

Sensitivity Based Generalization Error for Supervised Learning Problems with Application in Feature Selection

Daniel S. Yeung

Professor, the School of Computer Science and Engineering,
South China University of Technology, Guangzhou, Guangdong

Abstract. Generalization error model provides a theoretical support for a classifier's performance in terms of prediction accuracy. However, existing models give very loose error bounds.

This explains why classification systems generally rely on experimental validation for their claims on prediction accuracy. In this talk we will revisit this problem and explore the idea of developing a new generalization error model based on the assumption that only prediction accuracy on unseen points in a neighbourhood of a training point will be considered, since it will be unreasonable to require a classifier to accurately predict unseen points "far away" from training samples. The new error model makes use of the concept of sensitivity measure for multiplayer feedforward neural networks (Multilayer Perceptrons or Radial Basis Function Neural Networks). The new model will be applied to the feature reduction problem for RBFNN classifiers. A number of experimental results using datasets such as the UCI, the 99 KDD Cup, and text categorization, will be presented.