CHE 599: Introduction to Data Science for Engineers

Fall 2019 (3 credits)
Lecture: M/W/F 10 am - 10:50 am in Owen Hall 106.

🍎 Instructor:
Dr. Cory Simon [website]
📧 Cory.Simon@oregonstate.edu
Office hours: W 11 am - 12 am; KEC 2045.

📚Learning resources: We will use several textbooks as resources, all free:

- The Elements of Statistical Learning. Link
- Computational and Inferential Thinking. Link
- Statistics with Julia: Fundamentals for Data Science, Machine Learning and Artificial Intelligence. Link
- Learn Julia in 5 min. Link

Slack group

We shall communicate via Slack. Log in via https://oregonstate.enterprise.slack.com/ I encourage you to post questions and have discussions in public channels so that we can collaborate and learn from each other.

💬 Anonymous comment/feedback box: Submit here.

✔ Course Summary

As engineers, we often gather large amounts of data from which we aim to extract useful insights, infer, and predict. CHE 599 is an introduction to basic concepts forming the foundation of data science and machine learning. We emphasize the practical application of computational methodologies to, from data, extract useful knowledge and make inferences and predictions using Julia, a new programming language for numerical computing. Topics of study include computer programming in Julia, data visualization and communication, hypothesis and A/B testing via simulation, uncertainty estimation via bootstrap, dimensionality reduction, prediction (classification, regression), and fitting differential equation models to data. The goal is to

illustrate basic concepts and methodologies in data science and machine learning and equip students with an initial foundation for solving engineering problems through data and computation.

✔Course Specific Measurable Student Learning Outcomes

- query, clean, and summarize data using the Julia programming language in a Jupyter Notebook
- create publication-quality data visualizations that faithfully and effectively communicate the data with minimal cognitive burden on the audience
- test hypotheses via simulation, understand/interpret p-values
- conduct A/B tests via permutation tests
- estimate a statistic or model parameters and quantify uncertainty through bootstrapping
- understand the inner-workings of and implement a classification (e.g. k-nearest neighbors, logistic regression) and regression (linear regression, neural network) algorithm and assess its predictive performance via cross-validation
- reduce the dimensionality of an unlabeled data set via principal component analysis

Downloading Julia and Jupyter Notebook

Download and install Julia from here and Jupyter Notebook from here. You will also need Python and its libraries matplotlib and scikitlearn, which we call from Julia. I recommend the Anaconda distribution.

📘 Course content (subject to change)

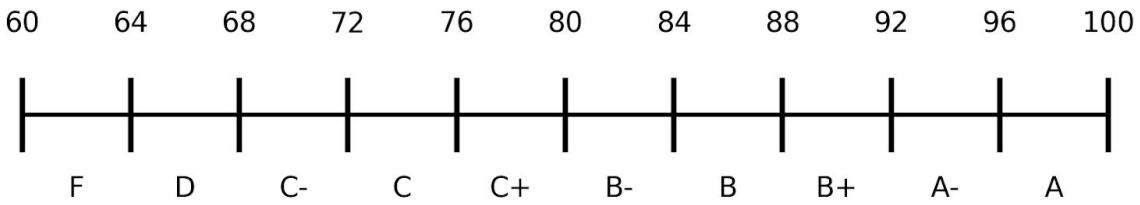
wk	lecture topic	resource	homework
1	introduction to computer programming in Julia		
2	querying, cleaning, summarizing data in DataFrames.jl data visualization (matplotlib)		
3	kernel density estimation hypothesis testing via conducting simulations understanding and interpreting p-values		
4	A/B testing via permutation tests parameter estimation and uncertainty quantification via bootstrapping		
5	optimization: gradient descent, the Nelder-Mead algorithm classification: the k-nearest neighbors algorithm (Scikitlearn.jl) decision boundaries, the bias-variance tradeoff, overfitting		
6	the K-folds cross-validation paradigm		

	classification: logistic regression (Scikitlearn.jl) regression: linear regression		
7	regression: neural networks dimensionality reduction: principal component analysis		
8	fitting differential equation models to data and quantifying uncertainty in the parameters via bootstrap		
9	working with and summarizing graphs in LightGraphs.jl Dijkstra's shortest path algorithm		
10	final project oral presentations		

📊 Evaluation of student performance

The final grade is comprised of:

homework 75%
final project + oral presentation 25%



📝 Homework

I encourage you to work with a homework partner and pair program. Each pair can turn in the same code/homework write-up. For each homework, please turn in an executive summary and the associated code to reproduce the plots.

🕐 Final project

Two options:
(a) propose your own data science project, using a data set from your research or from online, and employ methods learned in this class to extract

value from the data. Submit project proposals at the end of week 6 for feedback and so I can ensure it has a wide enough scope to be justified for a final project.
(b) I will propose a few projects from which you can choose if you have trouble coming up with your own project.

Turn in all code to reproduce the results of your project and a 5-page write-up structured like a scientific article. During the last week of class, you will give an oral presentation to the class on your final project.

More details and guidance will be provided week 6.

👮 Student Conduct Code / Academic or Scholarly Dishonesty

Please familiarize yourself with the student conduct expectations: http://studentlife.oregonstate.edu/code

Statement Regarding Students with Disabilities

Accommodations for students with disabilities are determined and approved by Disability Access Services (DAS). If you, as a student, believe you are eligible for accommodations but have not obtained approval please contact DAS immediately at 541-737-4098 or at http://ds.oregonstate.edu. DAS notifies students and faculty members of approved academic accommodations and coordinates implementation of those accommodations. While not required, students and faculty members are encouraged to discuss details of the implementation of individual accommodations.

The DAS Statement is posted online at ds.oregonstate.edu/faculty-advisors (4/14/16).

Reach Out for Success

University students encounter setbacks from time to time. If you encounter difficulties and need assistance, it's important to reach out. Consider discussing the situation with an instructor or academic advisor. Learn about

resources that assist with wellness and academic success at oregonstate.edu/ReachOut. If you are in immediate crisis, please contact the Crisis Text Line by texting OREGON to 741-741 or call the National Suicide Prevention Lifeline at 1-800-273-TALK (8255)

Student Evaluation of Courses

The online Student Evaluation of Teaching system opens to students the Wednesday of week 8 and closes the Sunday before Finals Week. Students will receive notification, instructions and the link through their ONID. They may also log into the system via Online Services. Course evaluation results are extremely important and used to help improve courses and the learning experience of future students. Responses are anonymous (unless a student chooses to “sign” their comments agreeing to relinquish anonymity) and unavailable to instructors until after grades have been posted. The results of scaled questions and signed comments go to both the instructor and their unit head/supervisor. Anonymous (unsigned) comments go to the instructor only.