

贝叶斯统计 2020 年考查题目

使用 logistic 回归模型分析数据, 即 $y_i \sim \text{Bernoulli}(\pi_i)$, $\text{logit}(\pi_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \doteq \mathbf{x}_i^T \boldsymbol{\beta}$, 其中 $\mathbf{x}_i = (1, x_{i1}, x_{i2})^T$ 和 $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ 。现有数据 (x_{i1}, x_{i2}, y_i) , $i = 1, \dots, n$, 并记 $\mathbf{y} = \{y_1, \dots, y_n\}$ 。假设 $\boldsymbol{\beta}$ 的先验分布为无信息先验, 即 $p(\boldsymbol{\beta}) \propto \text{常数}$ 。

1. 写出 $\boldsymbol{\beta}$ 的后验分布 $p(\boldsymbol{\beta}|\mathbf{y})$ 。
2. 叙述如何利用 Metropolis-Hastings 算法从 $p(\boldsymbol{\beta}|\mathbf{y})$ 抽取样本 $\boldsymbol{\beta}_1^{post}, \boldsymbol{\beta}_2^{post}, \dots$ 注: 一定要利用 MH 算法, 要把 $\boldsymbol{\beta}$ 看成一个整体更新, 而不能用 Gibbs sampler 把 $\boldsymbol{\beta}$ 分割抽样。
3. 用你熟悉的语言编写以上利用 MH 算法更新 $\boldsymbol{\beta}$ 的代码用于估计 $\boldsymbol{\beta}$ 。
4. 用你熟悉的语言编写代码, 计算该模型的 DIC 或 WAIC (二选一)。
5. 用 3. 和 4. 的代码分析数据集 Bayes_test.txt (样本量为 1000, 数据前两列分别为 x_1 和 x_2 , 第三列为 y , 取值为 0 或 1), 得到 $\boldsymbol{\beta}$ 的估计和模型的 DIC 或 WAIC。提醒: 请记得模型有截距项。
6. 记 $M_0: \text{logit}(\pi_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$, $M_1: \text{logit}(\pi_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2}$ 。用 3. 和 4. 的代码用模型 M_1 重新分析数据集 Bayes_test.txt。根据 DIC 或 WAIC, 哪个模型较优?
7. 用现有的软件包 (例如 R 语言中的 glm 函数) 分别用 M_0 和 M_1 再一次分析数据集, 并比较用现有软件包和用你代码得到结果的异同。

把上述内容整合成一份数据报告 (源代码另附)。注: 3. 和 4. 的代码可以自己写, 也利用现成的软件包或函数直接得到分析结果。