Abstract

The analysis of “Big Data” through the application of a new breed of analytical tools for manipulating and analyzing vast caches of data, is one of the cutting edge new areas. As a byproduct of the extensive use of the internet in collecting data on economic transactions, such data are growing exponentially every day. Machine Learning as a field of computer science has strong ties to mathematical optimization and delivers methods, theory and applications. It indeed helps in developing high-performance computer tools, which often provide useful predictions in the presence of challenging computational needs. However, the result is one that we might call “pure prediction” and is not necessarily based on substantive knowledge. Also, typical assumptions such as the data being independent and identically (or at least independently) distributed, are not satisfactory when dealing with time stamped data, which is driven by multiple “predictors” or “features”. Starting with a structural model for time series, we use Bayesian tools for model fitting, prediction and feature selection, thus extending some recent works along these lines for the univariate case. The Bayesian paradigm in this multivariate setting helps the model avoid over-fitting, as well as captures correlations among multiple target time series with various state components. The model provides needed flexibility in selecting a different set of components and available predictors for each target series.


Zoom details can be found at: https://stt.natsci.msu.edu/stt-colloquium-zoom-info/

To request an interpreter or other accommodations for people with disabilities, please call the Department of Statistics and Probability at 517-355-9589.