Bayesian Variable Selection
—— Stochastic Search Variable Selection

Dom Tsang

Email: 395091490@qq.com

School of Mathematical Sciences, South China Normal University

May 11, 2017

Contents

Contents

变量选择

- 一范数约束

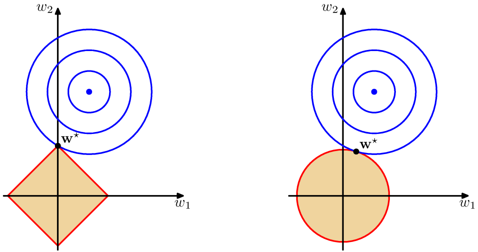

Figure 1: 正则化限制区域

机器学习中的范数规则化之(一)L0、L1 与 L2 范数
http://blog.csdn.net/zouxy09/article/details/24971995

- 逐步回归

分层贝叶斯模型示例

- 模型: $\mathbf{y}_{(P,1)} = \mathbf{H}_{(P,M)}\mathbf{x}_{(M,1)} + \mathbf{n}_{(P,1)}$ 其中, $\mathbf{n} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_p)$
- 似然函数: $f(\mathbf{y}|\mathbf{x},\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{P/2} \exp\left(-\frac{||\mathbf{y}-\mathbf{H}\mathbf{x}||^2}{2\sigma^2}\right)$
- 参数先验: $f(\mathbf{x}|w,a) = [(1-w)^{n_0}\prod_{i\in\mathcal{I}_0}\delta(x_i)][w^{n_1}\prod_{i\in\mathcal{I}_1}g_a(x_i)]$
  $\sigma^2|\nu,\gamma \sim \mathcal{IG}(\frac{\nu}{2},\frac{\gamma}{2})$
- 超参数的先验: $a|\alpha \sim \mathcal{IG}(\alpha_0,\alpha_1)$
  $w \sim \mathcal{U}([0,1])$
  $f(\gamma) \propto \frac{1}{\gamma}$
- 参数后验: $w|\mathbf{x} \sim \mathcal{BE}(1+n_1, 1+n_0)$
  $a|\mathbf{x},\alpha_0,\alpha_1 \sim \mathcal{IG}(||\mathbf{x}||_0+\alpha_0, ||\mathbf{x}||_1+\alpha_1)$
  $f(x_i|w,a,\sigma^2,\mathbf{x}_{-i},\mathbf{y}) \propto (1-w_i)\delta(x_i) + w_i\phi_+(x_i|\mu_i,\eta_i^2)$
  $\sigma^2|\mathbf{x},\mathbf{y} \sim \mathcal{IG}(\frac{P}{2},\frac{||\mathbf{y}-\mathbf{H}\mathbf{x}||^2}{2})$

算法

- 初始化 $\longleftarrow$ 从先验中
  1. 随机生成 $\mathbf{x}^{(0)}$
  2. 随机生成 $\sigma^{(0)}$
- 循环: for t = 1, 2, ..., do:
  1. 随机生成 $w^{(t)}$
  2. 随机生成 $a^{(t)}$
  3. For i = 1, ..., M $\longleftarrow$ 从后验中
     随机生成 $x_i^{(t)}$
  4. 随机生成 $\sigma^{2(t)}$
  5. t $\leftarrow$ t+1

Contents

模型

(广义) 线性模型

$$\boldsymbol{Y}|\boldsymbol{\beta},\sigma^2 \sim \mathcal{MVN}(\boldsymbol{X}\boldsymbol{\beta},\sigma^2 I) \tag{2.1}$$

其中, $\boldsymbol{Y} \in \mathcal{R}^n, \boldsymbol{X} \in \mathcal{R}^{n\times p}, \boldsymbol{\beta} = (\beta_1,...,\beta_p)^\top, \sigma^2$ 是一个标量. $\boldsymbol{\beta}$ 和 σ^2 都是未知参数.
*$\mathcal{MVN}(\boldsymbol{X}\boldsymbol{\beta},\sigma^2 I)$ 即是通常的正态分布 $\mathcal{N}(\boldsymbol{X}\boldsymbol{\beta},\sigma^2 I)$
* 式 (2.1) 的意思是 $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{n}$, 其中 $\sim \mathcal{N}(0,\sigma^2 I)$

β 的先验

- **使用两个不同方差的高斯分布的混合分布对 β 的每个成分进行建模**. 引入隐变量 $\gamma_i = 0$ 或 1 , 可将高斯分布混合成,

β_i (混合高斯)

$$\beta_i|\gamma_i \sim (1-\gamma_i)\mathcal{N}(0,\tau_i^2) + \gamma_i\mathcal{N}(0,c_i^2\tau_i^2) \tag{2.2}$$

$$P(\gamma_i = 1) = 1 - P(\gamma_i = 0) = \omega_i \tag{2.3}$$

* 式 (8) 表达的模型用大家相对熟悉的表达形式即是
$P(\beta_i) = P(\gamma_i = 0)P(\beta_i|\gamma_i = 0) + P(\gamma_i = 1)P(\beta_i|\gamma_i = 1)$
$= (1-\omega_i)\mathcal{N}(0,\tau_i^2) = \omega_i\mathcal{N}(0,c_i^2\tau_i^2)$

β 的先验

β (多元高斯先验)

$$\boldsymbol{\beta}|\boldsymbol{\gamma} \sim \mathcal{MVN}(0, \boldsymbol{D}_\gamma \boldsymbol{R} \boldsymbol{D}_\gamma) \tag{2.4}$$

其中 $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_p)^\top$, $\boldsymbol{R}$ 是先验协方差矩阵,
$\boldsymbol{D}_\gamma = diag[\{d_i\}_{i=1}^p]$, 其中 $d_i = (1-\gamma_i)\tau_i + \gamma_i c_i \tau_i$.

γ 的先验

γ (二项分布)

$$\boldsymbol{\gamma} \sim \prod_{i=1}^{p} \omega_i^{\gamma_i} (1 - \omega_i)^{1-\gamma_i} \tag{2.5}$$

σ^2 的先验

σ^2(逆 gamma 分布)

$$\sigma^2|\gamma \sim \mathcal{IG}(\frac{\nu_\gamma}{2}, \frac{\nu_\gamma\lambda_\gamma}{2}) \tag{2.6}$$

β 的后验

β 的后验

$$\begin{aligned}\boldsymbol{\beta}^{(j)} &\sim f(\boldsymbol{\beta}^{(j)}|\boldsymbol{Y},\sigma^{2(j-1)},\boldsymbol{\gamma}^{j-1}) \\ &= \mathcal{MVN}(A_{\gamma^{j-1}}(\sigma^{j-1})^{-2}\boldsymbol{X}^\top\boldsymbol{Y},A_{\gamma^{j-1}}) \end{aligned} \tag{2.7}$$

其中，$A_{\gamma^{j-1}} = ((\sigma^{j-1})^{-2}\boldsymbol{X}^\top\boldsymbol{X} + \boldsymbol{D}_{\gamma^{j-1}}^{-1}\boldsymbol{R}^{-1}\boldsymbol{D}_{\gamma^{j-1}}^{-1})$.

σ^2 的后验

σ^2 的后验

$$\sigma^{2(j)} \sim f(\sigma^{2(j)} | \boldsymbol{Y}, \boldsymbol{\beta}^j, \boldsymbol{\gamma}^{j-1}) = \mathcal{IG}(\frac{n + \nu_{\gamma^{j-1}}}{2}, \frac{|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}^j|^2 + \nu_{\gamma^{j-1}}\lambda_{\gamma^{j-1}}}{2}) \tag{2.8}$$

γ 的后验

$$\gamma_i^j \sim f(\gamma_i^j|\boldsymbol{Y},\boldsymbol{\beta}^j,\sigma^{2(j)},\gamma_{(\neq i)}^j) = f(\gamma_i^j|\boldsymbol{\beta}^j,\sigma^{2(j)},\gamma_{(\neq i)}^j) \tag{2.9}$$

每一个伯努利分布的概率为:

$$P(\gamma_i^j = 1|\boldsymbol{\beta}^{(j)},\sigma^{2(j)},\boldsymbol{\gamma}_{(\neq i)}^{(j)}) = \frac{a}{a+b} \tag{2.10}$$

其中,

$$a = f(\boldsymbol{\beta}^{(j)}|\boldsymbol{\gamma}_{(\neq i)}^{(j)}),\gamma_{(i)}^{(j)} = 1)f(\sigma^{2(j)}|\boldsymbol{\gamma}_{(\neq i)}^{(j)}),\gamma_{(i)}^{(j)} = 1)f(\boldsymbol{\gamma}_{(\neq i)}^{(j)}),\gamma_{(i)}^{(j)} = 1) \tag{2.11}$$

$$b = f(\boldsymbol{\beta}^{(j)}|\boldsymbol{\gamma}_{(\neq i)}^{(j)}),\gamma_{(i)}^{(j)} = 0)f(\sigma^{2(j)}|\boldsymbol{\gamma}_{(\neq i)}^{(j)}),\gamma_{(i)}^{(j)} = 0)f(\boldsymbol{\gamma}_{(\neq i)}^{(j)}),\gamma_{(i)}^{(j)} = 0) \tag{2.12}$$

γ 的后验

当 σ 的先验参数为常数时 ($\nu_\gamma \equiv \nu, \lambda_\gamma \equiv \lambda$), 式2.11, 2.12可进一步简化为

$$
\begin{aligned}
a &= f(\boldsymbol{\beta}^{(j)} | \boldsymbol{\gamma}_{(\neq i)}^{(j)}), \gamma_{(i)}^{(j)} = 1) f(\boldsymbol{\gamma}_{(\neq i)}^{(j)}), \gamma_{(i)}^{(j)} = 1) \\
&= f(\boldsymbol{\beta}^{(j)} | \boldsymbol{\gamma}_{(\neq i)}^{(j)}), \gamma_{(i)}^{(j)} = 1) \omega_i \qquad (2.13) \\
b &= f(\boldsymbol{\beta}^{(j)} | \boldsymbol{\gamma}_{(\neq i)}^{(j)}), \gamma_{(i)}^{(j)} = 0) f(\boldsymbol{\gamma}_{(\neq i)}^{(j)}), \gamma_{(i)}^{(j)} = 0) \\
&= f(\boldsymbol{\beta}^{(j)} | \boldsymbol{\gamma}_{(\neq i)}^{(j)}), \gamma_{(i)}^{(j)} = 0)(1 - \omega_i) \qquad (2.14)
\end{aligned}
$$

算法

SSVS

1. 初始化: $\boldsymbol{\beta}^0, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{c}, \boldsymbol{\tau}, \boldsymbol{R}, \boldsymbol{\omega}, \nu, \lambda$
2. 迭代: for t = 1, 2, ...:
   - 从式2.7中抽取 $\boldsymbol{\beta}$;
   - 从式2.8中抽取 σ^2;
   - 从式2.10中抽取 $\boldsymbol{\gamma}$.

Contents

问题描述

- 问题: 从模型集合中选择一个最优模型
- 在贝叶斯框架中, 模型选择问题被转化为参数估计的形式而不是寻找单一的最优模型, 贝叶斯方法将估计所有模型的后验概率. 在很多情况下, 这个问题从变量的层面被提出:例如, 估计模型中变量的边缘后验概率.

模型空间

在模型选择中, 模型空间的均匀分布被定义成:

$$f(m) = \frac{1}{|\mathcal{M}|}, forallm \in \mathcal{M} \tag{4.1}$$

模型的量化表示

- γ 的向量空间与模型空间建立起一个一一对应的映射:

$$m(\gamma) = \sum_{j=k}^{p} \gamma_j 2^{j-k} \tag{4.2}$$

模型的后验概率

$$\hat{f}(m|\boldsymbol{Y}) = \frac{1}{T-B}\sum_{t=B+1} I(m^{(t)} = m) \tag{4.3}$$

其中, T 和 B 分别是算法总共的迭代次数和预模拟运行 (burnin) 的迭代次数, $m^{(t)}$ 是模型在迭代次数 t 下的指示值.

Reference

O'Hara R B, Sillanpaa M J. A review of Bayesian variable selection methods: what, how and which[J]. Bayesian analysis, 2009, 4(1): 85-117.

George E I, McCulloch R E. Variable selection via Gibbs sampling[J]. Journal of the American Statistical Association, 1993, 88(423): 881-889.

Thanks!