Exercício: regressão múltipla sob a abordagem bayesiana

Disciplina: Modelagem Estatística
Instrutor: Luiz Max Carvalho
Monitor: Isaque Pim

Abril/2023

Notação: Como convenção adotamos $\mathbb{R} = (-\infty, \infty)$, $\mathbb{R}_+ = (0, \infty)$ e $\mathbb{N} = \{1, 2, \ldots\}$.

Motivação: Como vimos até aqui, o modelo linear (gaussiano) é extremamente útil para modelar a relação entre possíveis variáveis explanatórias (i.e. covariáveis) e uma variável dependente/resposta contínua. Até agora vimos como fazer inferência para esse modelo sob a ótica clássica/frequentista. Vamos então nos debruçar sobre o tratamento bayesiano do problema.

Tome $\boldsymbol{X}$ uma matriz real $n \times P$ e $\boldsymbol{Y} = \{Y_1, \ldots, Y_n\}^T \in \mathbb{R}^n$ um vetor contendo os valores da variável dependente.

Nosso modelo é

$$E[Y_i] =: \mu_i(\boldsymbol{\beta}) = \tilde{\boldsymbol{X}}_i^T \boldsymbol{\beta},$$

onde $\boldsymbol{\beta} \in \mathbb{R}^{P+1}$ é o vetor de coeficientes e parâmetro de interesse e $\tilde{\boldsymbol{X}}$ é uma matriz obtida adicionando uma coluna de uns, $\boldsymbol{X_0} = \{1, \ldots, 1\}^T$, a $\tilde{\boldsymbol{X}}$. Como antes, vamos assumir que os erros em torno do preditor linear são normalmente distribuídos com variância comum:

$$Y_i \overset{\text{i.i.d.}}{\sim} \text{Normal}(\mu_i(\boldsymbol{\beta}), \sigma^2),$$

com $\sigma^2 \in \mathbb{R}_+$ desconhecida. Do ponto de vista bayesiano, precisamos especificar uma distribuição de probabilidade conjunta para todas as quantidades desconhecidas θ do modelo. Neste caso, $\theta = (\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^{P+1} \times \mathbb{R}_+$.

Questões

1. Mostre que a verossimilhança $f_{\tilde{\boldsymbol{X}}}(\boldsymbol{y} \mid \theta)$ pode ser escrita na forma

$$f_{\tilde{\boldsymbol{X}}}(\boldsymbol{y} \mid \theta) = g_{\tilde{\boldsymbol{X}}}(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2) h_{\tilde{\boldsymbol{X}}}(\boldsymbol{y}|\sigma^2).$$

2. Utilize o resultado anterior para deduzir que a priori conjugada para este caso é da forma

$$\pi_{B,S}\left(\boldsymbol{\beta}, \sigma^2\right) = \pi_{B|S}\left(\boldsymbol{\beta} \mid \sigma^2\right) \pi_S(\sigma^2).$$

   Em particular, mostre que

$$\pi_{B,S}\left(\boldsymbol{\beta}, \sigma^2\right) \propto \left(\frac{1}{\sigma^2}\right)^{a+(P+1)/2+1} \times \exp\left(-\frac{1}{\sigma^2}\left\{b + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)^T \boldsymbol{V}_\beta^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)\right\}\right),$$

   onde $\boldsymbol{\mu}_\beta) \in \mathbb{R}^{P+1}$, $\boldsymbol{V}_\beta$ é uma matriz positiva definida e $a, b \in \mathbb{R}_+$.

   Dica: Que escolhas para $\pi_{B|S}$ e π_S eu preciso fazer?

3. A priori anterior chama-se **normal inversa gama** (NIG) e tem quatro parâmetros: $\boldsymbol{m}$, $\boldsymbol{V}$, a e b. Mostre que a posteriori de θ também é NIG e exiba seus hiperparâmetros.

4. **Distribuições marginais**

   Dica: Antes de começar os cálculos para essa seção, vale considerar a seguinte representação do nosso modelo:

$$\boldsymbol{y} = \tilde{\boldsymbol{X}}^T \boldsymbol{\beta} + \boldsymbol{\epsilon}_1, \text{ com } \boldsymbol{\epsilon}_1 \sim \text{MVN}_n\left(\boldsymbol{0}_n, \boldsymbol{\Sigma_1}\right),$$
$$\boldsymbol{\beta} = \boldsymbol{\mu}_\beta + \boldsymbol{\epsilon}_2, \text{ com } \boldsymbol{\epsilon}_2 \sim \text{MVN}_{P+1}\left(\boldsymbol{0}_{P+1}, \boldsymbol{\Sigma_2}\right),$$

   onde $\boldsymbol{\epsilon}_1$ e $\boldsymbol{\epsilon}_2$ são erros independentes.

   (a) Determine $\boldsymbol{\Sigma}_1$ e $\boldsymbol{\Sigma}_2$;

   (b) Compute a verossimilhança marginal com respeito a σ^2:

$$\tilde{f}_{\tilde{\boldsymbol{X}}}(\boldsymbol{y} \mid \sigma^2) := \int_{\mathbb{R}^{P+1}} f_{\tilde{\boldsymbol{X}}}(\boldsymbol{y} \mid \boldsymbol{b}, \sigma^2) \pi_{B|S}(\boldsymbol{b} \mid \sigma^2)\, d\boldsymbol{b}.$$

(c) Usando o item anterior, compute a verossimilhança marginal ou *preditiva a priori*:

$$m_{\tilde{\boldsymbol{X}}}(\boldsymbol{y}) := \int_0^\infty \tilde{f}_{\tilde{\boldsymbol{X}}}(\boldsymbol{y} \mid s)\pi_S(s)\, ds.$$

(d) Mostre $\bar{f}_{\tilde{\boldsymbol{X}}}(\boldsymbol{\beta} \mid \boldsymbol{y})$ e comente sobre como calcular, por exemplo, $\Pr(\beta_1 > a \mid \boldsymbol{y})$, para $a \in \mathbb{R}$.

5. Suponha que eu coletei uma nova matriz de desenho $m \times P$, $\boldsymbol{X}'$ e quero prever o valor de $\boldsymbol{y}'$ a partir do que eu aprendi usando $\boldsymbol{X}$ e $\boldsymbol{y}$. Compute

$$\bar{p}_{\tilde{\boldsymbol{X}}, \boldsymbol{X}'}(\boldsymbol{y}' \mid \boldsymbol{y}) := \int_{\mathbb{R}^{P+1} \times \mathbb{R}_+} p_{\tilde{\boldsymbol{X}}}(\boldsymbol{b}, s \mid \boldsymbol{y}) f_{\boldsymbol{X}'}(\boldsymbol{y}' \mid \boldsymbol{b}, s)\, d\boldsymbol{b} ds,$$

e esboce o seu gráfico para uma observação (linha de $\boldsymbol{X}'$) de um conjunto de dados da sua escolha.

Dica: use um conjunto de dados que você conheça bem. Bons exemplos são os bancos de 'peso ao nascer' e 'kid score', que já analisamos em sala.

Resultados úteis

Aqui estão enunciados alguns resultados úteis para o desenvolvimento das questões acima. Estes são dados sem demonstração, que você está convidada a fazer.

- **Completando o "quadrado" em múltiplas dimensões**: tome $\boldsymbol{A}$ matriz simétrica positiva definida $d \times d$ e $\boldsymbol{\alpha}, \boldsymbol{u} \in \mathbb{R}^d$. Vale que:

$$\boldsymbol{u}^T \boldsymbol{A} \boldsymbol{u} = (\boldsymbol{u} - \boldsymbol{A}^{-1}\boldsymbol{\alpha})^T \boldsymbol{A} (\boldsymbol{u} - \boldsymbol{A}^{-1}\boldsymbol{\alpha}) - \boldsymbol{\alpha}^T \boldsymbol{A} \boldsymbol{\alpha}. \tag{1}$$

  Dica: Expanda o produto e procure por cancelamentos de termos da forma $\boldsymbol{a}^T \boldsymbol{M}^{-1} \boldsymbol{a}$.

- **Sherman-Woodbury-Morrisson**: tome $\boldsymbol{A}$ matriz quadrada $d \times d$ inversível, $\boldsymbol{B}$ matriz $k \times d$, $\boldsymbol{C}$ matriz $d \times k$ e $\boldsymbol{D}$ matriz quadrada $k \times k$ inversível. Então

$$(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{D}\boldsymbol{C})^{-1} = \boldsymbol{A}^{-1} - \left(\boldsymbol{D}^{-1} + \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \boldsymbol{C}\boldsymbol{A}^{-1}.$$

- **Determinantes**: tome $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ e $\boldsymbol{D}$ como antes. Então,

$$\det(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{D}\boldsymbol{C}) = \det(\boldsymbol{A}) \det(\boldsymbol{D}) \det\left(\boldsymbol{D}^{-1} + \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right).$$