Lista 7

MI406-Regressão

Caio Gomes Alves

1 Questão 1

1.1 Pergunta

Seja $\hat{\beta}$ o estimador de mínimos quadrados de um modelo de regressão linear $\hat{\beta} = (X^\top X)^{-1} X^\top y$. Mostre que:

$$(y - X\beta)^\top (y - X\beta) = (y - X\hat{\beta})^\top (y - X\hat{\beta}) + (\hat{\beta} - \beta)^\top X^\top X(\hat{\beta} - \beta) \tag{1}$$

1.2 Resposta

Subtraindo $X\hat{\beta}$ da forma quadrática inicial teremos:

$$\begin{aligned}(y - X\beta)^\top (y - X\beta) &= (y - X\beta + X\hat{\beta} - X\hat{\beta})^\top (y - X\beta + X\hat{\beta} - X\hat{\beta}) \\ &= (y - X\hat{\beta})^\top (y - X\hat{\beta}) + (y - X\hat{\beta})^\top X(\hat{\beta} - \beta) + X^\top (\hat{\beta} - \beta)(y - X\hat{\beta}) + (\hat{\beta} - \beta)^\top X^\top X(\hat{\beta} - \beta)\end{aligned}$$

Como $(y - X\hat{\beta})$ são os resíduos de Mínimos Quadrados Ordinários, eles serão ortogonais às colunas de X, de modo que $(y - X\hat{\beta})^\top X = 0$, de modo que $(y - X\hat{\beta})^\top X(\hat{\beta} - \beta) = 0^\top (\hat{\beta} - \beta) = 0$. Além disso, podemos transpor o terceiro termo de modo que $(\hat{\beta} - \beta)^\top X^\top (y - X\hat{\beta}) = (\hat{\beta} - \beta)0 = 0$. Assim, ambos os termos centrais são zerados, de modo que teremos:

$$(y - X\beta)^\top (y - X\beta) = (y - X\hat{\beta})^\top (y - X\hat{\beta}) + (\hat{\beta} - \beta)^\top X^\top X(\hat{\beta} - \beta) \tag{2}$$

Que é a igualdade que tínhamos no início.

2 Questão 2

2.1 Pergunta

Considere o modelo de regressão linear:

$$y|\beta, \sigma^2 \sim N(X\beta, \sigma^2 I_n)$$

com $y \in \mathbb{R}^n, X \in \mathbb{R}^{n \times p}$, de posto completo, e parâmetros $\beta \in \mathbb{R}^p, \sigma^2 > 0$.

- **(a)** Usando a identidade do item 1, mostre que a verossimilhança pode ser escrita como:

$$\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(y - X\hat{\beta})^\top (y - X\hat{\beta})\right) \exp\left(-\frac{1}{2\sigma^2}(\hat{\beta} - \beta)^\top X^\top X(\hat{\beta} - \beta)\right)$$

- **(b)** Interprete o primeiro fator como função de σ^2, mas independentes de β, e o segundo como (parte de) uma densidade normal em β com variância proporcional a σ^2. Comente a utilidade dessa decomposição para a construção da distribuição a posteriori.

2.2 Resposta

a)

Utilizando o resultado obtido em (2) na função de distribuição da Normal Multivariada, teremos que a verossimilhança será:

$$\begin{aligned} p(y|\beta, \sigma^2) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\left[(y - X\hat{\beta})^\top (y - X\hat{\beta}) + (\hat{\beta} - \beta)^\top X^\top X(\hat{\beta} - \beta)\right]\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(y - X\hat{\beta})^\top (y - X\hat{\beta})\right) \exp\left(-\frac{1}{2\sigma^2}(\hat{\beta} - \beta)^\top X^\top X(\hat{\beta} - \beta)\right) \end{aligned}$$

Que utiliza a propriedade da função exponencial em que $\exp a + b = \exp(a)\exp(b)$, chegando assim na verossimilhança na forma explicitada.

b)

Como $\hat{\beta} = (X^\top X)^{-1}X^\top Y$, o primeiro termo pode ser reescrito como:

$$\exp\left(-\frac{1}{2\sigma^2}(y - X(X^\top X)^{-1}X^\top Y)^\top (y - X(X^\top X)^{-1}X^\top Y)\right)$$

Que não depende de β, representando assim a verossimilhança dos dados quando avaliados no ponto de $\beta = \hat{\beta}$. O segundo fator representa o núcleo (lide *kernel*) da parte exponencial de uma distribuição Normal Multivariada, com média $\hat{\beta}$ e variância $\sigma^2(X^\top X)^{-1}$, sendo assim proporcional à σ^2.

Essa decomposição é útil para a identificação da distribuição à posteriori de β, pois o primeiro termo não depende dele e pode ser considerado como uma constante de normalização para $p(\beta|y, \sigma^2)$, enquanto que o segundo termo indica que a distribuição à posteriori $p(\beta|\sigma^2)$ será Normal Multivariada, quando não houver priori informativa para β.

3 Questão 3

3.1 Pergunta

Considere o modelo de regressão linear:

$$y|\beta, \sigma^2 \sim N(X\beta, \sigma^2 I_n)$$

com $X \in \mathbb{R}^{n\times p}$ conhecido. Suponha as prioris independentes:

$$\begin{aligned}\beta &\sim \mathcal{N}(\mu_0, \Sigma_0)\\ \sigma^2 &\sim IG(a,b)\end{aligned}$$

- **(a)** Encontre a distribuição a posteriori $\pi(\beta, \sigma|y)$ a menos de uma constante de normalização.
- **(b)** Encontre a distribuição condicional a posteriori $\pi(\beta|y, \sigma^2)$.
- **(c)** Encontre a distribuição condicional a posteriori $\pi(\sigma^2|y, \beta)$.

3.2 Resposta

a)

Com os termos para as prioris e a verossimilhança dos dados, teremos que a posteriori será:

$$\pi(\beta, \sigma^2|y) \propto p(y|\beta, \sigma^2)\pi(\beta)\pi(\sigma^2)$$

Em que teremos a verossimilhança, dada por:

$$p(y|\beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^\top (y - X\beta)\right\}$$

E as distribuições a priori para β e σ^2:

$$\begin{aligned}\pi(\beta) &= (2\pi)^{-p/2}|\Sigma_0|^{-1/2} \exp\left\{-\frac{1}{2}(\beta - \mu_0)^\top \Sigma_0^{-1}(\beta - \mu_0)\right\}\\ \pi(\sigma^2) &= \frac{b^a}{\Gamma(a)}(\sigma^2)^{-(a+1)} \exp\left\{-\frac{b}{\sigma^2}\right\}\end{aligned}$$

Após multiplicarmos e agruparmos os termos que dependem de σ^2, teremos:

$$\pi(\beta, \sigma^2|y) \propto (\sigma^2)^{-(n/2+a+1)} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^\top (y - X\beta) - \frac{b}{\sigma^2}\right\} \cdot \exp\left\{-\frac{1}{2}(\beta - \mu_0)^\top \Sigma_0^{-1}(\beta - \mu_0)\right\}$$

Verifique que a primeira exponencial está com o mesmo formato do apresentado na questão 1, de modo que podemos utilizar a identidade apresentada. Assim, a parte exponencial com σ^2 pode ser reescrita como:

$$\exp\left\{-\frac{1}{2\sigma^2}\left[(y - X\hat{\beta})^\top (y - X\hat{\beta}) + (\beta - \hat{\beta})^\top X^\top X(\beta - \hat{\beta}) + 2b\right]\right\}$$

E a parte exponencial com β pode ser reescrita como:

$$\exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma^2}(\beta - \hat{\beta})^\top X^\top X(\beta - \hat{\beta}) + (\beta - \mu_0)^\top \Sigma_0^{-1}(\beta - \mu_0)\right]\right\}$$

Podemos definir os seguintes termos para fins de simplicidade:

$$\Sigma_n^{-1} = \frac{1}{\sigma^2} X^\top X + \Sigma_0^{-1}$$
$$\mu_n = \Sigma_n \left(\frac{1}{\sigma^2} X^\top X \hat{\beta} + \Sigma_0^{-1} \mu_0 \right)$$

Com isso, a distribuição a posteriori conjunta será dada por:

$$\pi(\beta, \sigma^2|y) \propto (\sigma^2)^{-(n/2+a+1)} \exp\left\{ -\frac{1}{\sigma^2} \left[\frac{1}{2}(y - X\hat{\beta})^\top (y - X\hat{\beta}) + b \right] \right\} \cdot \exp\left\{ -\frac{1}{2}(\beta - \mu_n)^\top \Sigma_n^{-1} (\beta - \mu_n) \right\}$$

b)

Para encontrarmos a distribuição condicional a posteriori $\pi(\beta|y, \sigma^2)$, devemos analisar a conjunta obtida anteriormente considerando σ^2 como uma constante. Para tanto, apenas a última exponencial será considerada, que como visto na questão 2 é o núcleo de uma Normal Multivariada, de forma que a posteriori de β será:

$$\pi(\beta|y, \sigma^2) \sim \mathcal{N}(\mu_n, \Sigma_n)$$

Onde μ_n e Σ_n são os valores já especificados anteriormente.

c)

De forma similar ao realizado anteriormente, temos que a posteriori condicional de σ^2 será dada avaliando a posteriori conjunta considerando β constante. Com isso, podemos chegar em:

$$\pi(\sigma^2|y, \beta) \propto (\sigma^2)^{-(n/2+a+1)} \exp\left\{ -\frac{1}{\sigma^2} \left[\frac{1}{2}(y - X\beta)^\top (y - X\beta) + b \right] \right\}$$

Que é o núcleo de uma Gamma Inversa, com os seguintes parâmetros:

$$\pi(\sigma^2|y, \beta) \sim IG\left(\frac{n}{2} + a, \frac{1}{2}(y - X\beta)^\top (y - X\beta) + b \right)$$

4 Questão 4

4.1 Pergunta

Considere o modelo de regressão linear:

$$y|\beta \sim N(X\beta, \sigma^2 I_n)$$

com $\sigma^2 > 0$ conhecido. Suponha que a priori sobre β seja:

$$\beta \sim N(0, \tau^2 (X^\top X)^{-1})$$

onde $\tau^2 > 0$ é uma constante conhecida.

- **(a)** Mostre que a distribuição a posteriori de $\beta|y$ é normal com:

$$\beta|y \sim \mathcal{N}\left(\frac{\tau^2}{\tau^2+\sigma^2}\hat{\beta}, \frac{\sigma^2\tau^2}{\tau^2+\sigma^2}(X^\top X)^{-1}\right)$$

Onde $\hat{\beta} = (X^\top X)^{-1}X^\top y$ é o estimador de mínimos quadrados.

- **(b)** Interprete a média a posteriori como um "encolhimento" de $\hat{\beta}$ em direção à origem. Qual o papel de τ^2 nesse encolhimento?
- **(c)** O que acontece com a distribuição a posteriori de β nos casos extremos:
  - $\tau^2 \to \infty$?
  - $\tau^2 \to 0$?

4.2 Resposta

a)

Sabemos que a distribuição a posteriori de $\beta|y$ será dada por:

$$\pi(\beta|y) \propto p(y|\beta)\pi(\beta)$$

A Verossimilhança para os dados será:

$$p(y|\beta) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^\top(y - X\beta)\right\}$$

Considerando a priori para beta sendo $\pi(\beta) \sim \mathcal{N}(0, \tau^2(X^\top X)^{-1})$, podemos focar apenas no núcleo da distribuição, chegando em:

$$\pi(\beta) \propto \exp\left\{-\frac{1}{2\tau^2}\beta^\top X^\top X\beta\right\}$$

Para a verossimilhança, podemos expandir o termo quadrático, de modo a chegar em:

$$(y - X\beta)^\top(y - X\beta) = y^\top y - 2\beta^\top X^\top y + \beta^\top X^\top X\beta$$

Multiplicando os dois termos, chegamos no exponencial combinado:

$$\exp\left(-\frac{1}{2\sigma^2}(y^\top y - 2\beta^\top X^\top y + \beta^\top X^\top X\beta) - \frac{1}{2\tau^2}(\beta^\top X^\top X\beta)\right)$$

Rearranjando os termos e eliminando os termos que não dependem de β, chegamos em:

$$\exp\left(\beta^\top\left(\frac{X^\top y}{\sigma^2}\right) - \frac{1}{2}\beta^\top X^\top X\left(\frac{\tau^2+\sigma^2}{\sigma^2\tau^2}\right)\beta\right)$$

Após completar quadrado, para que a expressão anterior pareça com o núcleo de uma distribuição Normal Multivariada, chegamos que a média a posteriori será:

$$\mu_p = \frac{\tau^2}{\tau^2+\sigma^2}\hat{\beta}$$

E a variância a posteriori será:

$$\Sigma_p = \left(\frac{\sigma^2 \tau^2}{\tau^2 + \sigma^2} \right) (X^\top X)^{-1}$$

Portanto, a distribuição a posteriori de $\beta | y$ será:

$$\beta | y \sim \mathcal{N} \left(\frac{\tau^2}{\tau^2 + \sigma^2} \hat{\beta}, \frac{\sigma^2 \tau^2}{\tau^2 + \sigma^2} (X^\top X)^{-1} \right)$$

b)

Como os valores de σ^2 e τ^2 são estritamente positivos, temos que $0 < \frac{\tau^2}{\tau^2+\sigma^2} < 1$, de modo que a média a posteriori será uma proporção do valor de $\hat{\beta}$. Assim, independentemente do valor estimado para $\hat{\beta}$, o valor a posteriori será reduzido e aproximado do valor 0 (que é o valor a priori para β).

c)

Caso o valor de $\tau^2 \to \infty$, a variância à priori para β será infinita, de modo que a distribuição será imprópria, e completamente não informativa. Com isso, o valor a posteriori para β não será encolhido para 0, de modo que os dados "dominam" a posteriori e a média será exatamente igual ao valor do Estimador de Mínimos Quadrados.

Caso o valor de $\tau^2 \to 0$, a variância à priori para β será igual a 0, dando a entender que há certeza que o valor de β será 0. Neste caso, a priori domina a verossimilhança, e independentemente do valor estimado para $\hat{\beta}$, o valor de β a posteriori será reduzido para 0.