Statistics Seminar

Friday, February 9, 12:00 pm - 1:00 pm CT
Zoom

Machine Learning for Time Series.
Part I: Pre-requisites.

Juana Sanchez, UCLA

Abstract: The deluge of data brought about by sensors and the Internet of Things has expanded the demand for time series analysis skills to include less experienced learners and researchers. But the proper analysis of time series data (data that is indexed by a time stamp and makes no sense without it) requires a skill set that is not usually included in the undergraduate curriculum. To speed up the execution of time series data analysis at large scale, contemporary practitioners and their methods have done the best they can do to convert the analysis of time series to something that may resemble our traditional regression models, random forests, or neural network models, methods that were not created for time series data but for iid data. That could make it easier for instructors to incorporate time series data analysis in the curriculum. The careful selection of features from a time series allows that: we just create a regular rectangular data set with variables that represent features of the single time series, for example. But feature extraction is easily understood and executed if the practitioner is already expert in time series data analysis. That expertise comes from the pre-data-deluge time series courses, rarely taught to undergraduate majors. Those courses prepared students to understand the nature of time series data, its autocorrelations, what is and is not important, and modeling and forecasting. That is still needed in order to make progress in contemporary time series data analysis, and it needs to be taught to undergraduate students to some extent. In fact, for some types of data, straightforward pre-data-deluge time series analysis would do very well. In this talk, we will cruise through the main issues to look at in time series data to prepare to do well informed machine learning for time series. At the same time, we will be offering an overview of the evolution of the R software for time series data analysis, and of undergraduate students' reactions to learning time series data analysis. Examples will illustrate all the points.

Bio: Dr. Juana Sanchez is a Senior Lecturer Emerita at UCLA Department of Statistics and Data Science. She is the author of "Time Series for Data Scientists" published by Cambridge University Press in 2023. (https://www.cambridge.org/gb/universitypress/subjects/statistics-probability/statistical-theory-and-methods/time-series-data-scientists-data-management-description-modeling-and-forecasting?format=HB) The book is accompanied by a web site, https://timeseriestime.org , where solutions to exercises and R programs used throughout the chapters and in the exercises, as well as errata sheet and supplements can be found. Juana has taught annually, since 1999, the course "Introduction to Time Series" to undergraduates at UCLA's Department of Statistics and Data Science. She has also published other books and numerous papers, is currently Editor of the Datasets and Stories section of the Journal of Statistics and Data Science Education, and has served in sections, chapters and committees of the ASA in the past. Throughout her career she has been actively engaged in the promotion of statistical literacy via national and international venues.
Webpage: http://statistics.ucla.edu/index.php/people1/all-faculty/7809-2/?smid=8829