Plano Aula 18

Markus Stein

09 May 2019

Statistical Learning = Inferência Estatística?

(Conectar com o exercício 1 da aula passada)

- **Supervisionado**: interesse em explicar o valor observaçdo y de uma **variável resposta** através da mensuração de uma **variável preditora** x (ou um vetor $\boldsymbol{x}$). Exemplo Modelos Lineares.
  - Exemplo: $Y = f(X) + \epsilon$, onde
    - Y é variável chamada resposta;
    - X é uma variável preditora;
    - ϵ é termo (erro) aleatório tal que $E(\epsilon) = 0$ e $Var(\epsilon) = \sigma^2$.
  - Como estimar f? se $\epsilon \sim Normal(\mu, \sigma^2)$ então $f(X) = \mu$, Basta estimar μ.
- **Não supervisionado**: **não possui variável resposta** Y definida. Exemplo problemas de classificação.
- Predição $\times$ estimação em **aprendizado estatístico**?
  - Estimar f e saber qual **preditor** influencia a **resposta**? Como é a relação entre cada **preditor** e **resposta**?
  - Predizer Y através de um $\hat{Y}$ que minimize $E[(Y - \hat{Y})^2]$, por exemplo.
  - *trade-off* entre **interpretação do modelo** e **precisão da predição**.

Família exponencial

- Definição 1: (**Família Exponencial Unidimensional**) (Bolfarine e Sandoval, definição 2.4.1, pg. 25) Dizemos que a distribuição da variável aleatória X, com f.m.p ou f.d.p. dada por $f(x;\theta)$, pertence à família exponencial unidimensional, se pudermos escrever f como

$$f(x;\theta) = e^{c(\theta)\,T(x) + d(\theta) + S(x)}\, I_A(x),$$

  onde
  - $c(\cdot)$ e $d(\cdot)$ são funções reais de θ;
  - $T(\cdot)$ e $S(\cdot)$ são funções reais de x;
  - A **não depende** de θ.

Exercício 1: Verificar qual(is) das seguintes distribuições pertence(m) à família exponencial. i) $X \sim Bernoulli(\theta)$, $X \sim Normal(\mu, 1)$, $X \sim Uniforme(0, \theta)$.

- Teorema 1: Família exponencial **unidimensional** para uma **amostra aleatória** $X_1, \ldots, X_n$ de X. (Bolfarine e Sandoval, teorema 2.4.1). Provar!!!

- Definição 2: (**Família Exponencial** k **Dimensional** (Bolfarine e Sandoval, definição 2.4.2, pg. 27) Dizemos que a distribuição da variável aleatória X, com f.m.p ou f.d.p. dada por $f(x;\boldsymbol{\theta})$, pertence à família exponencial k dimensional, se pudermos escrever f como

$$f(x;\boldsymbol{\theta}) = e^{\sum_{j=1}^{k} c_j(\theta)\, T_j(x) + d(\theta) + S(x)}\, I_A(x).$$

- Teorema 2: Família exponencial k **dimensional** para uma **amostra aleatória** (Notas de Aula, definição 2.13, pg. 37). Provar!!!

Informação de Fisher na Família Exponencial

Teorema 3: Seja X uma variável aleatória tal que sua f.d.p. ou f.m.p. $f(x;\theta)$ pertence à família exponencial, e a **informação individual de Fisher** dada por

$$I_1(\theta) = E\left\{\left[\frac{\partial}{\partial}\log f(X;\theta)\right]^2\right\},$$

então vale a **igualdade da informação**

$$E\left\{\left[\frac{\partial}{\partial\theta}\log f(X;\theta)\right]^2\right\} = -E\left[\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right]$$

Provar!!!

Tarefa 1: Fazer os exercícios acima e provar os teoremas.

Tarefa 2: Ler os "Slides_aula13" para a próxima aula.
