Model Selection Prediction for the Mixture of Gaussian Processes with RJMCMC

Zhe Qiang, Jinwen Ma

Model Selection Prediction for the Mixture of Gaussian Processes with RJMCMC

Zhe Qiang[1] and Jinwen Ma[2]

[1] Department of Mathematics, Northwest University, Xi'an, Shaanxi Province, 710069, P. R. China
zhe.qng@nwu.edu.cn
[2] Department of Information Science, School of Mathematical Sciences & LMAM, Peking University, Beijing, 100871, P. R. China
jwma@math.pku.edu.cn

Abstract. Repetition measurements from different sources often occur in data analysis which need to be model and keep track of the original sources. Moreover, data are usually collected as finite vectors which need to be considered as a sample from some certain continuous signal. Actually, these collected finite vectors can be effectively modeled by the mixture of Gaussian processes (MGP) and the key problem is how to make model selection on a given dataset. In fact, model selection prediction of MGP has been investigated by the RJMCMC method. However, the split and merge formula of the RJMCMC method are designed only for the univariables in the past. In this paper, we extend the split and merge formula to the situation of the multivariables. Moreover, we add a Metropolis-Hastings update rule after the RJMCMC process to speed up the convergence. It is demonstrated by simulation experiments that our improved RJMCMC method is feasible and effective.

Key words: RJMCMC, Mixture of Gaussian processes, multivariable regression.

1 Introduction

In real applications, data often come as measure curves or time series. Although these are gathered as finite points, it is valuable to consider them as sample or trajectories of stochastic processes. In [1], the batch curves are considered from a mixture of Gaussian processes to describe its heterogeneity. However, the problem of well-known model selection for the mixture model comes, which can be solved with BIC criterion [2]. But this method can only make inference about the parameters in a fixed dimension parameter subspace and is weak on exploring the low probability area between the neighbourhood subspaces.

Reversible Jump Markov Chain Monte Carlo(RJMCMC)[3] is one of the most important method among Markov chain Monte Carlo simulation methods. It can not only make inference about parameters but also make model selection prediction. Since this simulation method explores the different parameter

subspaces with varied dimension, it can prevent the solution to be trapped into local optimum. In [5], we proposed the RJMCMC method for automatic model selection or prediction for mixtures of Gaussian processes and demonstrated its effectiveness. Based on it, we try to extend the split and merge formula of the RJMCMC method from the single-input case to the multi-input case to adapt to more complex or general situation.

our paper is organized as follows. We introduce the mixture of Gaussian processes and their latent variables in Section 2. Section 3 reviews the five type of moves of the parameters and extends the split and formula to the multi-input situation. We demonstrate our improved RJMCMC on simulated data in section 4. Finally, we conclude briefly in section 5.

2 The Mixture of Gaussian Processes and Its Latent Variables

Suppose that M curves are given and the response $y_m(t)$ for the m-th curve is defined as:

$$y_m(t) = \tau_m(t) + \epsilon_m(t) \tag{1}$$

where $\tau_m(t)$ is a Gaussian process and $\epsilon_m(t)$ is a noise term. For the $m-$th curve, the time points at which we collect data are $(t_{m,1}, \cdots, t_{m,N_m})$ and the target values are $(y_{m,1}, \cdots, y_{m,N_m}) = (y_m(t_{m,1}), \cdots, y_m(t_{m,N_m}))$, where N_m is the total number of observations on this curve. Then this curve can be considered as a sample curve or trajectory of the corresponding Gaussian process [1].

Moreover, data are assumed coming from different source and mixture structure are constructed to model the heterogeneity. For the dataset $\{(x_{m,n}, y_{m,n}), m = 1, \cdots, M, n = 1, \cdots, N_m\}$, the likelihood is as follows

$$L(\boldsymbol{\theta}, \boldsymbol{p}|Y, X) = \prod_{m=1}^{M} \sum_{k=1}^{K} p_k \mathcal{N}(Y; \mu_k(X), \Sigma_k(X)) \tag{2}$$

where K is the total number of the mixtures, μ_k and Σ_k are the mean function and covariance function. Specially, the covariance function is defined as follows:

$$\Sigma_k(x_i, x_j; \theta_k) = v_k \exp\left(-w_k \frac{(x_i - x_j)^2}{2}\right) + \delta_{ij}\sigma_k^2 \tag{3}$$

here $\theta_k = (v_k, w_k, \sigma_k^2)$ and each parameters are positive on $\mathbb{R}$.

As usual, auxiliary variables [6] z_m is the index variable and $z_m = k$ indicate the m-th curve belong to the k-th component. Then the completion of model is:

$$Z_m \sim \mathcal{M}_k(1; \underbrace{\frac{1}{n}, \cdots, \frac{1}{n}}_{K repetitions}), \quad y_m|x_m, z_m, \boldsymbol{\theta} \sim \mathcal{N}(\mu_k(x_m), \Sigma(x_m; \theta_k)|z_m = k) \tag{4}$$

From the Bayesian point of view, the prior of parameters should also be set, and the posterior is proportional to the prior multiples likelihood. For convenience of the derivation, the prior is set to be the conjugate prior:

$$w_k \sim I\Gamma(\frac{1}{2}, \frac{1}{2}), \quad v_k \sim \mathcal{LN}(-1, 1^2), \quad \sigma_k^2 \sim \mathcal{LN}(-3, 3^2) \tag{5}$$

where $I\Gamma$ represents inverse gamma distribution and $\mathcal{LN}$ represents log normal distribution. The mixed proportion coefficient $\boldsymbol{\pi}$ are from Dirichlet distribution:

$$(\pi_1, \cdots, \pi_K) \sim Dir(1, \cdots, 1) \tag{6}$$

3 Model Selection Prediction with Reversible Jump Markov Chains Monte Carlo

Model selection is an important and difficulty problem on the mixture model. Moreover, the valley among the neibourhood modes cannot be detected with a general parameter estimation method. Thus, we utilize the Reversible Jump MCMC model to overcome this difficulty.

In fact, our algorithm contains five types of parameter moves:

(a) $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_K)$;
(b) $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_K)$, where $\theta_k = (v_k, w_k, \sigma_k^2)$;
(c) $\boldsymbol{z} = (z_1, \cdots, z_M)$;
(d) split and merge move;
(e) Metropolis-Hastings update.

Actually, the first three belong to the moves with the component number fixed and step (d) is the move with the component number varied, step (e) is also the move with the component number fixed.

3.1 The Moves with the Component Number Fixed for (a)(b)(c)

For the moves with the component bunber fixed, gibbs sampling and hybrid MCMC (one kind of Metropolis-Hastings) are adopt. For $\boldsymbol{\pi}$ and $\boldsymbol{z}$, Gibbs sampling are used as follows:

- sample z_m from its posterior $p(z_m = k|X, Y, \boldsymbol{\theta}, \boldsymbol{\pi}) \propto \pi_k p(\boldsymbol{y}_m|\theta_k, \boldsymbol{x}_m), \quad m = 1, \cdots, M, \quad k = 1, \cdots, K$
- sample π from the conditional distribution:$(\pi_1, \cdots, \pi_K)|\boldsymbol{z} \sim Dir(1+n_1, \cdots, 1+n_K)$

Here, $n_k = \sum_{m=1}^{M} \mathbb{I}_{z_m=k}$. In addition, $\boldsymbol{\theta}$ are sampled by hybrid MCMC from its posterior:

$$p(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{z}) \propto \prod_{k=1}^{K} p(\theta_k|\mathcal{D}_m, \boldsymbol{z}) \tag{7}$$

3.2 The Moves with the Component Number Changed

For the moves with the component number changed, we use the same rules given in [5] but extend the univariables to the multivariables $\boldsymbol{x} = (x_1, \cdots, x_D)$, then the off diagnoal entry becomes:

$$\pi_k \left(v_k \exp\left(-\frac{\sum_{d=1}^{D} w_{k,d}(x_{i,d} - x_{j,d})^2}{2} \right) \right) = \pi_{k_1} \left(v_{k_1} \exp\left(-\frac{\sum_{d=1}^{D} w_{k_1,d}(x_{i,d} - x_{j,d})^2}{2} \right) \right) + \pi_{k_2} \left(v_{k_2} \exp\left(-\frac{\sum_{d=1}^{D} w_{k_2,d}(x_{i,d} - x_{j,d})^2}{2} \right) \right) \quad (8)$$

Since each dimension x_d are independent, we can make the analysis for each variable separately and obtain the new detailed balanced framework:

$$\pi_k = \pi_{k_1} + \pi_{k_2} \quad (9a)$$

$$\pi_k \sigma_k^2 = \pi_{k_1} \sigma_{k_1}^2 + \pi_{k_2} \sigma_{k_2}^2 \quad (9b)$$

$$\pi_k v_k = \pi_{k_1} v_{k_1} + \pi_{k_2} v_{k_2} \quad (9c)$$

$$\pi_k v_k w_{k_d} = \pi_{k_1} v_{k_1} w_{k_1,d} + \pi_{k_2} v_{k_2} w_{k_2,d}, \quad d = 1, \cdots, D \quad (9d)$$

The merge moves are based on the above equations 9, and the split moves are follows: in order to match the dimension before and after split, $(d+3)$-dimension random variable $\boldsymbol{u} = (u_1, u_2, u_3, u_{4,1}, \cdots, u_{4,D})$ need to be generated each are generated from $Beta(2, 2)$. Then combining the detailed balanced equation 9, split formula is given as follows:

$$\pi_{k_1} = u_1 \pi_{k^*}, \quad \pi_{k_2} = (1 - u_1)\pi_{k^*}, u_1 \in (0, 1) \quad (10a)$$

$$\sigma_{k_1}^2 = u_2 \sigma_{k^*}^2 \frac{\pi_{k^*}}{\pi_{k_1}}, \quad \sigma_{k_2}^2 = (1 - u_2)\sigma_{k^*}^2 \frac{\pi_{k^*}}{\pi_{k_2}}, u_2 \in (0, 1) \quad (10b)$$

$$v_{k_1} = u_3 v_{k^*} \frac{\pi_{k^*}}{\pi_{k_1}}, \quad v_{k_2} = (1 - u_3) v_0^{k^*} \frac{\pi_{k^*}}{\pi_{k_2}}, u_3 \in (0, 1) \quad (10c)$$

$$w_{k_1,d} = \frac{1 - u_{4,d}}{u_3} w_k, \quad w_{k_2} = \frac{u_{4,d}}{1 - u_3} w_k, u_{4,d} \in (0, 1), \quad d = 1, \cdots, D \quad (10d)$$

In fact, parameters varies under split and merge moves which forms a birth-and-death Markov chain. The birth and death probability for this Markov chain is set to be: $d_1 = b_{k_{max}} = 0$, $d_{k_{max}} = b_1 = 1, d_k = b_k = 0.5, \forall k = 2, \cdots, k_{max} - 1$. Moreover, it is simulated by Metropolis-Hastings algirithm which is one kind of Markov chain Monte Carlo simulation methods and is called as Reversible Jump Markov chain Monte Carlo simulation in [3]. Therefore, the acceptance probability need to be calculated:

$$A = \prod_{m=1}^{M} \frac{l(Y_m|\theta_{k+1})}{l(Y|\theta_k)} \times \frac{d_{k+1}}{b_k} \times \frac{p(\theta_{k+1})}{p(\theta_k)} \times \frac{1}{Beta(\boldsymbol{u}|\theta_{k+1}, \theta_k)} \times \left| \frac{\partial \theta_{k+1}}{\partial(\theta_k, \boldsymbol{u})} \right| \quad (11)$$

here, k, k_1, k_2 are randomly chosen from $\{1, \cdots, k_{max}\}$.

3.3 Metropolis-Hastings Update

For the Metropolis-Hastings update, we adopt a penalized likelihood [4] as follows:

$$\log L^p(\theta, p|\mathcal{D}, k) = \log L(\theta, p|\mathcal{D}, k) - \mathcal{P} \tag{12}$$

where the penalization $\mathcal{P}$ is BIC penalization:

$$\mathcal{P}_{BIC} = \frac{\xi}{2}\log(N) \tag{13}$$

here ξ denote the number of model parameters $(4 * k)$ and N is the number of curve multiples the number of points on each curve. The acceptance probability of this Metropolis-Hastings is:

$$\min\{1, \frac{L^p(\theta_{new}|\mathcal{D})}{L^p(\theta_{old}|\mathcal{D})}\} \tag{14}$$

4 Convergence Diagnosing

In this subsection, we generate 9 curves as the dataset, which are trajectories of three Gaussian processes $GP(\mu_i, \theta_i),\ i = 1, 2, 3$, respectively. The mean functions are given by

$$\begin{aligned} \mu_1(x_1, x_2) &= f_1(x_1) + f_2(x_2), \\ \mu_2(x_1, x_2) &= f_2(x_1) + f_3(x_2), \\ \mu_3(x_1, x_2) &= f_3(x_1) + f_1(x_2). \end{aligned} \tag{15}$$

where

$$\begin{aligned} f_1(x) &= \exp(x/5) - 1.5 \\ f_2(x) &= \sin(x/4 * \pi) \\ f_3(x) &= -\sin(x/4 * \pi) \end{aligned} \tag{16}$$

and covariance parameters θ_i are given in Tab. 1. The convariants are :$x_1 =$

Table 1. The covariance parameters for the simulated data.

	w_1	w_2	v	σ^2
θ_1	1	0.2	0.2	0.0025
θ_2	0.5	0.5	1	0.001
θ_3	10	1.0	0.2	0.0005

$-4 : 0.08 : 4$ and $x_2 = 0 : 0.08 : 8$. Since the number of points in x_1 and x_2 are the same, there is a one-to-one map between x_1 and x_2 in our experiment. In Fig. 1, we plot the simulated curves. The x-axis is x_1 and y-axis is y. Each 3

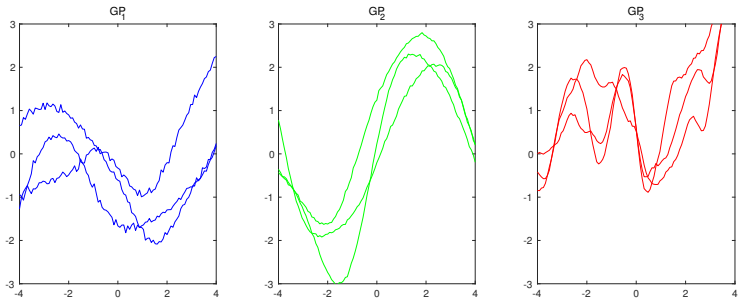

Fig. 1. The simulated dataset. The x-axis is x_1 and y-axis is y. Each 3 curves in one panel are trajectories of the same Gaussian process with common parameters of mean functions and covariance.

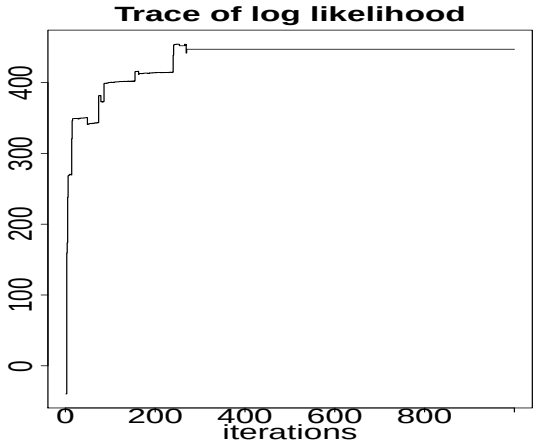

Fig. 2. The trace of log likelihood for 1000 iterations

curves in one panel are trajectories of the same Gaussian process with common parameters of mean functions and covariance.

Fig. 2 illustrates the trace of log likelihood for 1000 iterations of Markov chain Monte carlo simulation. It became stable after 400 iterations. In Fig. 3, we plot the trace of K for 1000 iterations and in Fig. 4 we plot the histogram of K for 1000 iterations. It can be seen that K becomes stable after 400 and stays on 3, which is compatible with our setting.

5 Conclusion

We have improved the RJMCM method for model selection prediction of MGPs on two aspects: first, we extend the split and merge formula for dealing with multi-input regression, which can be used to more complex and interesting data;

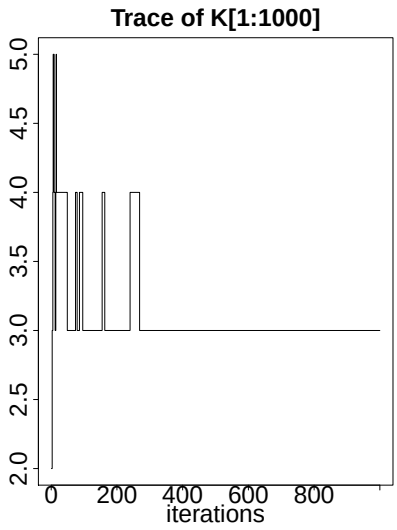

Fig. 3. The trace of K for 1000 iterations

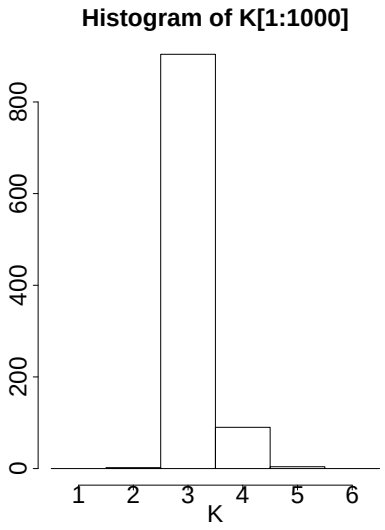

Fig. 4. The histogram of K for 1000 iterations

second, a Metropolis-Hasting update rule is added after the process of the moves, which can remarkably accelerate the convergence. The experimental results on simulated dataset demonstrate that the improved RJMCMC method is feasible and effective.

Acknowledgment

This work is supported by the Natural Science Foundation of China for Grant 61171138.

References

1. Shi, J. Q., Murray-Smith, R., Titterington, D. M. : Hierarchical Gaussian process mixtures for regression. Statistics and computing, 15(1), 31-41 (2005)
2. Shi, J. Q., B. Wang: Curve prediction and clustering with mixtures of Gaussian process functional regression models. Statistics and Computing 18(3): 267-283(2008).
3. Green, P. J.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika, 82(4), 711-732 (1995).
4. Andrieu, C., Freitas, N. D., Doucet, A. : Robust full Bayesian learning for radial basis networks. Neural computation, 13(10), 2359-2407 (2001)..
5. Qiang, Z., Ma, J. : Automatic model selection of the mixtures of Gaussian processes for regression. In Proceedings of the 12th International Symposium on Neural Networks (ISNN 2015), Jeju, South Korea, October 15-18, 2015, LNCS 9377, pp. 335-344.

6. Marin J M, Robert C P. : Bayesian core: a practical approach to computational Bayesian statistics. Springer Science & Business Media. (2007)
7. Marin J M, Robert C P.: Bayesian essentials with R. New York: Springer(2014).
8. Wu D, Chen Z, Ma J. An mcmc based em algorithm for mixtures of gaussian processes. International Symposium on Neural Networks. Springer, Cham, 2015: 327-334.
9. Wu D, Ma J. A Two-Layer Mixture Model of Gaussian Process Functional Regressions and Its MCMC EM Algorithm. IEEE Transactions on Neural Networks and Learning Systems, 2018.
10. Qiang Z, Luo J, Ma J. Curve clustering via the split learning of mixtures of Gaussian processes. 2016 IEEE 13th International Conference on Signal Processing (ICSP): 1089-1094.