Mathematics, Statistics, and Computer Science **@ UIC**

Statistics and Data Science Seminar

P-SVM: Efficient Parameter Selection for Support Vector Machines with Gaussian Kernels

Prof. Hsin-Hsiung Huang (University of Central Florida)

Abstract: Support Vector Machines (SVM) classifier is a popular classification method. However, most users may not well take tuning parameters selection because this step is time consuming. In practice, the tuning parameters are chosen by evaluating parameter candidates via cross validation. It is shown that the performance of SVM is sensitive to the values of tuning parameters. In some cases, SVM performs poorly due to the values of tuning parameters. However, selection of parameter values for SVM often relies on inefficient approaches such as extensive cross validation. To get around the problem, users may resort to anecdotal methods or default values set by software developers. However, these methods may compromise performance of classification accuracy. In this research, we propose an efficient algorithm called P-SVM for selecting the parameter pair, (gamma,C), of SVM with Gaussian kernels on metric data. P-SVM searches only a handful of percentiles of the squared Euclidean distances of data points to select the best pair of parameter values. Our motivation case study of business intelligence categorization demonstrates that P-SVM achieved a signi cant improvement in precision, recall, F-measure, and AUC from the default parameter values settled in Weka, a widely used data mining software. Applications of both simulation and publicly-available datasets also demonstrate that P-SVM achieves substantial improvement in computational time without loss of much classification accuracy.

Wednesday, April 26 at 4:00 PM in SEO 636