Bayesian Model-Assisted PRIM Algorithm

Longyang Wu and Hugh Chipman
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario N2L 3G1 Canada

The patient rule-induction method (PRIM) is a statistical learning method that seeks to locate regions in the feature space where the response variable has a high value. In this paper we present a Bayesian model-assisted PRIM algorithm. This algorithm can automatically cover promising regions in the feature space based on Bayes factor values and marginal posterior probabilities. The model-assisted part of the method begins with the specification of a flexible mean-variance shift model for quantitative response variables. This enables predictions, and has the potential to achieve even higher prediction power through the Bayesian model averaging. A simulated example is provided to illustrate the effectiveness of this new algorithm.

Introduction

The patient rule induction method (PRIM) is a novel statistical learning method for exploration of large high-dimensional data, proposed by Friedman and Fisher (1999). The PRIM algorithm has many potential predictive modelling applications in drug discovery, medical diagnostics, and so forth. Its goal is to locate hyper rectangular regions or "boxes" in the feature space where the response variable has a high value. Each box is defined by upper and lower bounds on one or more of the explanatory variables.

The algorithm identifies boxes sequentially. After finding one box, the data contained in that box is removed from the training data. The algorithm then begins to seek a new box in the remaining data. The search for each box consists of two steps: top-down peeling and bottom-up pasting. Top-down peeling starts with a candidate box containing all training data. It then compresses this candidate box along an upper or lower bound of an explanatory variable by a small amount to form a new candidate box, and provisionally peels off observations falling outside the new candidate box. The explanatory variable chosen for compression is the one resulting in the candidate box with the highest average, after the compression is performed. This top-down peeling process continues until the number of observation within the candidate box is below a predetermined threshold. Bottom-up pasting, the reverse of the top-down peeling, is applied to the minimal candidate box to expand along any boundary. Because top-down peeling is a greedy search, such an expansion is often possible. Bottom-up pasting continues until the expansion fails to increase the average. The analyst has to exercise expert knowledge to choose one candidate box from peeling trajectory, generated by the peeling and pasting, and repeat this process until there appears to be no boxes left.

The PRIM algorithm makes few statistical assumptions and does not explicitly state a model. Titterington (1999) and Smyth (1999) suggest, but do not implement, Bayesian approaches to formulate such a model.

We propose a Bayesian model-assisted PRIM algorithm, including a model for observations inside and outside boxes. The model resembles that used in Bayesian CART (Chipman et. al. 1998). The algorithm to identify boxes will be based on a search strategy via bumping (Tibshirani and Knight, 1999).

Motivation via Simulated Data

To illustrate the form of a PRIM model, we simulate from a simple two-dimensional predictive model. The underlying model for simulation is:

$$Y = f(x_1, x_2) + \epsilon, \tag{1}$$

where $x_1, x_2 \sim$ continuous Uniform(0, 12); $\epsilon \sim N(0,1)$, and

$$f(x_1,x_2) = \begin{cases} 10, \text{ if } 1 \le x_1 \le 4 \ \& \ 1 \le x_2 \le 4 & [B_1] \\ 10, \text{ if } 4 \le x_1 \le 7 \ \& \ 4 \le x_2 \le 7 & [B_2] \\ 10, \text{ if } 7 \le x_1 \le 10 \ \& \ 1 \le x_2 \le 4 & [B_3] \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

Note that x_1 and x_2 are rounded to one decimal place in the following analysis.

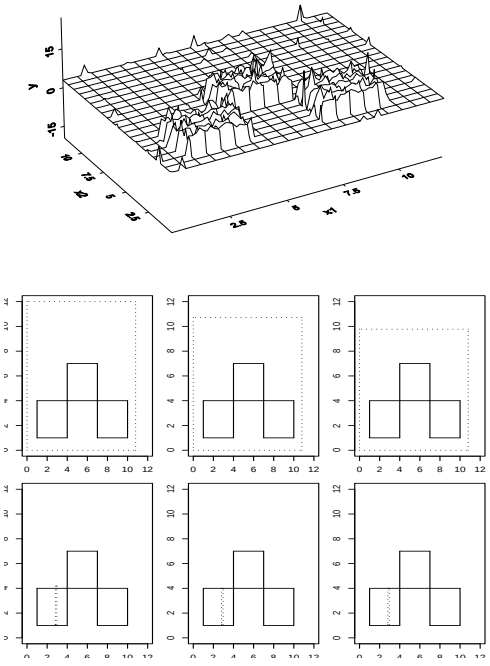

Figure 1: Top: the 3-dimensional plot of the podium data; Bottom: illustration of the PRIM algorithm.

The three boxes collectively define a "solution region" $R = \bigcup_{k=1}^{3} B_k$. The term solution region refers to the union of all boxes that constitute the model. The upper panel of Figure 1 displays 6,000 observations generated by the above model. The projection of $f(x_1, x_2)$ on the x_1 and x_2 plane resembles a podium, hence we refer to it as the podium data. The lower panel of Figure 1 shows six candidate boxes from a single peeling trajectory for inducing B_1. The first three are in the upper row, and the last three boxes are in the lower row.

The simulated model (2) is a PRIM model, with three boxes and the remainder of the predictor space a "background box". In this case, the boxes are disjoint, although in general they are permitted to intersect.

A Model for PRIM

To formulate a model, we need to be clear about what a prediction from PRIM is. The PRIM algorithm covers and removes the K boxes sequentially, i.e. the data points removed in B_1 are not available for B_2,...,B_K. Thus the prediction for any point $\mathbf{x}_i$ is based on the first box in the ordered list $B_1, B_2, \ldots, B_K, B_0$. The background box, B_0, catches all data points not in the solution region.

We use $k = 0, 1, ..., K$, to label the box membership of a response variable value. Let y_{kj} and $\mathbf{x}_{kj}$ denote the response and predictor vector of the jth observation removed by the kth box. Data points removed by the box B_k can be modelled by a distribution $p(y|\theta_k, \mathbf{x} \in B_k)$. The probability function represents a parametric family indexed by $\Theta = \{\theta_0, \theta_1, \theta_2, ...\theta_K\}$. Define $Y = (Y_0, Y_1, ..., Y_K)^{'}$, where $Y_k = (y_{k1}, y_{k2}, ..., y_{kn_k})$ is the vector of response variable values in box k. We assume that the values of $Y_0, Y_1, ..., Y_K$ are conditionally independent given (Θ, R, K) and y values within each box are iid. In this case, a general global model for PRIM is of the form

$$\begin{aligned} p(Y|X, \Theta, R, K) &= \prod_{k=0}^{K} p(Y_k|\theta_k, \mathbf{x} \in B_k) \\ &= \prod_{k=0}^{K} \prod_{j=1}^{n_k} p(y_{kj}|\theta_k, \mathbf{x} \in B_k), \quad (3) \end{aligned}$$

where $n_k, k = 0, ..., K$ is the number of observations inside the kth box.

For quantitative y values, the responses are assumed to be normal:

$$y_{k1}, ..., y_{kn_k}|\theta_k \text{ iid } \sim N(\mu_k, \sigma_k^2),\ k = 0, ..., K. \quad (4)$$

This model allows both the box mean μ and variance σ^2 vary across boxes.

A Bayesian Formulation for the PRIM Model

Since the PRIM model can be identified by (Θ, R, K), a Bayesian analysis of this model proceeds by specifying a prior probability distribution $p(\Theta, R, K)$. It is convenient to use the relationship $p(\Theta, R, K) = p(\Theta|R, K)p(R|K)p(K)$ and specify $p(K), p(R|K)$, and $p(\Theta|R, K)$ separately.

We assign a Poisson prior to the number of boxes, K, in the solution region. The prior probability of the solution region R is the joint distribution of these K boxes: $p(R|K) = p(B_1)p(B_2|B_1) \cdots p(B_K|B_1, B_2, ..., B_{K-1})$, since boxes found in early steps are not available as candidate boxes in the later steps. This random sampling without replacement process can be approximated by random sampling with replacement. The approximation is reasonable because the extremely large sample space for boxes.

If we assume independence among boxes, the prior $p(R|K)$ is $\prod_{k=1}^{K} p(B_k)$. We specify $p(B_k)$ in three stages. First, note that boxes can be grouped according to the number of constrained explanatory variables (out of a total of P variables). Without prior knowledge, we assume that a box has the equal probability $1/P$ belonging to one of the P groups. Second, if box B_k belongs to group m, then there are $\binom{P}{m}$ ways to choose m out of P explanatory variables to define box boundaries. We assume that each combination of boundaries occurs with equal probability of $\binom{P}{m}^{-1}$. Finally, we assume that the boundaries on the m explanatory variables are independent, and assign equal probability to all possible pairs of upper and lower boundary, excluding the pair $\{\min(x_j), \max(x_j)\}$. A data-dependent prior is used, assuming that only unique observed values along each predictor can be boundaries. Thus if variable x_j has u_j unique variables, there are $\binom{u_j}{2}-1$ possible boundary pairs.

Combining the three elements gives a prior probability distribution of B_k as the joint distribution of its group type g_k, its boundary combination c_k, and its m boundaries, s_{k1}, s_{k2}, ...,s_{km} as follows:

$$\begin{aligned} p(B_k) &= p(g_k, c_k, s_{k1}, s_{k2}, ..., s_{km}) \\ &= p(g_k)p(c_k|g_k)p(s_{k1}, s_{k2}, ..., s_{km}|g_k, c_k) \\ &= \frac{1}{P\binom{P}{m}} \prod_{i=1}^{m} \left[\binom{u_i}{2} - 1 \right]^{-1} . \end{aligned} \quad (5)$$

For the mean-variance shift model (4), perhaps the simplest prior specification is the standard conjugate form

$$\mu_0|\sigma_0^2, R, K \sim N(\tau, \sigma_0^2) \quad (6)$$

$$\mu_k|\sigma_k^2, R, K \sim N(\eta, c\sigma_k^2), \quad k = 1, ..., K \quad (7)$$

The scale parameter, $\sigma_k^2(\sigma_0^2)$, will take the standard conjugate prior distribution of a scaled inverse chi-square distribution as follows:

$$\sigma_0^2, \sigma_1^2, ..., \sigma_K^2|R, K \sim \frac{\nu\lambda}{\chi_\nu^2} , \quad (8)$$

where $\nu(\lambda)$ is the degrees of freedom (scale parameter). The pairs $(\mu_0, \sigma_0^2), ..., (\mu_K, \sigma_K^2)$ are assumed to be independently distributed. The additional parameter c ($c > 0$) in (7) allows for more flexibility in the model. There are situations that we might believe the values of response variable for observations inside the solution region has less variation than the observations outside the solution region (inside the background box). Then, we can incorporate this prior belief by taking the c value less than one. In the podium data we choose c equals one because of the absence of such prior belief.

A common interpretation of ν and λ in a scaled inverse chi-squared distribution is that they provide prior information about equivalent to v observations with averaged squared standard deviation λ. We choose v to be 5 in this study, a reasonably uninformative prior. Choice of λ is based on the belief that the error standard deviation within each box (including background box) ranges from being significantly smaller than the unconstrained sample standard deviation s, to being about the same magnitude. As in Chipman et al. (2002), we choose a quantile q such that $Pr(\sigma < s) = q$, and then use the implied value of λ, $\lambda = s^2\Phi_v^{-1}(1-q)/v$ where Φ_v is the cumulative distribution function for the chi-squared distribution with v degree of freedom. We tried the value $q = 0.9$ for podium data which corresponds to $\lambda = .1109s^2$.

After choosing v and λ, it is crucial that the choices of τ and η values should properly reflect the averages for background box and boxes where the response variable has a high value. If no prior knowledge about box averages is available, τ and η can be chosen as upper and lower quantile of the observed values. For the podium data, η is set to the 90th quantile (9.944) and τ the 50th quantile (0.323).

The Marginal Likelihood $p(Y|\mathbf{x},R,K)$

Under the general formulation of the PRIM model, the marginal likelihood of $p(Y|\mathbf{x},R,K)$ can be obtained by integrating out the model parameter Θ from (3) as follows:

$$\begin{aligned} p(Y|\mathbf{x},R,K) &= \int p(Y|\mathbf{x},\Theta,R,K)p(\Theta|R,K)d\Theta \\ &= \int \prod_{j=1}^{n_0} p(y_{0j}|\theta_0)p(\theta_0|R,K)d\theta_0 \\ &\quad \times \prod_{k=1}^{K} \int \prod_{j=1}^{n_k} p(y_{kj}|\theta_k)p(\theta_k|R,K)d\theta_k \end{aligned} \quad (9)$$

Conjugate priors (6), (7), and (8) enable analytic integration to obtain the marginal likelihood (Chipman et al. 1998):

$$
\begin{aligned}
p(Y|\mathbf{x},R,K) &= \frac{(\lambda\nu)^{-\frac{(K+1)\nu}{2}}\pi^{-\frac{n}{2}}}{\Gamma(\frac{\nu}{2})^{(K+1)}}\frac{\Gamma(\frac{\upsilon+n_0}{2})}{\sqrt{n_0+1}} \\
&\quad \times\left[\frac{n_0(\overline{y}_0-\tau)^2}{n_0+1}+s_0+\lambda\nu\right]^{-\frac{n_0+\nu}{2}} \\
&\quad \times\prod_{k=1}^{K}(\frac{\Gamma(\frac{\upsilon+n_k}{2})}{\sqrt{cn_k+1}}\times \qquad (10)\\
&\quad \left[\frac{n_k(\overline{y}_k-\eta)^2}{cn_k+1}+s_k+\lambda\nu\right]^{-\frac{n_k+\nu}{2}}),
\end{aligned}
$$

where $n=n_0+n_1+n_2+\cdots+n_K$ is the total number of observations in the dataset; $\overline{y}_0,\overline{y}_1,...,\overline{y}_K$ are the box averages for the background box and boxes in the solution region respectively. s_k is (n_k-1) times the sample variance of Y_k values for $k=0,1,\ldots,K$.

Choosing Box from the Peeling Trajectory

The Bayesian model-assisted PRIM algorithm provides an automated approach to choose one box from a peeling trajectory, based on Bayes factor values and marginal posterior probability for the solution region.

The marginal likelihood, $p(Y|\mathbf{x}\in\widetilde{B}_k,T_j)$, of n_k observations inside $\widetilde{B}_k$, where $\widetilde{B}_k$ is the kth box in the peeling trajectory T_j, is required to compute a Bayes factor. For the mean-variance shift model, it can be readily obtained by integrating out the parameter θ_k, after combining the sampling distribution (4), and prior distributions (7) and (8) as follows:

$$
\begin{aligned}
p(Y|\mathbf{x} \in \widetilde{B}_k,T_j) &= \int\prod_{j=1}^{n_k}p(y_{kj}|\theta_k)p(\theta_k|T_j)d\theta_k \\
&= \frac{(\lambda\nu)^{-\frac{\upsilon}{2}}\pi^{-\frac{n_k}{2}}}{\Gamma(\frac{\nu}{2})}\frac{\Gamma(\frac{\upsilon+n_k}{2})}{\sqrt{cn_k+1}} \qquad (11)\\
&\quad \times\left[\frac{n_k(\overline{y}_k-\eta)^2}{cn_k+1}+s_k+\lambda\nu\right]^{-\frac{n_k+\nu}{2}}.
\end{aligned}
$$

For observations inside box $\widetilde{B}_k$ in the peeling/pasting trajectory T_j, there are two competing models:

$$
\begin{aligned}
M_0 &: \text{ it is a background box with average of } \tau \\
M_1 &: \text{ it is a target box with average of } \eta
\end{aligned}
$$

The prior probabilities for M_0 and M_1 given T_j are $p(M_0|T_j)$ and $p(M_1|T_j)$ respectively. For boxes in the same peeling/pasting trajectory T_j, we assume that $p(M_0|T_j)=p(M_1|T_j)=\frac{1}{2}$, representing the absence of a prior preference for either model.

Based on marginal likelihood $p(Y|\mathbf{X}\in\widetilde{B}_k,T_j)$ from (11), the Bayes factor for M_1 against M_0 denoted as B_{10} (Kass and Raftery, 1995), can be readily obtained as follows:

$$
\begin{aligned}
B_{10} &= \frac{p(M_1|Y,\mathbf{x}\in\widetilde{B}_k,T_j)}{p(M_0|Y,\mathbf{x}\in\widetilde{B}_k,T_j)}\times\frac{p(M_0|T_j)}{p(M_1|T_j)} \\
&= \frac{p(Y|\mathbf{x}\in\widetilde{B}_k,M_1,T_j)}{p(Y|\mathbf{x}\in\widetilde{B}_k,M_0,T_j)} \qquad (12)\\
&= \left[\frac{(\overline{y}_k-\tau)^2+(\frac{cn_k+1}{n_k})(s_k+\lambda\nu)}{(\overline{y}_k-\eta)^2+(\frac{cn_k+1}{n_k})(s_k+\lambda\nu)}\right]^{\frac{n_k+\nu}{2}}.
\end{aligned}
$$

If the box average is close to τ, the Bayes factor is less than 1 and in favor of M_0; otherwise, if the box average is close to η, the Bayes factor is in favor of M_1. The box with the highest Bayes factor value is the most probable box. However, because of the trade-off between the box average and the box support (the proportion of observations inside a box), there are occasions that even the "true" box may not have the highest Bayes factor value in the peeling trajectory. This can be remedied by considering the marginal posterior probabilities for those candidate boxes with high Bayes factor values.

Given boxes in the current trajectory, the primary goal is to choose the optimal box from a list of candidate boxes. Therefore, we can treat each candidate box as a tentative solution region and evaluate its marginal posterior probability over the entire dataset. We can build a list of candidate boxes from the current trajectory T_j by arranging these boxes in desending order of their Bayes factor values. We denote the box with the highest Bayes factor value in the list as $\widetilde{R}_1$, the box with the second highest value as $\widetilde{R}_2$ and so forth. Then, we have a sequence of d boxes $\widetilde{R}_1,\cdots,\widetilde{R}_{d-1},\widetilde{R}_d$ where $d\geq 1$. If we treat each of them as the tentative solution region, then the marginal posterior probability $p(\widetilde{R}_i|\mathbf{x}\in\widetilde{R}_i)$ for the ith box is as follows:

$$
\begin{aligned}
p(\widetilde{R}_i|\mathbf{x}\in\widetilde{R}_i) &= p(\widetilde{R}_i)\int p(Y_0|\theta_0,\mathbf{x}\in\widetilde{R}_i^c)d\theta_0 \\
&\quad \times\int p(Y_1|\theta_1,\mathbf{x}\in\widetilde{R}_i)d\theta_1
\end{aligned}
$$

where $\widetilde{R}_i^c$ is the complement of the ith tentatively solution region, e.g. the tentatively background box; $\widetilde{R}_i$ is the ith tentatively solution region, and θ_0 and θ_1 are model parameter vectors for background box

and solution region. This marginal posterior probability can be obtained analytically from (10) by setting $K = 1$. The prior probability, $p(\widetilde{R}_i)$, for the single-box solution region is the same as the prior probability for the box $\widetilde{R}_i$.

Starting with the one with the highest Bayes factor value, we calculate and keep track of the marginal posterior probabilities of candidate boxes. The candidate box, from which the marginal posterior probability fails to increase, will be chosen as the optimal candidate box from the current list.

Bayesian Model-Assisted PRIM Algorithm

In this section, we propose a Bayesian model-assisted PRIM algorithm. Table 1 is the Bayesian version of the PRIM algorithm (Hastie et al., 2001). This model-assisted algorithm employs Bayes factor and marginal posterior probability to automatically select one box from the peeling trajectory. The new algorithm also provides a stopping criteria to terminate the algorithm when the Bayes factor value is less than unity.

The new algorithm in Table 1 can be further improved by the adoption of a stochastic search through multiple bumping. Starting with the training data set D, the algorithm produces multiple bootstrap samples and builds a peeling trajectory from each sample. Suppose we have J bootstrap samples. Let T_j be the trajectory from the jth sample and the box B_j^* be the best candidate box from the trajectory T_j by the criteria of Bayes factor value and marginal posterior probability. Then we have a group of boxes, B_j^*, $j = 1, 2, ..., J$, to compete for the estimation of the first box in the solution region.

A new candidate box list can be formed for these J boxes by ranking their Bayes factor values descendingly. Then, the marginal posterior probability of each candidate box is calculated. The box B_1, which we choose as the estimate for the first box in the solution region, is the one from which the marginal posterior probability of candidate box fails to increase. The training data set D is updated by removing observations inside B_1. The second box is covered in a similar fashion after the first box is removed, and so forth.

Bayesian model-assisted PRIM algorithm
1. Start with the maximal box containing all training data.
2. Peel off a small portion according to the peeling criteria.
3. Compute Bayes factor for the resultant box, and update the peeling candidates list based on Bayes factor value in descending order.
4. Repeat step 2 and 3 until some minimal number of observations remain in the box.
5. Compute the marginal posterior probabilities for boxes at the top of the list, and choose the one from which the marginal posterior probability fails to increase. Call the box B_1.
6. Remove the data in box B_1 from the training data and repeat steps 2-5 to obtain a second box, and continue until the highest Bayes factor for new candidates is less then 1.

Table 1: Bayesian model-assisted PRIM algorithm.

Results and Discussion

We applied this model-assisted algorithm to a sample of 6000 observations generated from the model (2). We use the mean-variance shift model (4) for observations inside and outside the solution region. The conjugate priors (6), (7) and (8) are adopted, and their hyperparameter values are chosen based on the data. Ten bootstrapped peeling trajectories are employed to identify a single box. We combine the Bayes factor values and marginal posterior probabilities for boxes in the candidate list and choose one box to be removed. Figure 2 displays the log of Bayes factor values vs. box averages for ten boxes. Each of these ten was the result of ten bumped datasets. Seven boxes that have significantly higher log Bayes factor values ($\geq$15) are chosen for further analysis.

Generally, if found boxes are nested or overlapped, they might be further grouped into different categories to reduce redundancies. We need the measurement of interbox dissimilarity (Friedman and Fisher, 1999) to group boxes. Two boxes are considered to be similar and put into same group if the interbox dissimilarity between the two is close to zero. After box grouping, we can obtain the minimal coverage box for each group. The boundaries of the minimal coverage box are the union of corresponding boundaries of boxes in the same group. The marginal posterior probabilities of all boxes (including the minimal coverage box) in the same group are calculated and the one with highest marginal posterior probability is chosen as the optimal box in the group.

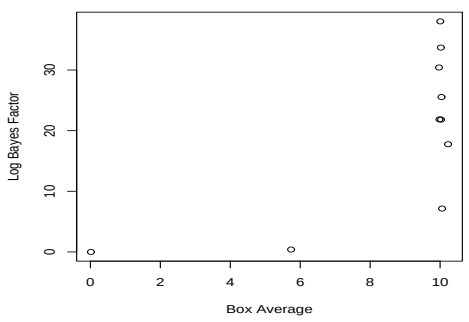

Figure 2: Log of Bayes factor vs. box average for the nine found boxes and the background box in podium data.

True Box	Estimated Box
$1.0 \leq x_1 \leq 4.0$ $1.0 \leq x_2 \leq 4.0$	$1.0 \leq x_1 \leq 4.0$ $1.0 \leq x_2 \leq 4.0$
$4.0 \leq x_1 \leq 7.0$ $4.0 \leq x_2 \leq 7.0$	$4.0 \leq x_1 \leq 7.0$ $4.0 \leq x_2 \leq 6.9$
$7.0 \leq x_1 \leq 10.0$ $1.0 \leq x_2 \leq 4.0$	$7.0 \leq x_1 \leq 10.0$ $1.0 \leq x_2 \leq 4.0$

Table 2: The comparison of the true and the estimated solution region in the podium data.

In the case of the podium data, the three groupings of the seven boxes are very obvious. The minimal coverage box for each group can be obtained easily as well. For the podium data, the minimal coverage box in each group has the highest marginal posterior probability and is chosen as the estimate for box in the solution region. Table 2 presents the estimates obtained by the Bayesian model-assisted PRIM algorithm. It seems that this model-assisted approach is very effective and finds all boxes in the solution region for the podium data.

It is not required to specify the hyperparameter in the Poisson prior for the number of boxes, K, if the aim is only to find the solution region as we did in the forgoing analysis of podium data. Such a specification, however, is usually necessary if we want to do prediction or Bayesian model averaging. Results from early application of PRIM (e.g. Friedman and Fisher, 1999) are helpful in this regard.

One advantage of Bayesian model-assisted PRIM is to facilitate automatic implementation of PRIM and avoid the frequent user interactions which may be indispensable in the non-model based PRIM. In turn, this will provide more accurate approximations to regions where the response variable has the maxima. This model-assisted algorithm employs Bayes factor to avoid the peeling/pasting trajectory being trapped into local maxima when the response variable has multiple maximum regions. The marginal posterior probabilities calculated for top candidate boxes from a single trajectory allows the algorithm to automatically choose the optimal box.

One potential important application of this mode-assisted PRIM is the AIDS Antiviral Database from National Cancer Institute (NCI). In this data, there are 29,812 chemical compounds, of which only 608 are active against the HIV virus. Results from our preliminary analysis are promising. We plan to carry out a thorough analysis and report the results elsewhere.

References

[1] Chipman, H. A., George, E. I., and McCulloch, R. E. (1998), "Bayesian CART Model Search (with discussion)", *Journal of the American Statistical Association*, 93, 935-960.

[2] Chipman, H., George, E. and McCulloch, R. (2002),"Bayesian Treed Models", *Machine Learning*, 48, 299-320.

[3] Friedman, J. and Fisher, N. (1999). "Bump hunting in high-dimensional data" (with discussion), *Statistics and Computing*, *9 123-143*.

[4] Hastie, T., Tibshirani, R. and Friedman, J (2001), *The elements of Statistical Learning, Data Mining, Inference, and Prediction*, Springer.

[5] Kass, R. and Raftery, A. (1995). "Bayes factors". *Journal of the American Statistical Association*, 90, 773–795.

[6] Smyth, P. (1999) "Discussion of 'Bump hunting in high-dimensional data"' *Statistics and Computing*, *9*,149-150.

[7] Tibshirani, R. and Knight, K. (1999) "Model Search and Inference by Bootstrap 'Bumping"', *Journal of Computational and Graphical Statistics*, 8, 671-686.

[8] Titterington, D. M. (1999) "Discussion of 'Bump hunting in high-dimensional data"' *Statistics and Computing*, 148-149.