STAT 8810 Assignment 3

Due: November 20th, 2017

	Team Member	Team Member
team 1	9	7
team 2	1	6
team 3	8	3
team 4	2	4

Question 1

Show that the Gibbs step for $\theta_j \in (\theta_1, \ldots, \theta_k)$ is a special case of a Metropolis-Hastings step with an acceptance probability of 1.

Question 2

Many of the models we have seen rely on conjugacy between Normal distributions and between the Normal and Gamma distributions. We derived some simple cases in the class notes, here we will look at their more general forms.

Consider data $\mathbf{Y}$ distributed as

$$\mathbf{Y}|\lambda^{-1}\mathbf{R} \sim N\left(\mathbf{F}\boldsymbol{\beta}, \lambda^{-1}\mathbf{R}\right)$$

and priors

$$\boldsymbol{\beta} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

and

$$\lambda \sim Gamma(\alpha, \beta),$$

where $\mathbf{Y}$ is an $n \times 1$ vector, λ is a scalar, $\boldsymbol{\beta}$ is $p \times 1$, $\mathbf{F}$ is $n \times p$, $\mathbf{R}$ is $n \times n$, $\boldsymbol{\mu}$ is $p \times 1$ and $\boldsymbol{\Sigma}$ is $p \times p$.

(a) Derive the full conditional distribution $\boldsymbol{\beta}|\mathbf{Y}, \cdot$.

(b) Derive the full conditional distribution $\lambda|\mathbf{Y}, \cdot$.

Question 3

Write functions to fit the Bayesian Gaussian Process model including linear trend,

$$\mathbf{Y}|\boldsymbol{\beta}, \lambda, \mathbf{R} \sim N\left(\mathbf{F}\boldsymbol{\beta}, \lambda^{-1}\mathbf{R}\right)$$

$$\boldsymbol{\beta} \sim N\left(\boldsymbol{\mu}, \lambda_{\mu}^{-1}\mathbf{I}\right)$$

where $\mathbf{I}$ is the $p \times p$ identity matrix and

$$\mathbf{R}_{ij} = \prod_{d=1}^{D} \rho_d^{(x_{id}-x_{jd})^2}$$

where $\mathbf{x}_i = (x_{i1}, \ldots, x_{iD}) \in \mathbb{R}^D$ and

$$\rho_d \sim \text{Beta}(\alpha, \beta).$$

In addition, write functions to draw samples from the posterior predictive distribution,

$$\pi(y(\mathbf{x})|\mathbf{Y}).$$

Test your functions by simulating "fake" data in $p = 1$ dimensions by drawing from the GP for particular settings of the prior parameters that you choose, and checking that the resulting posterior distributions concentrate on the parameter settings you used to simulate your data.

Question 4

On the website I have included the dataset `co2plume.dat` (you may load it using `R`'s `load()` command). This dataset consists of outputs from a deterministic simulator used to study the dispersion of CO_2 emissions from coal-fired power plants. The response is CO_2 concentration and the inputs are `time` and `stack_inerts`, a reaction parameter.

(a) Using your code from question (3), fit the Bayesian GP model with linear trend. Thoughtfully motivate your choice of priors. Perform some simple diagnostic checks of the posterior distribution to ensure the MCMC algorithm is mixing well. Using your fitted model, predict the held-out observations provided in the `co2holdout.dat` file. Summarize your out-of-sample predictions in terms of MSE and in terms of the empirical coverage of a 95% credible interval from your posterior predictive distribution.

(b) Repeat the analysis in part (a) using the Bayesian treed gaussian process model available as the function `btgp()` in the `R` library `tgp`. Here, you may use the default priors as shown in class.

(c) Repeat the analysis in part (a) using the Bayesian single-tree model available in `R` library `BayesTree`.

(d) Repeat the analysis in part (a) using the BART model available in `R` library `BayesTree`.

Question 5

Consider data $\mathbf{Y}$ generated according to the process

$$Y(\mathbf{x}) = f(\mathbf{x}) + s(\mathbf{x})Z$$

where Z is i.i.d. Normal$(0, 1)$ and $\mathbf{x} \in \mathbb{R}^D$, and where we assume that the mean function is a realization of a single-tree process

$$f(\mathbf{x}) = g(\mathbf{x}; \mathcal{T}, \mathcal{M})$$

with terminal node parameters $\mathcal{M} = (\mu_1, \ldots)$ having conjugate prior distributions $\mu_j \sim N(0, \tau^2)$ and the variance is also modeled as a realization of a single-tree process,

$$s(\mathbf{x})^2 \sim h(\mathbf{x}; \mathcal{T}', \mathcal{M}')$$

with terminal node parameters $\mathcal{M}' = (s_1^2, \ldots)$ having conjugate prior distributions $s_j^2 \sim \chi^{-2}(\nu, \lambda)$. Both trees are regularized by the depth-penalizing prior

$$\alpha(1 + d)^{-\beta}$$

for any node with depth d. The setup then is identical to the single-tree model discussed in class except the variance is now also modeled as a function of $\mathbf{x}$ using a single tree.

The posterior of this model can then be expressed as

$$\pi(\mathcal{T}, \mathcal{M}, \mathcal{T}', \mathcal{M}'|\mathbf{Y}) \propto L(\mathbf{Y}|\mathcal{T}, \mathcal{M}, \mathcal{T}', \mathcal{M}')\pi(\mathcal{M}|\mathcal{T})\pi(\mathcal{T})\pi(\mathcal{M}'|\mathcal{T}')\pi(\mathcal{T}'),$$

where $\pi(\mathcal{M}|\mathcal{T}) = \prod_{j=1}^{m} \pi(\mu_j)$ where $m = |\mathcal{M}|$ is the number of terminal nodes in tree $\mathcal{T}$ and $\pi(\mathcal{M}'|\mathcal{T}') = \prod_{j=1}^{m'} \pi(s_j^2)$ where $m' = |\mathcal{M}'|$ is the number of terminal nodes in tree $\mathcal{T}'$.

(a) Derive the full conditional for the μ_j, $\pi(\mu_j|\cdot)$.

(b) Derive the integrated likelihood, $\int_{\mu_j} L(\mu_j|\cdot)\pi(\mu_j)d\mu_j$.

(c) Derive the full conditional for the s_j^2, $\pi(s_j^2|\cdot)$.

(d) Derive the integrated likelihood, $\int_{s_j^2} L(s_j^2|\cdot)\pi(s_j^2)ds_j^2$.