September 05, 2019

1 Regression Spline

Assume that the range of x is $[a, b]$. Let the point

$$a < \xi_1 < \cdots < \xi_K < b$$

be a partion of the interval $[a, b]$ $\{\xi_1, \ldots, \xi_K\}$ are called knots.

1.1 Radial Basis Function

A RBF φ is a real valued function whose value depends only on the distance from origin. A real function $\varphi : [0, \infty) \to \mathbb{R}$ with a metric on space $\|\cdot\| : V \to [0, \infty)$ a function $\varphi_c = \varphi(\|\mathbf{x} - \mathbf{c}\|)$ is said to be a radial kernel centered at c. A radial function and the associated radial kernels are said to be radial basis function

We use radial basis functions defined by

$$\mathbf{b}(u) = \left\{ u, \left| \frac{u - \tau_1}{c} \right|^3, \cdots, \left| \frac{u - \tau_K}{c} \right|^3 \right\}$$

where c is sample standard deviation

2 Simulation

Let

$$y = \sum_{l=1}^{4} f_l(X_l) + \sum_{k=1}^{4} Z_k \theta_k + e$$

$$\begin{aligned}
f_1(x) &= 3exp(-30(x - 0.3)^2) + exp(-50(x - 0.7)^2) \\
f_2(x) &= sin(2\pi x) \\
f_3(x) &= x \\
f_4(x) &= 0 \\
\theta_1 &= 0.6 \\
\theta_2 &= -1 \\
\theta_3 &= \theta_4 = 0
\end{aligned}$$

Make spline and centerize the data we can get $\tilde{y}$

$$\tilde{y} = y - \bar{y} = b_1(X_1)\beta_1 + b_2(X_2)\beta_2 + b_3(X_3)\beta_3 + b_4(X_4)\beta_4 + \sum_{k=1}^{4} Z_k\theta_k + e$$

2.1 MFVB method

Setting prior as

$$\begin{aligned}
Y|\tau,\beta &\sim N(X\beta, \sigma^2 \cdot I_N) \\
\beta_i|\gamma_i &\sim^{ind} N(0, \sigma^2_\beta) \text{ for } i = 1, \dots p \\
\sigma^2_\beta &\sim Inverse - Gamma(a, b) \\
\sigma^2 &\sim Gamma(c, d)
\end{aligned}$$

By Baye's rule

$$p(\tau, \gamma, \beta|Y) \propto p(Y|\tau, \beta)p(\beta|\gamma)p(\tau)p(\gamma)$$

Then variational distribution is

$$p(\tau, \gamma, \mu|Y) \approx q(\tau, \gamma, \mu) = q_1(\tau)q_2(\gamma)q_3(\mu)$$

we can maximize ELBO by coordinate descent algorithm

$$\begin{aligned}
q_1^*(\sigma^2) &= E_{q_2,q_3}[p(\sigma^2, \gamma, \beta|Y)] \propto E_{q_2,q_3}[p(Y|\sigma^2, \beta)p(\tau)] \\
q_2^*(\sigma^2_\beta) &= E_{q_1,q_3}[p(\tau, \sigma^2_\beta, \beta|Y)] \propto E_{q_1,q_3}[p(\beta|\sigma^2_\beta)p(\sigma^2_\beta)] \\
q_3^*(\beta) &= E_{q_1,q_2}[p(\tau, \gamma, \beta|Y)] \propto E_{q_1,q_2}[p(Y|\tau, \beta)p(\beta|\gamma)]
\end{aligned}$$

Then

$$\begin{aligned}
q_1^* &\sim Gamma\left(c + \frac{N+1}{2}, d + \frac{1}{2}\left\{Y'Y - E_{q3}[\beta'](X'Y)\right\} + tr\left[X(var_{q3}[\beta] + E_{q3}[\beta]E_{q3}[\beta'])X'\right]\right) \\
q_2^* &\sim \prod_{i=1}^{p} Inverse - Gamma(a + \frac{1}{2}, b + \frac{1}{2}\{var_{q3}[\beta]_{i,i} + E_{q3}[\beta_i]^2\}) \\
q_3^* &\sim N\left(E_{q1}[\sigma^2]\Sigma X'Y, \left(diag(E_{q2}[\sigma^2_\beta]) + E_{q1}[\sigma^2]X'X\right)^{-1} = \Sigma\right)
\end{aligned}$$

2.2 MFVB method with variable selection

Variable selection model is

$$Y = X\Gamma\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I)$$

where

$$\begin{aligned}
Y|\beta, \sigma^2, \gamma &\sim N(X\Gamma\beta, \sigma^2 I) \\
\sigma^2 &\sim Inverse - Gamma(A, B) \quad A = 0,\ B = 0 \\
\beta_j &\sim N(0, \sigma_\beta^2) \\
\gamma_j &\sim Bernoulli(\rho) \quad \rho = 0.5
\end{aligned}$$

2.3 MFVB method with variable selection hierarchical model

Variable selection hierarchical model is

$$Y = X\Gamma\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I)$$

where

$$\begin{aligned}
Y|\beta, \sigma^2, \gamma &\sim N(Z\Gamma\beta, \sigma^2 I) \\
\sigma^2 &\sim Inverse - Gamma(A, B) \quad A = 0,\ B = 0 \\
\beta_j &\sim N(0, \sigma_\beta^2) \\
\gamma_j &\sim Bernoulli(\rho) \\
\rho &\sim Beta(C, D) \quad C = 1,\ D = 1.4
\end{aligned}$$

$$p(\beta, \sigma^2, \sigma_\beta^2, \Gamma|Y) \approx q(\beta, \sigma^2, \sigma_\beta^2, \Gamma) = \prod_{j=1}^{p} q_1^*(\beta_j) q_2^*(\sigma^2) q_3^*(\rho) \prod_{j=1}^{p} q_4^*(\gamma_j)$$

Use coordinate ascent algorithm, q density of β is

$$
\begin{aligned}
q_1^*(\beta) &\propto E_{-q_1}\left[p(\beta,\sigma^2,\sigma_\beta^2,\Gamma,\rho,Y)\right] \\
&\propto E_{-q_1}\left[\exp\left(-\frac{1}{2\sigma^2}\left(Y-Z\Gamma\beta\right)'\left(Y-Z\Gamma\beta\right)-\frac{1}{2}\sum_{j=1}^{p}\frac{\beta_j^2}{\sigma_\beta^2}\right)\right] \\
&\propto \exp\left(-\frac{1}{2}\beta'\left\langle D\right\rangle-\frac{1}{2}\left\langle\frac{1}{\sigma^2}\right\rangle\left(\beta'\left\langle\Gamma\right\rangle' Z'Z\left\langle\Gamma\right\rangle\beta-2\beta'\left\langle\Gamma\right\rangle' Z'Y\right)\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\beta'\left(\left\langle D\right\rangle+\left\langle\frac{1}{\sigma^2}\right\rangle\left\langle\Gamma\right\rangle' Z'Z\left\langle\Gamma\right\rangle\right)\beta-2\left\langle\frac{1}{\sigma^2}\right\rangle\beta'\left\langle\Gamma\right\rangle' Z'Y\right]\right) \\
&\sim N(\mu,\Sigma)
\end{aligned}
$$

Where $D = diag(\frac{1}{\sigma_\beta^2})$, <> means expectation under q functions and

$$
\Sigma = \left(\left\langle D\right\rangle+\beta'\left\langle\frac{1}{\sigma^2}\right\rangle\left\langle\Gamma\right\rangle' Z'Z\left\langle\Gamma\right\rangle\right)^{-1}, \quad \mu = \left\langle\frac{1}{\sigma^2}\right\rangle\Sigma\left\langle\Gamma\right\rangle' Z'Y
$$

q density of σ^2 is

$$
\begin{aligned}
q_2^*(\sigma^2) &\propto E_{-q_2}\left[p(\beta,\sigma^2,\sigma_\beta^2,\Gamma,\rho,Y)\right] \\
&\propto E_{-q_2}\left[\left(\sigma^2\right)^{-\left(\frac{n}{2}+a\right)-1}\exp\left(-\frac{1}{\sigma^2}\left(b+\left(Y-Z\Gamma\beta\right)'\left(Y-Z\Gamma\beta\right)\right)\right)\right] \\
&\propto Inverse-Gamma\left(a+\frac{n}{2},b+\frac{1}{2}\left(Y'Y-2\left\langle\beta\right\rangle'\left\langle\Gamma\right\rangle' Z'Y+tr\left(\left(Z'Z\odot\Omega\right)\left(\mu\mu'+\Sigma\right)\right)\right)\right)
\end{aligned}
$$

Where $\odot$ is hadamard product and

- $\gamma = (\gamma_1,\ldots,\gamma_p)$
- $\Omega = \left\langle\gamma\right\rangle\left\langle\gamma\right\rangle' + \left\langle\Gamma\right\rangle\odot\left(I-\left\langle\Gamma\right\rangle\right)$

q density of ρ is

$$
\begin{aligned}
q_3^*(\rho) &\propto E_{-q_3}\left[p(\beta,\sigma^2,\sigma_\beta^2,\Gamma,\rho,Y)\right] \\
&\propto \rho^{C-1}\left(1-\rho\right)^{D-1}\prod_{j=1}^{p}\rho^{\gamma_j}\left(1-\rho\right)^{1-\gamma_j} \\
&\propto \rho^{C+\sum_{j=1}^{p}\gamma_j-1}\left(1-\rho\right)^{D+p-\sum_{j=1}^{p}\gamma_j-1} \\
&\sim Beta(C+\sum_{j=1}^{p}\gamma_j, D+p-\sum_{j=1}^{p}\gamma_j)
\end{aligned}
$$

q density of γ is

$$q_4^*(\gamma) E_{-q_4}\left[\propto \prod_{j=1}^{p} \rho^{\gamma_j}(1-\rho)^{-\gamma_j} \exp(-\frac{1}{2}\frac{1}{\sigma^2}\left(\beta'\Gamma'Z'Z\Gamma\beta - 2\beta'\Gamma'Z'y\right))\right]$$

$$\propto \exp\left[\gamma_j\left(\langle \log\left(\rho/(1-\rho)\right)\rangle - \frac{1}{2}\left\langle\frac{1}{\sigma^2}\right\rangle \langle\beta_j^2\rangle Z_j'Z_j + \left\langle\frac{1}{\sigma^2}\right\rangle Z_j'\left[Y\mu_j - X_{-j}\langle\Gamma_{-j}\rangle\left(\mu_{-j}\mu_j + \Sigma_{-j,j}\right)\right]\right)\right]$$

Where

- X_j means jth coloumn of X
- X_{-j} means without jth column
- $X_{-i,j}$ means jth column without ith component
- μ_j is jth component of vector and mu_j means without jth component

2.3.1 Posterior of rho

$$f_1(x) = 3exp(-30(x-0.3)^2) + exp(-50(x-0.7)^2)$$

$$f_2(x) = sin(2\pi x)$$

$$f_3(x) = x$$

$$f_4(x) = 0$$

$\rho \sim$	$Beta(1, 1.4)$	$Beta(1, 1)$	$Beta(1, 2)$	$Beta(2, 1)$	$Beta(2, 2)$
f1 $q^*(\rho) \sim$	beta(30.0 2.4)	beta(31.0 1.0)	beta(4.0, 29.0)	beta(32.0 1.0)	beta(22.8,10.2)
f2 $q^*(\rho) \sim$	beta(5.0 27.4)	beta(31.0 1.0)	beta(3.0, 30.0)	beta(32.0 1.0)	beta(26.9 6.1)
f3 $q^*(\rho) \sim$	beta(4.0, 28.4)	beta(31.0 1.0)	beta(4.0, 30.0)	beta(32.0 1.0)	beta(26.9 6.1)
f4 $q^*(\rho) \sim$	beta(1.0 31.4)	beta(1.0 31.0)	beta(1.0 32.0)	beta(32.0 1.0)	beta(30.0 3.0)

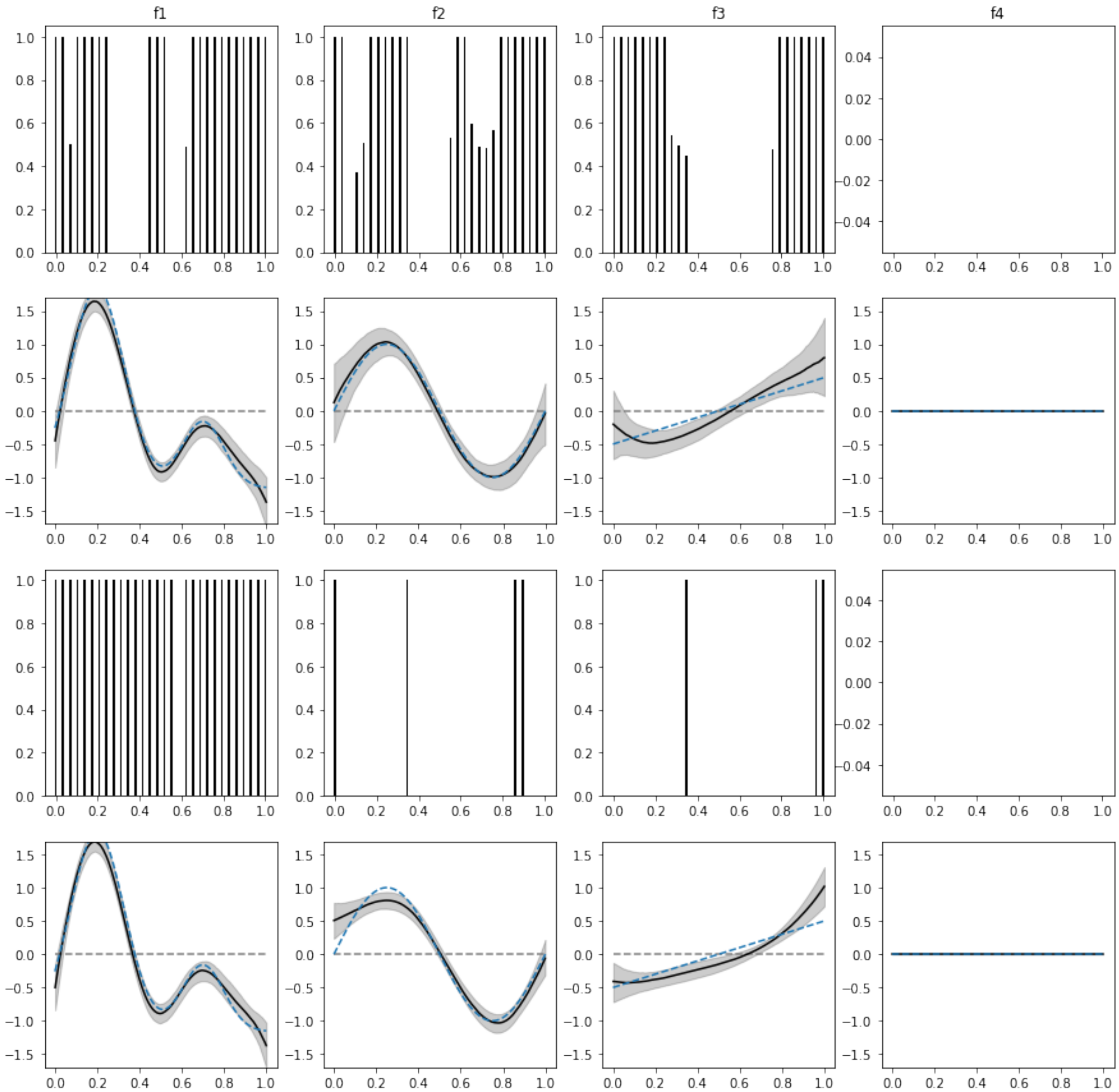

Figure 1: Compare the variable selection model without Beta prior and without Beta prior. First and Second row has plat prior $\rho = 0.5$. Third and Forth row has beta prior $\rho \sim Beta(1, 1.4)$