August 22, 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

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

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

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

$$f_3(x) = x, f_4(x) = 0$$

$$\theta_1 = 0.6$$

$$\theta_2 = -1$$

$$\theta_3 = \theta_4 = 0$$

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

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

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

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

Then

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

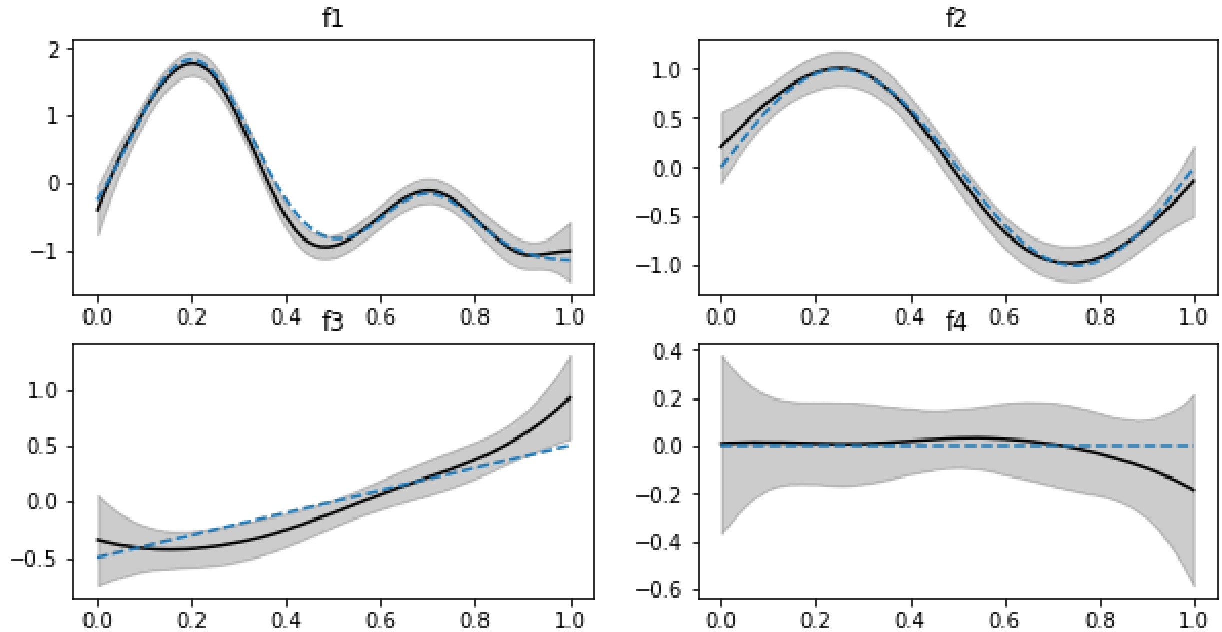

Figure 1: estimation of non linear term with MFVB method

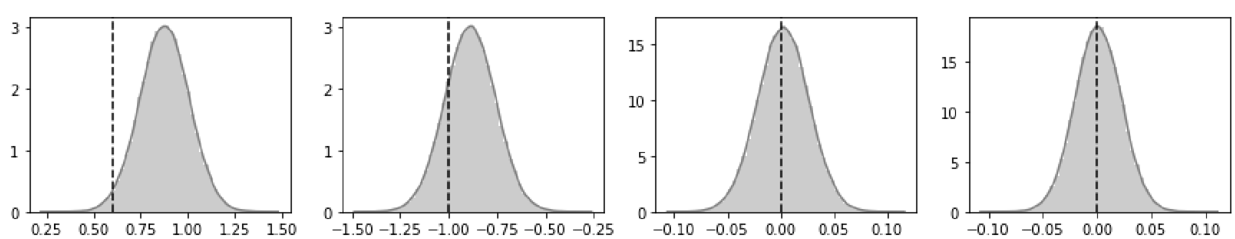

Figure 2: estimation of linear term with MFVB method

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

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

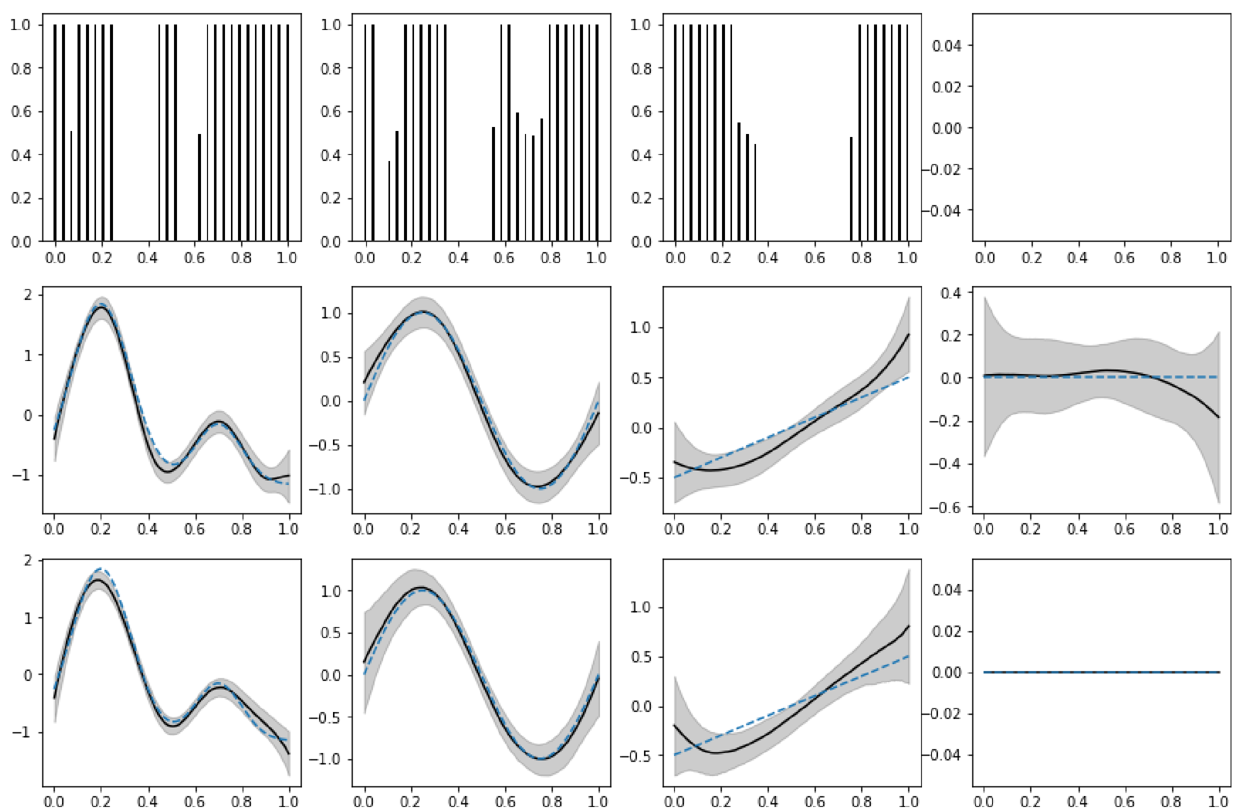

Figure 3: MFVB variable selection non linear term

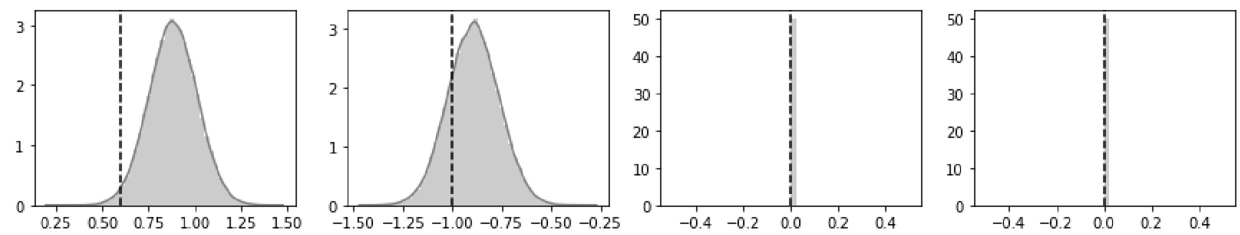

Figure 4: MFVB variable selection non linear term

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

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

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

Use coordinate ascent algorithm, q density of β is

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

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

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

q density of σ^2 is

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

Where *odot* is hadamard product and

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

q density of ρ is

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

q density of γ is

$$\begin{aligned} q_4^*(\gamma) E_{-q_4} & \left[\propto \prod_{j=1}^{p}\rho^{\gamma_j}(1-\rho)^{-\gamma_j}\exp\left(-\frac{1}{2}\frac{1}{\sigma^2}\left(\beta'\Gamma'Z'Z\Gamma\beta - 2\beta'\Gamma'Z'y\right)\right)\right] \\ & \propto \exp\left[\gamma_j\left(\langle\log\left(\rho/(1-\rho)\right)\rangle - \frac{1}{2}\left\langle\frac{1}{\sigma^2}\right\rangle\langle\beta_j^2\rangle Z_j'Z_j + \left\langle\frac{1}{\sigma^2}\right\rangle Z_j'\left[Y\mu_j - X_{-j}\langle\Gamma_{-j}\rangle\left(\mu_{-j}\mu_j + \Sigma_{-j,j}\right)\right]\right)\right] \end{aligned}$$

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

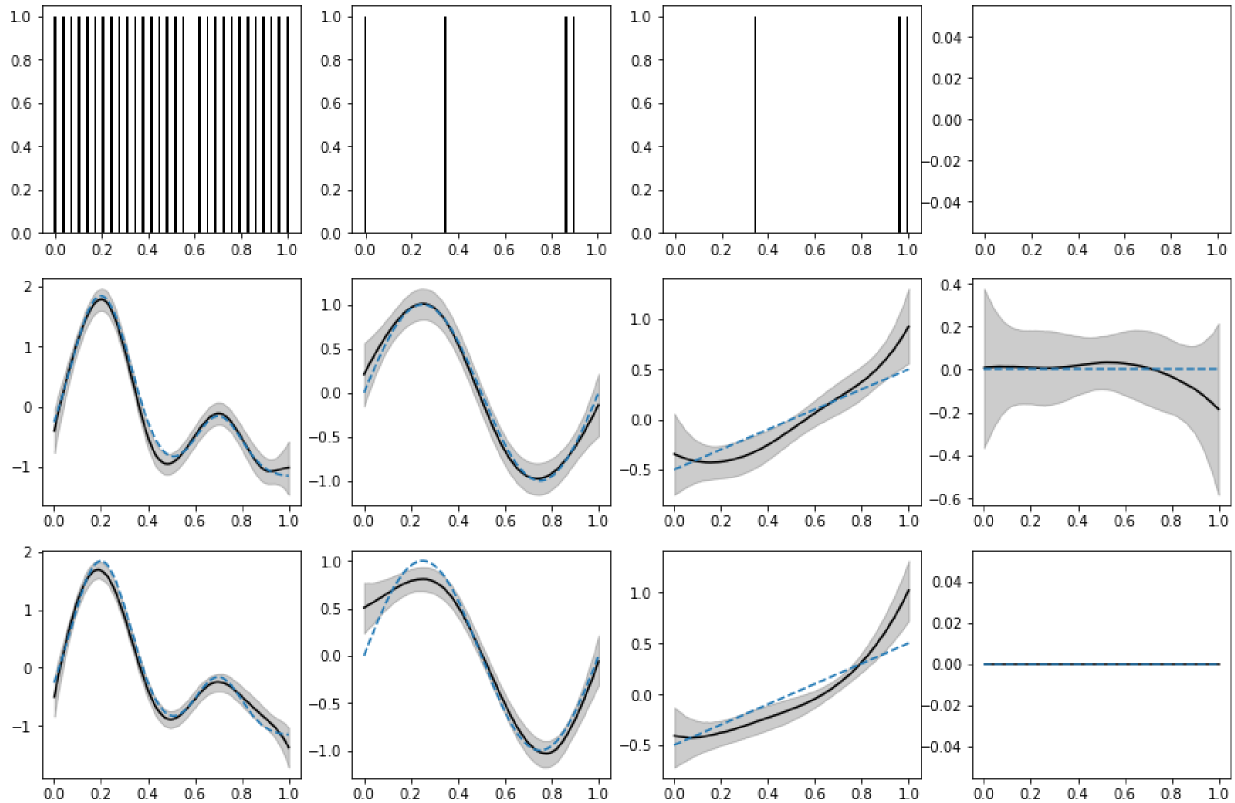

Figure 5: hierarchical model MFVB variable selection non linear term

3 Real Data

$$
\begin{aligned}
log(\text{beta-carotene}) = & f_1(\text{AGE}) + f_1(\text{BMI}) + f_1(\text{CALORIES}) + f_1(\text{FAT}) + \\
& + f_1(\text{FIBER}) + f_1(\text{ALCOHOL}) + f_1(\text{CHOL}) + f_1(\text{BETADIET}) + \\
& + f_1(\text{RETDIET}) + \beta_1\text{SEX} + \beta_2\text{SMOTESTAT} + \beta_3\text{VITUSE} + \epsilon
\end{aligned}
$$

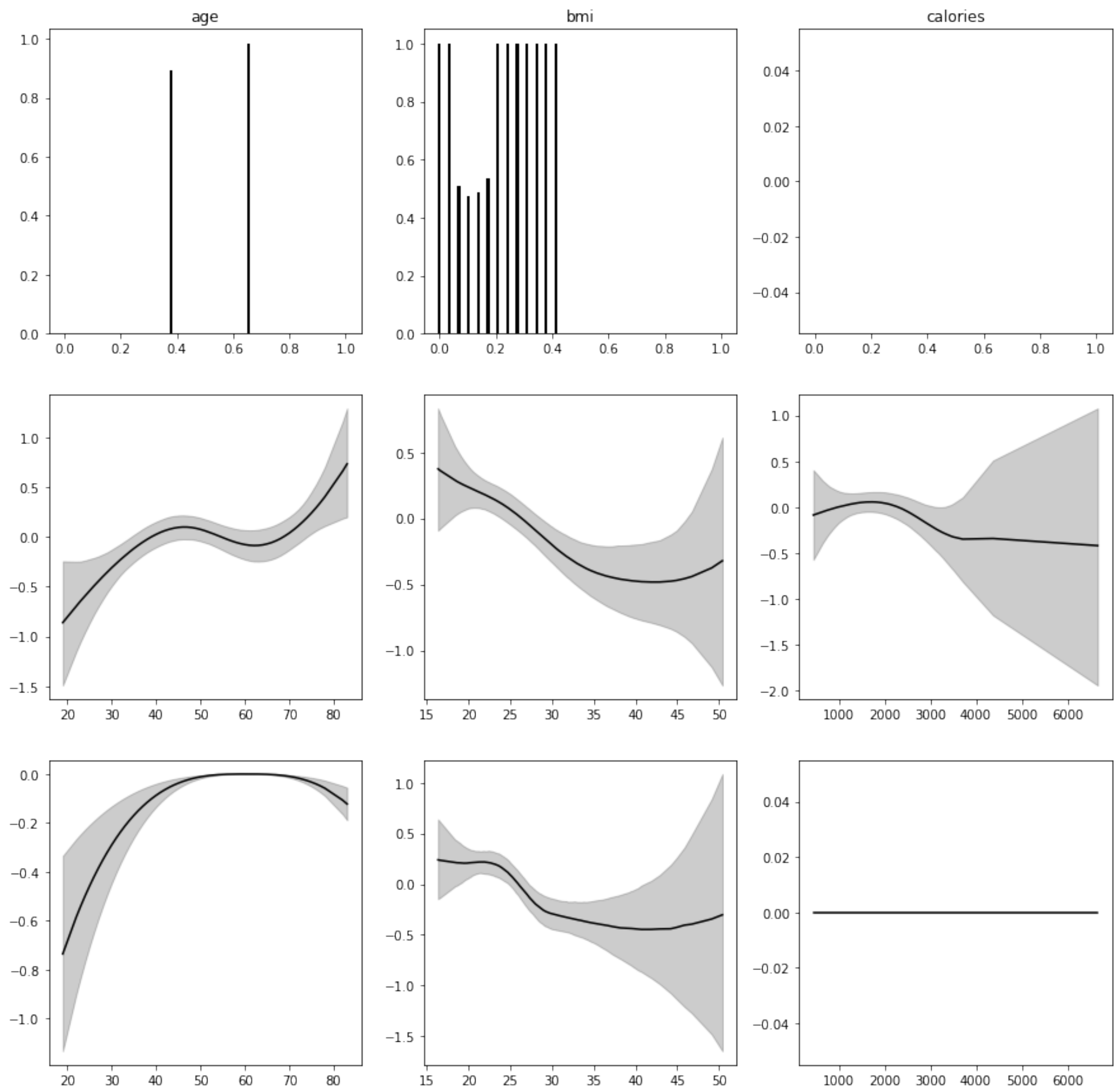

Figure 6: