THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

DEPARTMENT OF APPLIED MATHEMATICS
應 用 數 學 系

Dr. GUO Xin

Research interests include:

- Machine learning
- Mathematical biology

On-going ECS project

Mathematical analysis for coordinate kernel polynomial-based learning schemes that produce sparse approximations

Abstract

Learning theory considers problems of extracting pattern, model, and knowledge from data. For high dimensional data, the presence of inactive variables not only blurs the true model, but also spoils the accuracy of the inferred predictors. The proposed project focuses on the mathematical analysis of two promising kernel-based non-linear variable selection algorithms for removing irrelevant dimensions and identifying variable interactions. The first algorithm assigns one reproducing kernel to each variable and uses the Hadamard power of the weighted sum of the kernels to train the predictor. Sparsity enhancing regularization is employed to eliminate the weights of inactive dimensions. A surprising pilot simulation shows that with linear complexity, the algorithm is able to capture two- or even more-way interactions. The second algorithm uses the l^1 penalty on both the gradient and the Hessian. A data dependent hypothesis space is used to simplify computations. We will investigate the complexity of the reproducing kernel space based on coordinate kernel polynomials, and develop bound on covering numbers. Techniques from approximation theory will be applied. We will develop the theory of model selection consistency and the rate of convergence of the algorithms, of which two the relation will also be studied. We will develop mathematical foundation for the interaction selection consistency observed in simulation. For different coefficient vectors for the kernel polynomials, the relation between the corresponding reproducing kernel spaces will be investigated.