Bayesian Analysis (2011) **6**, Number 1, pp. 43–48

Rejoinder

Nicholas G. Polson* and Steven L. Scott†

We thank all the discussants for their insights and comments on the article. Due to the subject matter specialization, *Bayesian Analysis* has a more homogeneous readership than journals that cater to a more general audience, so it is not surprising to find substantial agreement among the discussants and ourselves. Of course, readers may be disappointed by the lack of blood-sport normally associated with discussion articles. We apologize for this, and promise to write a more provocative article in the future.

1 Mallick et al.

Mallick *et al.* rightly point out that our focus on posterior inference for model parameters is only indirectly related to the classification performance that typically interests SVM users. The simulations provided by Mallick *et al.* are a welcome correction to our omission. The simulations show that the SVM criterion can in fact reduce the misclassification error compared to probit regression. Many Bayesians (including us) approach support vector machines with a wary suspicion that they are simply logistic regression's poor, non-probabilistic cousin. Simulations like this are useful data exercises that should force us to update that viewpoint. We have replicated the simulations in Table 1 with logistic regression in place of probit. The logistic regression and the SVM were both run, using spike-and-slab priors, on the spam data set from Section 5. We used the algorithm from Tüchler (2008) for the logit model.

Prediction is a common theme among the discussants. Lindley (1968) provides the theoretical analysis of prediction-based Bayesian variable selection in the presence of costs, as well as a beautiful discussion of the faults of commonly used classical procedures. The upshot is that, to select variables for a model that predicts best (in an MSE sense), one needs to find the linear combination that best fills in for the linear combination of variables that you leave out. Brown et al. (1998, 1999, 2002) illustrate the advantages of this framework in large scale predictive regression systems. This approach trades-off the cost of variable inclusion with the gain in MSE predictive power. We are not presently in a position to provide the equivalent predictive analysis for SVM's but Hans' proposal of basing prediction on the posterior mean via the linear combination $E(\beta|y)'x_f$ for a future covariate x_f seems sensible. Implementing the Lindley analysis requires some posterior standard errors, which we can directly obtain from our MCMC algorithm.

Another interesting direction for future research is showing the interplay between sparse estimators, variable selection, and prediction in the original Mallows (1973) C_p paper. That paper also contains a very useful discussion of the C_L criteria, corresponding to a linear Bayes ridge rule. Again our representation makes such a discussion

*Booth School of Business, Chicago, IL, mailto:ngp@chicagobooth.edu
†Google Corporation, mailto:stevescott@google.com

 DOI:10.1214/11-BA601REJ