:00-8:00 p.m.
A7	Any day	4:20-6:30 p.m.	Monday	6:00-8:00 p.m.

The final examination periods for all courses are two hours long. Exams for multi-slot classes will be scheduled during the exam slot for the first class meeting time only.