ENGINEERING SCIENCES & APPLIED MATH--FEATURED SPEAKER

Building Predictive Models from Data: Examples from Sports and Public Health

Harish Bhat

School of Natural Sciences, University of California, Merced

February 12, 2018 • 4 pm – 5pm
M416 (ESAM Conference Room), Tech

Many modern problems in the statistical sciences call for predictive, data-driven models. To answer this call in real-world applications, several challenges must be met, including (for instance) non-normality, high-dimensionality, temporal dependence, and imbalanced data. In this talk, I will give two examples that illustrate these problems and corresponding solutions. In the first example, I will present a continuous-time Markov chain model for basketball games. In this model, each state corresponds to a unit of 5 players on the court, and transitions represent substitutions. Naive maximum likelihood estimation yields Markov chains with absorbing states, resulting in poor predictions of playing times for each unit. We develop and apply a sparsity-promoting linear programming method that removes absorbing states, improving long-term predictive performance. In the second example, I will present neural networks to predict adolescent suicide attempts. We train these models using anonymized electronic health records for over 500,000 unique California residents. I will describe a cumulative encoding to deal with patient records of varying length, and I will show how the model's predictive power increases as we obtain more data on an individual patient. For both problems, I will highlight common elements in the end-to-end procedures used to go from data to predictions, and how the results obtained thus far motivate future work.

Note: Cookies will be served at 3:30