

Special Colloquium

Probabilistic Approaches to Machine Learning on Tensor Data

Qing Mai (Florida State University)

Abstract: In contemporary scientific research, it is often of great interest to predict a categorical response based on a high-dimensional tensor (i.e. multi-dimensional array). Motivated by applications in science and engineering, we propose two probabilistic methods for machine learning on tensor data in the supervised and the unsupervised context, respectively. For supervised problems, we develop a comprehensive discriminant analysis model, called the CATCH model. The CATCH model integrates the information from the tensor and additional covariates to predict the categorical outcome with high accuracy. We further consider unsupervised problems, where no categorical response is available even on the training data. A doubly-enhanced EM (DEEM) algorithm is proposed for model-based tensor clustering, in which both the E-step and the M-step are carefully tailored for tensor data. CATCH and DEEM are developed under explicit statistical models with clear interpretations. They aggressively take advantage of the tensor structure and sparsity to tackle the new computational and statistical challenges arising from the intimidating tensor dimensions. Efficient algorithms are developed to solve the related optimization problems. Under mild conditions, CATCH and DEEM are shown to be consistent even when the dimension of each mode grows at an exponential rate of the sample size. Numerical studies also strongly support the application of CATCH and DEEM. Finally, we discuss how these developments in tensor data advance vector data analysis, such as differential network analysis.

Friday, January 31 at 3:00 PM in 636 SEO