Probabilistic Time Series Analysis

Instructor: Cristina Savin, csavin@nyu.edu

Section leaders:
- Ashwin Siripurapu (in-person for section 002 - blended), ars991@nyu.edu
- Jiyuan Lu (remote for section 002 - blended), jl11046@nyu.edu
- Colin Bredenberg (remote for section 018), cjb617@nyu.edu

Graders:
- Jiyuan Lu - jl11046@nyu.edu
- Yiqiu (Artie) Shen - ys1001@nyu.edu

Web: https://github.com/savinteachingorg/pTSAFall2020.git

DS-GA 3001.001 **Lecture** (Cap: 80)
Tuesday from 2:00pm-3:40pm
Room: EDUC_176 (capacity = 54)
Instruction Mode: Blended

DS-GA 3001.002 **Lab** (Cap: 40)
Wednesday from 3:30pm-4:20pm
Room: 60FA_150 (capacity = 17)
Instruction Mode: Blended

DS-GA 3001.018 **Lab** (Cap: 40)
Wednesday from 9am-9:50am
Instruction Mode: Online

Office hours: twice a week, online; details TBD based on student needs.

Enrollment cap: 80

Target audience: intended primarily as an elective for MSDS students, with a small fraction of PhD students (CDS, CNS); a few seniors (Courant, Shanghai) have also successfully taken it in the past.

Prerequisites: Linear algebra, Basic probability.

Overview

This graduate level course covers fundamental probabilistic models for characterizing data with dependencies over time, and their use for predicting future outcomes. The methods covered 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 ARIMA, Kalman filtering, Hidden Markov Models (HMM), recurrent neural networks (RNNs), Gaussian Processes, and their application to several kinds of data. In the process of covering these models, we review key principles of probabilistic inference and learning, including exact and approximate inference (variational, sampling) and maximum likelihood learning (e.g. expectation maximization, EM). The lectures and homeworks emphasize basic mathematical understanding, while the labs and project allow students to develop practical skills in implementing different algorithms and to generalize the methods learned in class to real world problems. One invited guest lecturer each year provides insights into states of the art research and applications.

Course aims:

- Understanding statistical assumptions made by different models so as to be able to reason about the applicability of different tools to a given dataset
- Being able to derive from scratch fundamental algorithms for inference and learning, including Kalman filtering/smoothing, alpha-beta for HMMs, EM, basic back propagation through time
- Being able to implement in python all these fundamental algorithms
- Developing a clear overview of how different models relate to one another and how the fundamental models can be generalized to capture more complex statistical dependence in the data (mixtures, introducing nonlinearities, combining latent and AR structure, etc)

Grading

Problem sets (25%) There are 5 problem sets (each contributing 5% of final grade), distributed across the semester. Students are given 2 weeks to solve each set. Each set covers a specific section of material: 1) ARIMA models, 2) Kalman, 3) HMMs, 4) Gaussian processes and 5) Spectral methods. Most exercises are small derivations similar to those covered in the lecture, with some twists to check for understanding. One problem in each set has a higher level of difficulty.

Midterm (20%) Covers ARIMA, Kalman, HMMs and their generalizations.

Projects (25%): 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. Examples of past projects are listed on github.

Labs (20%): implementing basic algorithms; we use *python* for all the lab work. The lab solutions have to be handed in by the end of the week. They are graded separately, the best scoring 7 contribute to the final grade.

Participation (10%): pre-lecture and in-lecture small quizzes, piazza activity, engagement during lectures, labs, and office hours.

Logistics

Github course repository provides access to lecture notes, lab solutions, latex templates for reports, and the up-to-date schedule. Make sure to check it regularly.

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.

Videos covering key concepts are posted in advance, students are expected to have watched those before the in person lecture and to be able to answer a short quiz about them. Video of the in person lectures will be simultaneously streamed via zoom, with recording linked on NYU Classes afterwards.

Date	Lecture title	Assignments	Further reading
Sept. 8th	1. Logistics. Introduction. Basic statistics for characterizing time series.		Shumway and Stoffer

Date	Lecture title	Assignments	Further reading
Sept. 9	No lab. Recap Bayes, basic graphical models as online video, classes meet on Mo schedule.		
Sept. 15th	2. AR basic inference, learning	Problem set 1	Handout 1
Sept. 16th	Lab 1: AR		
Sept. 22th	3. ARIMA		Shumway and Stoffer
Sept. 23th	Lab 2: ARIMA		
Sept. 29th	4. Kalman filtering, EM	Problem set 2	Handouts 2 and 3, Bishop
Sept. 30th	Lab 3: Kalman filtering inference		
Oct. 6	5. Particle filtering		Handout 4
Oct. 7	Lab 4: Kalman EM		
Oct.13	6. Hidden Markov Models	Project proposals due	Handout 5, Bishop
Oct.14	Lab 5: Particle filtering		
Oct.20	7. Links between models, generalizations. Recap of key concepts for midterm	Problem set 3	Roweis & Ghahramani (1999), Bishop
Oct.21	Lab 6: HMMs		Handout 6, Bishop
Oct.27	**Mid-term exam**		
Oct.28	No lab		
Nov. 3rd	Introduction to Gaussian Processes		Rassmussen & Williams, Turner review
Nov. 4th	Lab 7: GP regression		
Nov.10th	9. GP advanced topics (guest lecture: A.Wilson)	Problem set 4	ICML 2020 tutorial on bayesian deep nets
Nov.11th	No lab, group work on projects		
Nov. 17th	10. Deep learning for time series		Handout 7
Nov. 18th	Lab 8: RNNs		
Nov.24th	11. Deep learning part 2, links between models		
Nov. 25th	No lab, group work on projects		
Dec. 1st	12. Spectral methods	Problem set 5	

Date	Lecture title	Assignments	Further reading
Dec. 2nd	Lab 9: Spectral methods		Shumway and Stoffer
Dec. 8th	**Final projects poster session**	Poster & written report due 1 week later	

Bibliography
There is no required textbook. Handouts provided for each section.

Core material

- *Time series analysis and its applications*, by Shumway and Stoffer, 4th edition (free online pdf)
- *Pattern recognition and machine learning*, Bishop
- *Gaussian processes,* Rassmussen & Williams

Policies
Students should try to solve problems on their own first. If stuck, one can discuss homework questions with colleagues, but should write up the final solution individually. Any violation will be penalized with a **zero** score for the assignment and referred to the DGS. Credit should be explicitly given for any external code use.

Late submission penalties: 20% points off for *each* day of delay.