DS-GA 3001.001 Special Topics in Data Science: Probabilistic Time Series Analysis

Instructor: Cristina Savin
Section Leaders: Caroline Haimerl and Yiqiu Shen

Overview

This graduate level course presents fundamental tools for characterizing data with statistical dependencies over time, and using this knowledge for predicting future outcomes. These methods have broad applications from econometrics to neuroscience.

The course emphasizes generative models for time series, and inference and learning in such models. We will cover range of approaches including Kalman Filter, HMMs, AR(I)MA, Gaussian Processes, and their application to several kinds of data.

Note: information presented is tentative, syllabus may be subject to change as course progresses.

Course details
Lecture (Section 008)
Mo 2.00pm - 3.40pm
60 5th Avenue, Room 110

Lab (Section 009)
Wed 3:30pm-4:20pm
60 5th Avenue, Room 110

Enrollment cap: 79

Office hours
CS: Mo 4:00-5:00pm. Location: 60 5th ave, 6th floor, room 608.
CH/YS: TBD. Location: 60 5th ave, 6th floor, room 660.

Grading
problem sets (25%) +
midterm exam (25%) +
final project (25%) +
lab(20%)+
participation(5%)

Participation: piazza, engagement during lectures, labs, and office hours

Piazza
We will use Piazza for announcements, and discussions about the course. Interactions on Piazza, particularly good answers to other students' questions, will count toward the participation grade.

Projects
Work in groups of 2-3 students.* Topics are flexible, including applying know algorithms to an interesting dataset, reviewing and implementing a state of the art solution, to improving an existing algorithm. Project proposals due in week 4.
*Check with CS if you are considering working individually or in a larger group.

Video
Video recordings of the lectures will be available via NYU Classes. Class attendance is still required.

Date	Lecture title	Assignments
Sept. 9	Lecture 1: Logistics. Introduction. Basic statistics for characterizing time series.	
Sept. 11	Recap basic Bayes, graphical models	
Sept. 16	Lecture 2: AR(I)MA	Problem set 1
Sept. 18	Lab 1: ARIMA	
Sept. 23	Lecture 3: LDS; Kalman filtering	Projects initial discussion
Sept. 25	Lab 2: Inference in LDS	
Sept. 30	Lecture 4: Particle filtering	Problem set 2
Oct. 2	Lab 3: LSD parameter learning	
Oct. 7	Lecture 5: Hidden Markov Models	Project proposal due
Oct. 9	Lab 4: Particle filtering	
Oct. 14	FALL RECESS	
Oct.15 (TUE!)	Lecture 6: a unified view of linear models	Problem set 3
Oct.16	Lab 5: HMMs	
Oct.21	Guest lecture	
Oct.23	No lab. Office hours	
Oct.28	**Mid-term exam**	

Date	Lecture title	Assignments
Oct.28	No lab	
Nov. 4	Introduction to Gaussian Processes	Problem set 4
Nov. 6	Lab GP	
Nov.11	GP advanced topics. Intro to RNNs	
Nov.13	No lab. Work on projects	
Nov.18	Deep learning for time series	
Nov. 20	RNNS lab	
Nov. 25	Spectral methods 1	Problem set 5
Nov. 27	Thanksgiving	
Dec. 2	Spectral methods 2	
Dec. 4	Lab spectral methods	
Dec. 9	Final projects presentation	Project reports due Dec.15
Dec. 11	No lab	

Bibliography
There is no required textbook.
Required reading will be provided on the course page in electronic format.
We use *python* for all the lab work.

Core material
- Time series analysis and its applications, by Shumway and Stoffer, 4th edition
- Pattern recognition and machine learning, Bishop
- Gaussian processes Rassmussen & Williams

Useful extras
- Review notes from Stanford's machine learning class
- Sam Roweis's probability review
- Carlos Ferndandez's notes on Statistics and Probability for Data Science DS-GA 1002

Policies
Try to solve problems on your own first. If you get stuck, you can discuss homework questions with colleagues, but *you need to write up the final solution individually*.
Credit should be explicitly given for any code you use that you did not write yourselves.
Late submission penalties: 20% points off for each extra day of delay.