

Syllabus for DS-GA 3001.009 Special Topics in Data Science: Applied Statistics

Semester: 2023-2024 Fall

Instructor: Yanjun Han, yh5107@nyu.edu

Section Leader: Yuxiao Wen, yw3210@nyu.edu

Lecture: Tue 4:55 - 6:35 PM at CDS 150 (60 Fifth Ave)

Lab: Fri 2:45 - 3:35 PM at Silver Center 401 (100 Washington Square East)

Number of Credits: 3

Course Description

This course provides a rigorous introduction to classical models and tools in applied statistics. About 50% of the lectures use exponential family structure to motivate generalized linear models and other useful applied techniques, and the other half cover models and methods in semiparametric (~20%) and nonparametric statistics (~30%). The main topics include exponential family, generalized linear model, bootstrap, survival analysis, missing data analysis, empirical Bayes, model selection, semiparametric models, average treatment effect, double robust estimation, density estimation, nonparametric regression, local polynomials and splines, wavelets methods, isotonic regression.

Prerequisites

- Undergraduate level probability and statistics
- Calculus I and Linear Algebra
- Some experience in programming
- Some prerequisites may be waived with permission from the instructor

Learning Outcomes

- Get acquainted with classical statistical models
- Learn mathematical and statistical tools for data science
- Understand ideas and philosophy in statistical modeling
- Familiar with programming in R via real-data examples

Contact Hours

There will be 14 weeks in total for the semester. Each week comprises 100 minutes of lectures; 50 minutes of recitations; and 2 hours of office hours for the instructor and section leader.

Schedule of Classes (Weekly Schedule)

Date	Activity	Homework
9/5/2023	Course logistics & review of basic statistics	
9/12/2023	Properties of exponential families	Homework 1 out (Due: 9/21)
9/19/2023	Parameter estimation and inference	Homework 2 out (Due: 9/28)
9/26/2023	Generalized linear models	Homework 3 out (Due: 10/5)
10/3/2023	Survival analysis and Cox model	Homework 4 out (Due: 10/12)
10/10/2023	No class (Monday schedule - Columbus Day)	
10/17/2023	Missing data and EM algorithm	Homework 5 out (Due: 11/2)
10/24/2023	Empirical Bayes	
10/31/2023	In-class midterm exam	
11/7/2023	Semiparametric models and efficient score/influence	Homework 6 out (Due: 11/16)
11/14/2023	Double robust estimation of average treatment effect	Homework 7 out (Due: 11/30)
11/21/2023	Nonparametric density estimation & regression	
11/28/2023	Local polynomial regression and splines	Homework 8 out (Due: 12/7)
12/5/2023	Fourier and wavelet methods	Homework 9 out (Due: 12/14)
12/12/2023	Isotonic regression & course recap	

Course Materials

Lecture notes will be provided after each lecture.

Additional reading materials:

- B. Efron, “Exponential Families in Theory and Practice” (<http://www.cs.columbia.edu/~blei/fogm/2020F/readings/Efron2018.pdf>);
- G. James, D. Witten, T. Hastie, R. Tibshirani, “An Introduction to Statistical Learning” (https://hastie.su.domains/ISLR2/ISLRv2_corrected_June_2023.pdf);
- L. Wasserman, “All of statistics” (<https://egrcc.github.io/docs/math/all-of-statistics.pdf>).

Assignments

- There will be 9 homework assignments in total;
- Homework will be released and collected on Gradescope;
- The homework with the lowest score will be dropped;
- Only 2 late homework submissions are accepted, each with 3 late days.

Grading

Final grade = $40\% * \text{homework} + \max(30\% * \text{midterm} + 30\% * \text{final}, 60\% * \text{final})$.

The numerical score will be normalized before conversion to letter grade.

Student Responsibilities

Students are expected to read/view assigned material of the class for which they are scheduled, attend class, participate in class, complete assignments, complete projects, and ask for help early if they are having trouble. Students are expected to be respectful of their fellow students and the instructor. This class is a collaborative learning experience.

Instructor Responsibilities

The instructor is expected to read/view the material prior to the class for which they are scheduled, prepare and deliver high quality introductions to the material, prepare exercises and assignments that are relevant to the course topic, and provide comments on assignments and projects intended to help students develop their abilities to work the course material.

Academic Honesty

NYU students and faculty to maintain the highest standards of academic honesty. Students can find information on the core principles and standards in the University’s policy on academic integrity, which is accessible at

<http://www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/academic-integrity-html>.

Student Accessibility

Academic accommodations are available to any student with a chronic, psychological, visual, mobility, learning disability, or who is deaf or hard of hearing. Students should please register with the Moses Center for Student Accessibility at 212-998-4980.

The Moses Center for Student Accessibility

726 Broadway, 2nd Floor

New York, NY 10003-6675

Telephone: 212-998-4980

Email: mosescsa@nyu.edu

Website: <https://www.nyu.edu/students/communities-and-groups/student-accessibility.html>