Parametric Empirical Bayes Model Selection - Some Theory, Methods and Simulation

Nitai Mukhopadhyay
Eli Lilly and Company

and

Jayanta Ghosh
Purdue University

Abstract

For nested models within the PEB framework of george and Foster (Biometrika,2000), we study the performance of AIC, BIC and several relatively new PEB rules under 0-1 and prediction loss, through asymptotics and simulation. By way of optimality we introduce a new notion of consistency for 0-1 loss and an oracle or lower bound for prediction loss.The BIC does badly , AIC does well for the prediction problem with least squares estimates. The structure and performance of PEB rules depend on the loss function. Properly chosen they rend to outperform other rules.

1 Introduction

Our starting point is a paper by George and Foster (2000), abbreviated henceforth as [6]. [6] propose a number of new methods using PEB (Parametric Empirical Bayes) ideas on model selection as a tool for selecting variables in a linear model. An attractive property of the new methods is that they use penalized likelihood rules with the penalty coefficient depending on data, unlike the classical AIC, due to Akaike (1973), and BIC, due to Schwartz (1978), which use constant penalty coefficients. The penalty for a model dimension q is usually λq, where λ is a penalty coefficient. [6] compare different methods through simulation.

Our major contribution is to supplement this with some theoretical work for both prediction loss and 0-1 loss. The former is supposed to be relevant in soft science, where one only wants to make good prediction, and the latter is relevant in hard science, where one wants to know the truth. It is known in model selection literature that these different goals lead to different notions of optimality.

Our theory is based on the assumption that we have nested , orthogonal models – a situation that would arise if one tries to fit an orthogonal polynomial of unknown degree. This special case receives special attention in [6].

Our paper is based on Chapter 4 of Mukhopadhyay (2000), subsequently referred to as [9]. A related paper is Berger, Ghosh and Mukhopadhyay, (2003), which shows the inadequacy of BIC in high dimensional problems.