UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

BIOSTATISTICS SEMINAR

Lorin Crawford
Department of Statistical Science
Duke University

Bayesian Approximate Kernel Regression with Variable Selection

Nonlinear kernel regression models are often used in statistics and machine learning due to greater accuracy than linear models. Variable selection for kernel regression models is a challenge partly because, unlike the linear regression setting, there is no clear concept of an effect size for regression coefficients. In this paper, we propose a novel framework that provides an analog of the effect size of each explanatory variable for Bayesian kernel regression models when the kernel is shift-invariant---for example the Gaussian kernel. We use function analytic properties of shift-invariant reproducing kernel Hilbert spaces (RKHS) to define a linear vector space that (1) captures nonlinear structure and (2) can be projected onto the original explanatory variables. The projection onto the original explanatory variables serves as the analog of effect sizes. The specific function analytic property we use is that shift-invariant kernel functions can be approximated via random Fourier bases. Based on the random Fourier expansion we propose a computationally efficient class of Bayesian approximate kernel regression (BAKR) models for both nonlinear regression and binary classification for which one can compute an analog of effect sizes. By adapting some classical results in compressive sensing we state conditions under which BAKR can recover a sparse set of effect sizes, simultaneous variable selection and regression. We illustrate the utility of BAKR by examining, in some detail, two important problems in statistical genetics: genomic selection (predicting phenotype from genotype) and association mapping (inference of significant variables or loci). State-of-the-art methods for genomic selection and association mapping are based on kernel regression and linear models, respectively. BAKR is the first method that is competitive in both settings. We will also outline how our proposed framework was used to reveal a MYC-driven transcriptional program in BRAF-mutant melanomas that have become resistant to MAP kinase (MAPK) inhibitors.

Thursday, September 8, 2016

3:30 pm - 4:30 pm

Blue Cross Blue Shield Auditorium

0001 Michael Hooker Research Center