1 Aula 5/4

Estatística Bayesiana: Teoria da decisão

Modelo Bayesiano O modelo Bayesiano é composto por três elementos:

1. Distribuição preditiva: $(X_1, \cdots, X_n)$ é um vetor aleatório dependente, com

$$p(x_1, \cdots, x_n) = p(x_{\rho(1)}, \cdots, x_{\rho(n)}),$$

onde $\rho(1), \cdots, \rho(n)$ representa qualquer permutação dos índices $1, \cdots, n$. A propriedade probabilística acima é chamada permutabilidade e sugere que a ordem em que observamos $X_1, \cdots, X_n$ não é relevante para a especificação do modelo.

2. Modelo a priori: Sob a hipótese de permutabilidade do vetor aleatório $X_1, \cdots, X_n$, o teorema da representação de de Finetti garante a existência de uma variável "latente" que, condicional a ela, o vetor de observações $X_1, \cdots, X_n$ é independente, isto é, sendo θ a variável "latente", temos

$$p(x_1, \cdots, x_n|\theta) = \prod_{i=1}^{n} p(x_i|\theta).$$

Pondera-se que, embora exista essa ideia de variável "latente", θ tem interpretação limite em função da lei forte dos grandes números.

3. Modelo a posteriori: Com base no teorema de Bayes, a obtenção de dados sugere atualizar a informação inicial. Mais especificamente, no instante inicial apenas a informação a priori está disponível. A medida que novas informações chegam, o procedimento de aprendizado é atualizado com base no teorema de Bayes.

Para motivar a discussão sobre o procedimento inferencial sob o paradigma Bayesiano, vejamos o seguinte exemplo:

Exemplo (Bayes, 1764) Uma bola é lançada sobre uma linha de comprimento 1 com probabilidade uniforme. Suponha que a bola lançada pare no ponto θ. Em seguida, uma segunda bola é lançada sob as mesmas condições da primeira, e define-se: $X = 1$, se a bola para antes do ponto θ, e $X = 0$, caso contrário. Pergunta: Qual inferência podemos fazer para θ?

Por simplicidade, suponha que o valor observado foi $X = 1$. Assim, o estimador de máxima verossimilhança de θ é $\hat{\theta} = 1$. Note que $P(X > \hat{\theta}) = 0$, e portanto testes de hipóteses do tipo

$$\mathrm{H}_0 : \theta = \theta_0 \text{ versus } \mathrm{H}_1 : \theta = \theta_0,$$

para $\theta_0 \in (0, 1)$, sempre rejeitará a hipótese H_0, seja qual for o nível de significância do teste (o p-valor para o teste acima, $P(X > \hat{\theta})$, é nulo).

Procedimentos inferenciais Bayesianos (usualmente) decorrem de métodos envolvendo teoria da decisão, cujo principal objetivo é minimizar alguma estrutura de perda.

Teoria da decisão: Elementos básicos Uma teoria da decisão é composta pela tripla:

1. Espaço de decisões (associado ao espaço paramétrico);
2. Conjunto de possíveis ações, A, disponíveis para a tomada de decisão;
3. Uma função perda L, sendo $L(\theta, a)$ a perda associada a tomada de decisão a, quando o valor verdadeiro do parâmetro é θ.

O problema prático, onde θ é desconhecido, sugere ser impossível calcular $L(\theta, a)$. Como alternativa a essa problemática, podemos tomar decisões em função do valor esperado da função perda, ou seja, através da minimização de

$$E[L(\theta, a)|X],$$

onde $X = (X_1, \cdots, X_n)$ representa um conjunto de dados observados. A decisão 'a' que minimiza $E[L(\theta, a)|X]$ é denominada estimador de Bayes para a função perda $L(\theta, a)$.

Algumas funções perda usuais, são:

(i) Perda Quadrática: $L(\theta, a) = (\theta - a)^2$

(ii) Perda Absoluta: $L(\theta, a) = |\theta - a|$

(iii) Perda 0-1: $L(\theta, a) = 0$, se $|\theta - a| < \epsilon$ e $L(\theta, a) = 1$, se $|\theta - a| \geq \epsilon$.

A seguir vamos demonstrar quais os estimadores de Bayes para os casos (i) e (ii). No caso (iii), é direto que o estimador de Bayes é a moda a posteriori de θ.

Demonstração estimador de Bayes sob perda quadrática

$$\begin{aligned}
E(L(\theta, a)|X) &= E[(\theta - a)^2|X] = E[(\theta - E(\theta|X) + E(\theta|X) - a)^2|X] \\
&= E[(\theta - E(\theta|X))^2|X] + 2E[\theta - E(\theta|X)|X](E(\theta|X) - a) + E[(E(\theta|X) - a)^2|X] \\
&= Var(\theta|X) + (E(\theta|X) - a)^2,
\end{aligned}$$

e, portanto, a decisão ótima é $a^* = E(\theta|X)$.

2 Aula 7/4

Demonstração estimador de Bayes sob perda absoluta

$$
\begin{aligned}
E(L(\theta,a)|X) &= E[|\theta-a| \mid X] = \int_{-\infty}^{\infty} |\theta-a| p(\theta|X)d\theta \\
&= \int_{-\infty}^{a} (a-\theta)p(\theta|X)d\theta + \int_{a}^{\infty} (\theta-a)p(\theta|X)d\theta \\
&= a\left[\int_{-\infty}^{a} p(\theta|X)d\theta - \int_{a}^{\infty} p(\theta|X)d\theta\right] - \int_{-\infty}^{a} \theta p(\theta|X)d\theta + \int_{a}^{\infty} \theta p(\theta|X)d\theta \\
&= a\left[\int_{-\infty}^{a} p(\theta|X)d\theta - \left(1 - \int_{-\infty}^{a} p(\theta|X)d\theta\right)\right] \\
&\quad - \int_{-\infty}^{a} \theta p(\theta|X)d\theta + \int_{a}^{\infty} \theta p(\theta|X)d\theta + \int_{-\infty}^{a} \theta p(\theta|X)d\theta - \int_{-\infty}^{a} \theta p(\theta|X)d\theta \\
&= a\left[2\int_{-\infty}^{a} p(\theta|X)d\theta - 1\right] + E(\theta|X) - 2\int_{-\infty}^{a} \theta p(\theta|X)d\theta.
\end{aligned}
$$

Agora podemos derivar a função acima. Assim, temos

$$
\begin{aligned}
\frac{\partial}{\partial a} E(L(\theta,a)|X) &= \left[a\left[2\int_{-\infty}^{a} p(\theta|X)d\theta - 1\right] + E(\theta|X) - 2\int_{-\infty}^{a} \theta p(\theta|X)d\theta\right]' \\
&= \left[2\int_{-\infty}^{a} p(\theta|X)d\theta - 1\right] + a2p(a|X) - 2ap(a|X) \\
&= 2\int_{-\infty}^{a} p(\theta|X)d\theta - 1.
\end{aligned}
$$

Daí, temos que a decisão ótima é $a^* = \text{Mediana}(\theta|X)$.

Teste de hipóteses Considere um modelo estatístico $f(x|\theta)$ com $\theta \in \Theta$. Seja Θ_0 e Θ_1 uma partição de Θ, ou seja, $\Theta_0 \cup \Theta_1 = \Theta$ e $\Theta_0 \cap \Theta_1 = \emptyset$. Assuma que estamos interessados em testar

$$\mathrm{H}_0 : \theta \in \Theta_0 \text{ versus } \mathrm{H}_1 : \theta \in \Theta_1.$$

Defina a seguinte estrutura de perda

$L(\theta,a)$	Decisão	
	$\theta \in \Theta_0$	$\theta \in \Theta_1$
$\theta \in \Theta_0$	0	α
$\theta \in \Theta_1$	β	0

sendo $\alpha, \beta > 0$. Sob essa formulação, qual o estimador de Bayes associado?

O objetivo é minimizar $E[L(\theta,a)|X]$. Assim, temos

$$\begin{aligned} E[L(\theta,a)|X] &= \int_{\theta\in\Theta} L(\theta,a)p(\theta|X)d\theta = \int_{\theta\in\Theta_0} L(\theta,a)p(\theta|X)d\theta + \int_{\theta\in\Theta_1} L(\theta,a)p(\theta|X)d\theta \\ &= \alpha I_{\{\theta\in\Theta_1\}}(a)\int_{\theta\in\Theta_0} p(\theta|X)d\theta + \beta I_{\{\theta\in\Theta_0\}}(a)\int_{\theta\in\Theta_1} p(\theta|X)d\theta. \end{aligned}$$

Portanto

$$E[L(\theta,a)|X] = \begin{cases} \alpha P(\theta\in\Theta_0|X), \text{ se decidimos por H}_1, \\ \beta P(\theta\in\Theta_1|X), \text{ se decidimos por H}_0. \end{cases}$$

Logo, H_0 é preferível a H_1 quando $\beta P(\theta\in\Theta_1|X) < \alpha P(\theta\in\Theta_0|X)$. Segue que

$$\begin{aligned} \beta P(\theta\in\Theta_1|X) &< \alpha P(\theta\in\Theta_0|X) \ \ (\Leftrightarrow) \\ \beta(1-P(\theta\in\Theta_0|X)) &< \alpha P(\theta\in\Theta_0|X) \ \ (\Leftrightarrow) \\ \beta &< \alpha P(\theta\in\Theta_0|X) + \beta P(\theta\in\Theta_0|X) \ \ (\Leftrightarrow) \\ P(\theta\in\Theta_0|X) &> \frac{\beta}{\alpha+\beta}. \end{aligned}$$

Conclui-se assim que H_0 é preferível a H_1 quando $P(\theta\in\Theta_0|X) > \frac{\beta}{\alpha+\beta}$.