

CSE 351: Introduction to Data Science

Syllabus

Term: Spring 2020

Instructor: Pravin Pawar (pravin.pawar@sunykorea.ac.kr, Office B424, +82-32-626-1227, +82-10-8692-4908)

Lectures: Mon & Wed 10:30 AM -11:50 AM

Office Hours: Tue & Thu 10:30-12:30 pm in B424 or by appointment

Course Homepage: <http://ppawar.github.io/Spring2020/CSE351-S20/index.html>

Teaching assistants (TA) are available for tutoring in the CS Commons in B Building on the 4th floor. The TA schedule will be posted after the semester begins.

Course Description

This multidisciplinary course introduces both theoretical concepts and practical approaches to extract knowledge from data. Topics include linear algebra, probability, statistics, machine learning, and programming. Using large data sets collected from real-world problems in areas of science, technology, and medicine, we introduce how to preprocess data, identify the best model that describes the data, make predictions, evaluate the results, and finally report the results using proper visualization methods. This course also teaches state-of-the-art tools for data analysis, such as Python and its scientific libraries.

Prerequisites

CSE214 or CSE260; AMS 310; CSE Major

Textbooks

Required Texts:

(Textbook 1) Joel Grus, Data science from scratch: first principles with python. O'Reilly Media, 2019.

(Textbook 2) Ian H. Witten, Frank Eibe, Mark A. Hall, and Christopher J. Pal. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

Reference Texts:

(Reference book 1) Foster Provost and Tom Fawcett. Data Science for Business: What you need to know about data mining and data-analytic thinking. O'Reilly Media, Inc., 2013.

Grading

The course provides a total of 500 points distributed across the following categories.

- **Programming assignments (30%, 5 assignments given)** = 30% (150 points). Web pages will be designed and developed by students in various ways using different technologies for these assignments.
- **Class exercises and surprise quizzes (~10, 2% each)** = 20% (100 points). There will a number of class exercises and quizzes given during the lecture on the material covered in class that week. These exercises and quizzes are intended to make sure students are keeping up with the material.
- **Mid-term exams (30%, 2 exams given)** = 30% (150 points). These will be written exams.
- **Final exam/hackathon** = 20% (100 points).

Your final grade is calculated by the total points you accumulate in the class.

- **Important note:** You must attain a grade of at least 50% on exams and 50% on assignments to achieve a grade higher than C-. Additionally, if you have over 20% unexcused absences, the final course grade will be an F.

Makeup examinations will only be given for extenuating circumstances (e.g. hospital admission) or for verified, officially sanctioned university activities. All makeup examinations may be oral.

Regrading

Should you discover what you think is an error in grading your work, you have exactly one week after the grades are made available to you to request a regrade - no exceptions.

To promote consistency of grading, questions and concerns about grading should be addressed first to the TA and then, if that does not resolve the issue, to the instructor.

You are welcome to contact the TA by email or come to their office hours. If you would like to speak with the TA in person and have a schedule conflict with their office hours, you are welcome to make an appointment to meet the TA at another time.

For the final exam, there will be a special office hour designated to resolve any grade queries or disputes. This will be announced after the final exam. The final exam papers will not be returned.

Tentative Weekly Class Schedule

The following **tentative** course schedule provides topics, problem sets, quiz dates, and exam dates. Check back frequently.

Date	Activity	Topics	Readings	Course Material
Mon, Feb 24, 2020	Lecture	Course Introduction		
Wed, Feb 26, 2020	Lecture Exercise/quiz			
Mon, Mar 2, 2020	Lecture			
Wed, Mar 4, 2020	Lecture Exercise/quiz			

Mon, Mar 9, 2020	Lecture			
Wed, Mar 11, 2020	Lecture Exercise/quiz	Assignment 1		
Mon, Mar 16, 2020	Lecture			
Wed, Mar 18, 2020	Lecture Exercise/quiz			
Mon, Mar 23, 2020	Lecture			
Wed, Mar 25, 2020	Lecture Exercise/quiz			
Mon, Mar 30, 2020	Lecture	Assignment 2		
Wed, Apr 1, 2020	Lecture Exercise/quiz			
Mon, Apr 6, 2020	Lecture	Midterm 1		
Wed, Apr 8, 2020	Lecture Exercise/quiz			
Mon, Apr 13, 2020	Lecture			
Wed, Apr 15, 2020	[No Class – Parliamentary elections Day]			
Mon, Apr 20, 2020	Lecture			
Wed, Apr 22, 2020	Lecture Exercise/quiz	Assignment 3		
Mon, Apr 27, 2020	Lecture			
Wed, Apr 29, 2020	Lecture Exercise/quiz			
Mon, May 4, 2020	[No Class - Adjustment Day]			
Tue, May 5, 2020	[No Class - Children's Day]			
Mon, May 11, 2020	Lecture	Midterm 2		
Wed, May 13, 2020	Lecture Exercise/quiz	Assignment 4		
Mon, May 18, 2020	Lecture			

Wed, May 20, 2020	Lecture Exercise/quiz			
Mon, May 25, 2020	Lecture			
wed, May 27, 2020	Lecture Exercise/quiz			
Mon, Jun 1, 2020	Lecture			
Wed, Jun 3, 2020	Lecture Exercise/quiz	Review for final		
Final	TBD			
Final Grades	TBD	Final Grades		

Academic Integrity: Cooperation vs Cheating

Working with others on assignments is a good way to learn the material and we encourage it. However, there are limits to the degree of cooperation that we will permit.

When working on programming assignments, you must work only with others whose understanding of the material is approximately equal to yours. In this situation, working together to find a good approach for solving a programming problem is cooperation; listening while someone dictates a solution is cheating. You must limit collaboration to a high-level discussion of solution strategies, and stop short of actually writing down a group answer. Anything that you hand in, whether it is a written problem or a computer program, must be written in your own words. If you base your solution on any other written solution, you are cheating.

It is okay to help other students, within limits. If you are asked for help by another student, two things that are absolutely forbidden are to show that student your solution or to put your hands on that student's keyboard or paper. That isn't helping; that is facilitating cheating! Instead answer questions, give tips, help with tools, explain Python, point out a bug, and/or give encouragement. In other words, interact with other students the way that the TAs do.

When taking an exam, you must work completely independently of everyone else. Any collaboration here, of course, is cheating.

We do not distinguish between cheaters who copy others' work and cheaters who allow their work to be copied.

If you cheat, you will be referred to the appropriate office at the University. If you have any questions about what constitutes cheating, please ask.

The statement from the university on academic integrity is: Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Faculty members are required to report any suspected instances of academic dishonesty to the Academic Judiciary Committee or the Department of Academic Affairs, Campus Building A, Room 201, (032) 626-1121.

Students with Disabilities

If you have a physical, psychological, medical or learning disability that may impact your course work, please contact the Department of Student Affairs, Campus Building A, Room 207, (032) 626-1190. They will determine with you what accommodations, if any, are necessary and appropriate. All information and documentation is confidential.

Critical Incident Management

SUNY Korea expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Department of Academic Affairs any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn.