

Model averaging for linear regression

Erkki P. Liski

University of Tampere, Finland

Abstract

Estimators formed after model selection really are like mixtures of many potential estimators. Sometimes it is advantageous to smooth estimators across several models, rather than rely only on the model that is suggested by a single selection criterion. The main theme of this paper is the problem of selecting of weights for averaging across least squares estimates obtained from a set of models. Existing model average (MA) methods are based on exponential AIC or BIC weights and on Mallows' criterion.

We propose selecting the weights by using the MDL (Minimum Description Length) criterion and compare this new model average estimator with those based on AIC and BIC weights. The model is homoscedastic linear regression and the set of potential explanatory variables is countable infinite. For MA estimation we construct first a sequence of approximating models and then derive an MA estimator of the regression coefficients as a convex combination of estimators for single models. The implied "hat" matrix P_w is symmetric but generally not idempotent. An MA estimator shows a certain resemblance to a shrinkage estimator and P_w has a central role in the structure of an MA estimator. Some of the properties of P_w are discussed.

Keywords

Approximating models, MDL criterion, Shrinkage estimator, Hat matrix.

References:

- Burnham, K.P. and D.R. Anderson (2002). *Model Selection and Multi-model Inference*. New York: Springer-Verlag.
- Hjort, L.H. and G. Claeskens (2003). Frequentist Model Average Estimators. *J. Amer. Statist. Assoc.* 98, 879–899.
- Liski, E.P. (2006). Normalized ML and the MDL Principle for Variable Selection in Linear Regression. In: Liski, E.P., Isotalo, J., Niemel, J., Puntanen, S., and Styan, G.P.H. (eds), *Festschrift for Tarmo Pukkila on His 60th Birthday* (pp. 159–172). Tampere: Department of Mathematics, Statistics and Philosophy.
- Rissanen, J. (2000). MDL Denoising. *IEEE Trans. Inform. Theory, IT-46, No. 1*, 2537–2543.