Sparse Bayesian Modelling of categorial predictors

Helga Wagner

Department of Applied Statistics,
Johannes Kepler Universität, Linz

Sparse modelling and variable selection is one of the most important issues in regression type models, as in applications often a large number of covariates on comparably few subjects are available. Estimation of regression effects in such *large p, small n* problems is ill-conditioned: estimated regression effects typically have large standard errors, estimation results are instable and fitted models have no good predictive performance. To identify those regressors which have a non-negligible effect, many methods have been developed. In a Bayesian approach, variable selection methods often rely on specifying spike and slab priors on the regression effects. These priors are mixtures of two components: the spike is centered at zero with very small variance and the slab is comparably flat. The finite mixture structure allows classification of effects as (practically) zero or as non-zero.
Often covariates are categorial, measured either on an ordinal or a nominal scale. The usual strategy to include a categorial covariate in a regression type model is to define one of the levels as the baseline category and introduce dummy variables for all other levels. Hence the effect of a categorial covariate is not captured by a single but by a group of regression coefficients. Routine application of variable selection methods is inappropriate as these allow only selection of single regression coefficients. However, for categorical predictors sparsity cannot only be achieved by restricting coefficients to zero but also by fusing levels with essentially the same effect.
In this talk I will show how sparse modelling for the effect a categorial predictor can be achieved by appropriate prior distributions. To achieve fusion of effects we specify spike and slab prior distributions on all level effect differences. As a second approach we consider a modification of the standard spike and slab prior where the slab component is replaced by location mixture distribution. For both priors Bayesian inference relies on MCMC methods. Performance of these methods will be illustrated on simulated as well as real data.