Multivariate Bayesian variable selection and prediction

P. J. Brown[*], M. Vannucci[†] and T. Fearn[‡]

Abstract

The multivariate regression model is considered with p regressors. A latent vector with p binary entries serves to identify one of two types of regression coefficients, those close to zero and those not. Specialising our general distributional setting to the linear model with Gaussian errors and using natural conjugate prior distributions, we derive the marginal posterior distribution of the binary latent vector. Fast algorithms aid its direct computation, and in high dimensions these are supplemented by a Markov chain Monte Carlo approach to sampling from the known posterior distribution. Problems with hundreds of regressor variables become quite feasible. We give a simple method of assigning the hyperparameters of the prior distribution. The posterior predictive distribution is derived and the approach illustrated on compositional analysis of data involving three sugars with 160 near-infrared absorbances as regressors.

KEYWORDS: MULTIVARIATE REGRESSION; CONJUGATE DISTRIBUTIONS; BAYESIAN SELECTION; LATENT VARIABLES; PREDICTION; MARKOV CHAIN MONTE CARLO; MODEL AVERAGING.

[*]University of Kent at Canterbury, Institute of Mathematics and Statistics, Cornwallis Building, Canterbury, CT2 7NF, UK

[†]University of Kent at Canterbury

[‡]University College London, UK