ECE 29595–WNG:
Introduction to Data Science

Spring 2019

Lectures: Fridays, 1:30–2:20, WALC 3154
Course web page: https://engineering.purdue.edu/datamind/datascience/19spring/
Piazza discussion page: https://piazza.com/purdue/spring2019/ece29595/
Course calendar: https://calendar.google.com/calendar/embed?src=24sq2b7k9g8rncvi2hmrc2egbo%40group.calendar.google.com&ctz=America%2FNew_York

Instructors:
       Milind Kulkarni (**milind@purdue.edu**)
       Office: EE 324A
       Office Hours: Tuesdays, 10–11; Wednesdays, 1:30–2:30; or by appointment

TAs:
       Nikhil Hegde (hegden@purdue.edu)

TA Office Hours: TBD

Course Outcomes: A student who successfully fulfills the course requirements will have demonstrated:
1. The ability to write data analyses in Python
2. The ability to build statistical models and use them for prediction
3. The ability to design analyses/models to solve engineering problems

These outcomes are extremely high level. In more detail, after taking this course you will be able to:

- Explain data analysis and modeling algorithms like sampling, estimation, regression
- Write basic data analyses in Python, taking advantage of language features such as higher-order functions (map/reduce) and complex data structures (including NumPy arrays, kd-trees, etc.)
- Use these tools to propose, design, and implement a set of data analyses to solve engineering problems, then visualize and present the results.

Course assessment: The achievement of the course objectives will be through 5–6 programming assignments (covering objectives 1 and 2), and a mini-project (covering objective 3), that will have two components: a project proposal, and a final report.

Course grading: Grades will be assigned as follows:
      60% — Programming assignments (10–12% per assignment)
      35% — Project (10% proposal, 25% final report)
      5% — Class participation/attendance

Class participation will be assessed by attendance: because there are only 15 class sessions, attendance is required. You are allowed 2 unexcused absences, but all other absences must be for a valid reason (e.g., illness with documentation from PUSH, excused absence for an interview, etc.)

Programming assignments: Programming assignments will be due approximately every two weeks. They will test the concepts covered in class, both programming and statistical. The following rules apply to *all* programming assignments:

1. All programs should run correctly in python 2.7.10 (the version of python available on ecegrid), with versions ≥0.22.0 of pandas, ≥1.14.0 of numpy, and ≥1.0.0 of scipy. While you may need to install these yourself for development purposes, you can assume that our grading environment will have them.
2. Your assignment submissions should be either (a) a Jupyter notebook (and any other accompanying files) with code that will produce the required output *and writeup*; or (b) a python script (or scripts) that will produce the required output when run as well as a *separate* writeup.
3. Unless otherwise specified, assignments are due at 11:59 PM on the deadline.

Programming assignments will be submitted via GitHub Classroom (**https://classroom.github.com**). As such, *you are required to have a GitHub account.* These can be obtained for free at **https://github.com**.

Please fill out the form here: **https://goo.gl/forms/CscnGcV8bIMski1z2** to provide your GitHub account information.

Late submission policy: Except for medical and family emergencies (accompanied by verification), there will be *no individual extensions* granted for programming assignments. Late submissions will be scaled according to lateness, docking 10% from your score per day late, up to a maximum of 50%. Submissions more than 5 days late will be assigned a score of 0.

Course Discussion: This term we will be using Piazza for class discussion. If you have questions about the course or the project, we encourage you to post them on Piazza. It's a shared discussion forum, where your question can be answered by the instructors, the TAs or your fellow students!

Find our class's Piazza page at: **https://piazza.com/purdue/spring2019/ece29595/**

Students who are active participants on Piazza may receive class participation bonus points (Over and above the 5% attendance score)

Email: Questions about course material or programming assignments *should be posted to Piazza* or raised during lecture or office hours. The professor and TAs will not answer programming questions via email. This is to allow other students who might have similar questions to benefit from our answers. Of course, if you have questions of a personal or confidential nature, we welcome your email.

Course announcements: Course announcements, including changes in due dates, course topics, programming assignment details, etc., will be communicated in three ways:

1. Updates to the relevant webpage(s)
2. Announcement posts on Piazza and Blackboard
3. Changes to due dates on course calendar

Course Topics: Below is a rough list of topics covered in the course:

1. Python Basics (loops, functions, arrays, lists)
2. Histograms: sample vs. population, optimal bin width
3. Higher-order functions, closures, map/reduce
4. Distribution: random variables, distribution, probability
5. N-grams
6. Estimation: mean, variance, confidence intervals, bootstrap
7. Data manipulation with pandas
8. Regression: linear regression, prediction
9. Dealing with missing data
10. Classification
11. Clustering
12. Neural networks

Academic Honesty: Unless expressly allowed, you are expected to complete all assignments by yourself. However, you are allowed to discuss general issues with other students (programming techniques, clearing up confusion about requirements, etc.). You may discuss particular algorithmic issues on Piazza (but do not copy code!). *We will be using software designed to catch plagiarism in programming assignments, and all students found sharing solutions will be reported to the Dean of students.*

Punishments for academic dishonesty are severe, including receiving an F in the course or being expelled from the University. By departmental rules, all instances of cheating will be reported to the Dean. On the first instance of cheating, students will receive a 0 on the assignment; the second instance of cheating will result in a failure of the course.

Campus Interruptions: In the event of a major campus emergency, course requirements, deadlines and grading percentages are subject to changes that may be necessitated by a revised semester calendar or other circumstances beyond the instructor's control. In such an event, information will be provided through the course website and email.