Comment on Article by Polson and Scott

Babak Shahbaba^{*}, Yaming Yu[†] and David A. van Dyk[‡]

1 Introduction

Polson and Scott's paper presents the enlightening observation that the standard SVM can be embedded into a statistical latent variable model. This is aligned with other recent work, in which the penalty term in the convex optimization for several popular non-Bayesian models has been replaced by a prior distribution in order to develop an alternative Bayesian approach. See, for example, the Bayesian lasso model by Park and Casella (2008) and Hans (2009), and the Bayesian bridge regression model by Armagan (2009). Following the work of Andrews and Mallows (1974) and West (1987), the prior distributions used in these methods are expressed as scale mixtures of normal distributions. For example, in bridge regression (Frank and Friedman 1993), which includes both ridge regression (Hoerl and Kennard 1970) and lasso (Tibshirani 1996) as special cases, the regression parameters are estimated by minimizing the penalized residual sum of squares (using centered data),

$$\hat{\beta} = \arg\min_{\beta} \left[(y - X\beta)^T (y - X\beta) + \lambda \sum_{j=1}^p |\beta_j|^{\gamma} \right]$$

where $\beta = (\beta_1, \ldots, \beta_p)$. In the Bayesian framework, the penalty term can be replaced by a prior distribution of the form $P(\beta) \propto \exp(-\lambda |\beta_j|^{\gamma})$. When $0 < \gamma \leq 2$, the penalty can be represented as a scale mixture of normal distributions (West 1987).

The current paper follows a similar approach in its replacement of the regularization term in SVM with a prior distribution. The authors also followed similar steps to specify the likelihood since unlike ridge regression, the likelihood is not readily available for SVM. In particular, they insightfully replace the part of the objective function that depends on the data with $\exp[-2\sum_{i=1}^{n} \max(1-y_iX_i^T\beta, 0)]$ and use results from Andrews and Mallows (1974).

2 A Bayesian SVM model or a Bayesian model with SVM properties

While the authors presented the critical first step of formulating a Bayesian model that encompasses SVM, in general it is not necessary to limit our choice of prior distributions in a Bayesian model by forcing mathematical compatibility with the penalty term in the

© 2011 International Society for Bayesian Analysis

^{*}Department of Statistics, University of California, Irvine, CA, mailto:babaks@uci.edu

[†]Department of Statistics, University of California, Irvine, CA, mailto:yamingy@ics.uci.edu

[‡]Department of Statistics, University of California, Irvine, CA, mailto:dvd@ics.uci.edu