

Fall 2020

## CS 301-003: Introduction to Data Science

Pan Xu

Follow this and additional works at: <https://digitalcommons.njit.edu/cs-syllabi>

---

### Recommended Citation

Xu, Pan, "CS 301-003: Introduction to Data Science" (2020). *Computer Science Syllabi*. 161.  
<https://digitalcommons.njit.edu/cs-syllabi/161>

This Syllabus is brought to you for free and open access by the NJIT Syllabi at Digital Commons @ NJIT. It has been accepted for inclusion in Computer Science Syllabi by an authorized administrator of Digital Commons @ NJIT. For more information, please contact [digitalcommons@njit.edu](mailto:digitalcommons@njit.edu).

# CS 301 (003), Introduction to Data Science, Fall 2020

## 1 Basic information

**Instructor:** Pan Xu, GITC 4310, pxu@njit.edu. Here is the website: <https://sites.google.com/site/panxupi/about-me--1?authuser=0>. Office Hours: by appointment. Feel free to email to the instructor to schedule an appointment.

**Teaching assistant (TA):** Erik Lampe, erl2@njit.edu. Office hours: Tues 3 PM to 4 PM; Thurs 10 AM to 11 AM. Please click this link to enter the TA's online chatting room: <https://njit.webex.com/meet/erl2>. Feel free to email to TA to schedule an online appointment other than TA's office hours.

**Class schedule:** Wed/Friday, 9 AM to 10:20 AM. The class is delivered via the synchronous online manner. Please click this link to enter the online classroom: <https://njit.webex.com/meet/pxunjit.edu>.

**Course overview:** The course will focus on the math and algorithmic side of data science. Specifically, we will discuss several basic models and fundamental algorithms related to data science, including Bayes' models, linear regression, streaming algorithms, reinforcement learning, online learning, and online matching algorithms in E-Commerce like crowdsourcing and ridesharings.

**Prerequisite:** CS 114, Math 333. This course will require good knowledge in probability theory, statistics, linear algebra, and calculus II. Also, it assumes students have basic programming skills.

**Textbooks:** No textbooks are required.

## 2 Course content and detailed schedule

There are several components to this course as follows:

1. Instructor lectures.
2. Individual assignments: There will be expectedly four homework assignments, involving both algorithm design and implementation.
3. Final project: Students will expect to complete and submit a summary report individually.
4. Midterm and final exam.
5. Class participation.

Dates	Topics
W1	Introduction to data science
W2	Probability theory and statistics (HW1 will be posted)
W3	Probability theory and statistics
W4	Basic Python tools for data science, Major Rule
W5	Linear regression, Naive Bayes
W6	Naive Bayes, Reinforcement learning (HW2 will be posted)
W7	Reinforcement learning
W8	Reinforcement learning (Project will be posted)
W9-11	Streaming algorithms and Multi-Armed Bandits (HW3 will be posted on W10)
W12-14	Online learning (HW4 will be posted on W13)
W15	Review for the final (There will be only one class on Wed, Dec. 9)
W16	Final exam

### 3 Submission and grading of assignments

Items	Grade (% of final grade)	(Posted date, Due date)
HW1	7.5	(W2,W4)
HW2	7.5	(W6,W8)
HW3	7.5	(W10, W12)
HW4	7.5	(W13, W15)
Final Project	15	(W8,W15)
Midterm	20	W8
Final Exam	30	W16
Class Participation	5	

Here are a few important remarks.

- Generally, all HWs will be posted and due on the Fridays of the week except the last one: HW4 will be due on the Wed of W15 such that the solution can be posted well before the final exam date.
- Midterm will take place on Friday, Oct. 23, from 9 to 10:20 AM, and the final will be on Wed, Dec. 16, from 9 AM to 10:20 AM. Both are open books and notes.
- 20% late submission penalty if submitted within 24 hours of the deadline. Beyond that, late submission is not accepted. Unless otherwise stated, all submissions are due by 11:59 PM (EST) on the due date.
- For all HWs and the final project, students will expect to upload a PDF on Canvas by the deadline. Note that only PDF files are accepted. It is strongly encouraged that all students should use Latex editors to type their solutions including all equations and math symbols in a proper way. A very useful online cloud-based Latex editor is overleaf: <https://www.overleaf.com/>.
- The part of class participation will be graded as follows: every student will be expected to scribe at least one lecture offered starting from the second week. The submission is due one

week after the starting time of the lecture scribed. All students should send their scribed note (only PDF accepted) directly to the instructor with the subject line “Scribing assignment of CS 301”.

- Grading scale: A: Top 20%; B+: Top 21-40% B: Top 41-60%; C+/C/D/F/W: TBD.

## 4 Statement on academic integrity

Academic Integrity is the cornerstone of higher education and is central to the ideals of this course and the university. Cheating is strictly prohibited and devalues the degree that you are working on. As a member of the NJIT community, it is your responsibility to protect your educational investment by knowing and following the academic code of integrity policy that is found at: <http://www5.njit.edu/policies/sites/policies/files/academic-integrity-code.pdf>.

Please note that it is my professional obligation and responsibility to report any academic misconduct to the Dean of Students Office. Any student found in violation of the code by cheating, plagiarizing or using any online software inappropriately will result in disciplinary action. This may include a failing grade of F, and/or suspension or dismissal from the university. If you have any questions about the code of Academic Integrity, please contact the Dean of Students Office at [dos@njit.edu](mailto:dos@njit.edu).