Cycle de Conférences du Laboratoire MIPS
2013-2014

16 mai 2014 à 10h30
Faculté des Sciences et Techniques FST6 (Amphis) – Salle 6.01

BAYESIAN ERROR ESTIMATION FOR
CLASSIFIER MODEL SELECTION

Dr Heikki Huttunen

Department of Signal Processing, Tampere University of Technology, Tampere, Finland
heikki.huttunen@tut.fi

The estimation of classification error is a critical step in classifier design, and closely related to model selection. Typical model selection procedures are either based on estimating the error (e.g., cross-validation, bootstrap, holdout, etc.) or information theoretic principles (e.g., AIC, BIC, MDL). The problem with the former approach is that the traditional counting-based estimators are both computationally expensive and inaccurate. On the other hand, the latter approach optimizes a measure that is not directly connected to the prediction error and often requires a careful selection of hyperparameters.

In this talk we concentrate on the recently proposed Bayesian Error Estimator (BEE), and on its uses for model selection among a family of generalized linear models. More specifically, we will show that the estimator is more accurate than the traditional error estimation approaches when selecting the best model along the regularization path of a LASSO regularized logistic regression model. Moreover, the BEE estimates the error directly from the training set, thus avoiding multiple training stages typical of cross-validation procedures.

As a case study, we will describe the anatomy of our submission into the IEEE MLSP 2013 Bird sound classification competition (https://www.kaggle.com/c/mlsp-2013-birds). The method was essentially a BEE-selected generalized linear model with BoW-like features calculated from a sparse dictionary representation calculated with the SPAMS toolbox developed by INRIA.

Biography:
Mr. Heikki Huttunen received the Dr.Eng. degree at Tampere University of Technology, Finland, in 1999. From 2003 until 2005 he was with Visy Oy, developing an automatic car license plate recognition system that is currently in use in four continents. Since 2005 he has held various positions in Tampere University of Technology. His research interests include computer vision and machine learning, and their practical applications, and is an author of over 90 scientific articles. In 2014 he received the EEPOS grant from the French Institute in Finland for research visits into French universities.

Recent publications:
T. Manninen, H. Huttunen, P. Ruusuvuori and M. Nykter, "Leukemia Prediction Using Sparse Logistic Regression," PLOS ONE 8(8): e72932. August 2013.

H. Huttunen, T. Manninen and J. Tohka, "Bayesian Error Estimation and Model Selection in Sparse Logistic Regression," IEEE MLSP, Southampton, UK, Sept. 2013.

Briggs et al., "The 9th annual MLSP competition: New methods for acoustic classification of multiple simultaneous bird species in a noisy environment," IEEE MLSP, Southampton, UK, Sept. 2013.

H. Kikuchi, S. Kataoka, S. Muramatsu, H. Huttunen, "Color-tone Similarity of Digital Images," IEEE ICIP, Melbourne, Australia, September 2013.

N. Aghaeepour, G. Finak, The FlowCAP Consortium, The DREAM Consortium, H. Hoos, T. Mosmann, R. Brinkman, R. Gottardo and R. Scheuermann, "Critical assessment of automated flow cytometry data analysis techniques," Nature Methods, February 2013.

H. Huttunen, T. Manninen, J-P. Kauppi and J. Tohka, "Mind Reading with Regularized Multinomial Logistic Regression," in Machine Vision and Applications, pp. 1-15, November 2012.

P. Ruusuvuori, T. Manninen, and H. Huttunen, "Image Segmentation Using Sparse Logistic Regression with Spatial Prior," in EUSIPCO 2012, Bucharest, Romania, August 27-31, 2012.

H. Huttunen, H. Oinonen, J. Selinummi, P. Ruusuvuori, and V. Voipio, "Polynomial Pulp Fiber Modeling," in EUSIPCO 2012, Bucharest, Romania, August 27-31, 2012.