VARIABLE SELECTION FOR REGRUEEION MODELS

林扬涛

Email: 871636062@qq.com

School of Mathematical Sciences, South China Normal University

May 12, 2017

CONTENTS

模型

(广义)线性模型

$$\mathbf{y} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2 \sim \mathcal{N}_p(\mathbf{X}\vartheta, \sigma^2 I) \tag{1.1}$$

$\boldsymbol{Y} \in \mathcal{R}^{n\times 1}, \mathbf{X} = [\mathbf{X_1}, ..., \mathbf{X_p}] \in \mathcal{R}^{n\times p}$， $\boldsymbol{\vartheta} = (\beta_1\gamma_1, ..., \beta_p\gamma_p)^T$，
$\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2$都是未知参数。$\sigma^2$ 是一个标量。

似然函数

似然函数

$$f(\mathbf{y} \mid \boldsymbol{\beta}, \gamma, \sigma^2) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) \mathbf{exp} \left(-\frac{(\mathbf{y}_i - \boldsymbol{x}_i \vartheta)^2}{2\sigma^2} \right)$$

$$= \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \mathbf{exp} \left(-\frac{(\mathbf{y} - \mathbf{X}\vartheta)^T (\mathbf{y} - \mathbf{X}\vartheta)}{2\sigma^2} \right)$$

参数$\beta, \gamma_j, \sigma^2$的先验

β(多元高斯先验)

$$\beta \sim \mathcal{N}_p(\beta_0, \boldsymbol{D_0}) \tag{1.2}$$

γ_j(二项分布)

$$\gamma_j \sim B(1, p_j) \tag{1.3}$$

σ^2(逆伽马分布)

$$\sigma^2 \sim \mathcal{IG}(\frac{\alpha}{2}, \frac{\eta}{2}) \tag{1.4}$$

参数β的后验

多元高斯分布的一个重要性质

如果两组变量是联合高斯分布,那么一组变量为条件,另一组变量同样是高斯分布.类似地,任何一个变量的边缘分布也是高斯分布

参数β的后验

- $\mathbf{y} \mid \boldsymbol{\beta} \sim \mathcal{N}_p(\mu, \Sigma)$, 其中$\mu = \mathbf{X}^* \boldsymbol{\beta}, \Sigma = \sigma^2 I, \mathbf{X}^* = [\gamma_1 \mathbf{x_1}, ..., \gamma_1 \mathbf{x_1}]$
- $\boldsymbol{\beta} \sim \mathcal{N}_p(\boldsymbol{\beta_0}, \boldsymbol{D_0})$
- 贝叶斯定理$\Rightarrow$

$$\begin{aligned}\ln p((\boldsymbol{\beta}, \mathbf{y})^T) &= \ln(p(\mathbf{y} \mid \boldsymbol{\beta}) \cdot p(\boldsymbol{\beta})) \\ &= \ln p(\mathbf{y} \mid \boldsymbol{\beta}) + \ln p(\boldsymbol{\beta}) \\ &= -\frac{1}{2}(\mathbf{y} - \mu)^T \Sigma^{-1} (\mathbf{y} - \mu) - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \boldsymbol{D_0}^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + C\end{aligned}$$

其中,常数C表示与$\mathbf{y}, \boldsymbol{\beta}$无关的项.令$z = (\boldsymbol{\beta}, \mathbf{y})^T$,则上式为$z$的分量的一个二次函数,因此$p(z)$是一个高斯分布.为找到这个高斯分布的精度,我们考虑上式的前两项中的二次项.

参数β的后验

$$
\begin{aligned}
&-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}+\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mu+\frac{1}{2}\mu^T\Sigma^{-1}\mathbf{y}-\frac{1}{2}\mu^T\Sigma^{-1}\mu-\frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{D}_{\mathbf{0}}^{-1}\boldsymbol{\beta}\\
&=-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}+\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{X}^*\boldsymbol{\beta}+\frac{1}{2}(\mathbf{X}^*\boldsymbol{\beta})^T\Sigma^{-1}\mathbf{y}-\frac{1}{2}(\mathbf{X}^*\boldsymbol{\beta})^T\Sigma^{-1}\mathbf{X}^*\boldsymbol{\beta}-\frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{D}_{\mathbf{0}}^{-1}\boldsymbol{\beta}\\
&=-\frac{1}{2}\boldsymbol{\beta}^T(\boldsymbol{D}_{\mathbf{0}}^{-1}+\mathbf{X}^*\Sigma^{-1}\mathbf{X}^*)\boldsymbol{\beta}-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}+\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{X}^*\boldsymbol{\beta}+\frac{1}{2}\boldsymbol{\beta}^T\mathbf{X}^{*T}\Sigma^{-1}\mathbf{y}\\
&=\begin{pmatrix}\boldsymbol{\beta}\\ \mathbf{y}\end{pmatrix}^T\begin{pmatrix}\boldsymbol{D}_{\mathbf{0}}^{-1}+\mathbf{X}^*\Sigma^{-1}\mathbf{X}^* & -\mathbf{X}^{*T}\Sigma^{-1}\\ -\Sigma^{-1}\mathbf{X}^* & \Sigma^{-1}\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}\\ \mathbf{y}\end{pmatrix}\\
&=z^TRz
\end{aligned}
$$

其中R为z上的高斯分布的精度矩阵(协方差的逆矩阵)

参数β的后验

- 根据公式对R求逆得

$$cov[z] = \begin{pmatrix} \boldsymbol{D_0} & \boldsymbol{D_0}\mathbf{X}^{*T} \\ \mathbf{X}^{*}\boldsymbol{D_0} & \sigma^2 I + \mathbf{X}^{*}\boldsymbol{D_0}\mathbf{X}^{*} \end{pmatrix}$$

- 根据条件高斯分布得β的后验服从高斯分布

$$\boldsymbol{\beta} \sim \mathcal{N}_p(\widetilde{\boldsymbol{\beta}}, \boldsymbol{D})$$

- 根据边缘高斯分布中条件概率的期望和方差得

$$\widetilde{\boldsymbol{\beta}} = (\boldsymbol{D_0}^{-1} + \sigma^{-2}\mathbf{X}^{*T}\mathbf{X}^{*})^{-1}(\boldsymbol{D_0}^{-1}\boldsymbol{\beta_0} + \sigma^{-2}\mathbf{X}^{*T}\mathbf{y})$$
$$\boldsymbol{D} = (\boldsymbol{D_0}^{-1} + \sigma^{-2}\mathbf{X}^{*T}\mathbf{X}^{*})^{-1}$$

参数γ_j的后验

二项分布

$$\gamma_j \mid \gamma_{-j}, \boldsymbol{\beta}, \sigma^2, \mathbf{y} \sim B(1, \widetilde{p}_j) \tag{1.5}$$

- $\widetilde{p}_j = c_j/(c_j + d_j)$
- $c_j = p_j \mathbf{exp}\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta}_j^*)^T(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta}_j^*)\}$
- $d_j = (1 - p_j)\mathbf{exp}\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta}_j^{**})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta}_j^{**})\}$

其中$\boldsymbol{\vartheta}_j^*$由$\boldsymbol{\vartheta}$中第$j$个数被$\beta_j$替换得到
其中$\boldsymbol{\vartheta}_j^{**}$由$\boldsymbol{\vartheta}$中第$j$个数被0 替换得到

参数σ^2的后验

二项分布

$$
\begin{aligned}
f(\sigma^2 \mid \mathbf{y}, \boldsymbol{\beta}, \gamma) &= f(\mathbf{y} \mid \sigma^2) f(\sigma^2) \\
&\propto \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \mathbf{exp}\left(-\frac{(\mathbf{y}-\mathbf{X}\vartheta)^T(\mathbf{y}-\mathbf{X}\vartheta)}{2\sigma^2}\right) \mathcal{IG}\left(\frac{\alpha}{2}, \frac{\eta}{2}\right) \\
&\propto (\sigma^2)^{-\frac{\alpha+n}{2}-1} \mathbf{exp}\left(\frac{\eta + (\mathbf{y}-\mathbf{X}\boldsymbol{\vartheta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\vartheta})}{2\sigma^2}\right)
\end{aligned}
$$

故

$$
\sigma^2 \mid \mathbf{y}, \boldsymbol{\beta}, \gamma \sim \mathcal{IG}\left(\frac{\alpha+n}{2}, \frac{\eta + (\mathbf{y}-\mathbf{X}\boldsymbol{\vartheta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\vartheta})}{2}\right)
$$

算法

SSVS2

1. 初始化:$\boldsymbol{\beta_0}, \boldsymbol{D_0}, \gamma, \alpha, \eta, p_j$
2. 迭代: for t=1,2, ...
   - 存储$\boldsymbol{\beta}$
   - 存储$\boldsymbol{\gamma}$
   - 存储$\boldsymbol{\sigma^2}$
3. 统计$\boldsymbol{\gamma}$

实现

示例

$$\mathbf{y} = \mathbf{x_4} + 1.2\mathbf{x_5} + \epsilon \tag{1.6}$$

其中$\epsilon \sim \mathcal{N}_{60}(0, \sigma^2 I)$,$\sigma = 2.5$