3^a Prova - FECD B - 2013/02

Renato Assunção

Suponha que $Y_1, \ldots, Y_n$ forme uma amostra aleatória de v.a.'s i.i.d. com a seguintes densidade de probabilidade (caso contínuo): $f(y;\theta) = \theta e^{-\theta y}$ para $y > 0$ com $\theta \in (0, \infty)$ (densidade exponencial, contínua).

1. Encontre a estatística sufficiente para estimar o parâetro θ. DICA: Use o teorema da fatoração de Neyman-Fisher.
2. Mostre que o MLE de θ é uma função da estatística suficiente.

Extraído de CS281: Advanced Machine Learning, 2013, Harvard Let $p(k)$ be a one-dimensional discrete distribution that we wish to approximate, with support on non-negative integers. One way to do this is to fit an approximating distribution $q(k)$ that minimizes the Kullback-Leibler divergence between $p(k)$ and the approximation $q(k)$. We know that

$$KL(p||q) = \sum_{k=0}^{\infty} p(k) \log\left(\frac{p(k)}{q(k)}\right) = \mathbb{E}_p \log\left(\frac{p(X)}{q(X)}\right)$$

3. Show that when we select a model $q(k)$ as a Poisson(λ) distribution, this KL divergence is minimized by setting λ equal to the expected value of X under $p(k)$. OBS: Take the derivative of KL with respect to λ. For a Poisson(λ), $\mathbb{P}(X = k) = \lambda^k e^{-\lambda}/k!$.

IC para regressão logística com múltiplos regressores. Observamos $Y_1, \ldots, Y_n$ v.a.'s binárias independentes. A probabilidade $p_i = \mathbb{P}(Y_i = 1)$ varia nos exemplos i em função de k atributos medidos em cada exemplo: regressores ou variáveis independentes. Assumimos que

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik})}} = \frac{1}{1 + e^{-\mathbf{x}_i^t \,\cdot\, \boldsymbol{\theta}}}$$

onde $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ik})$ é o vetor de atributos medidos no exemplo i e $\boldsymbol{\theta} = (\beta_0, \beta_1, \ldots, \beta_k)$ é o vetor de parâmetros desconhecidos. Defina

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} \qquad \boldsymbol{\theta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

We obtain the MLE by means fo the Newton-Raphson algorithm:

$$\boldsymbol{\theta}^{\text{new}} = \boldsymbol{\theta}^{\text{old}} - \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t}\right)^{-1} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

onde as derivadas são avaliadas usando-se o valor corrente $\boldsymbol{\theta}^{\text{old}}$. It is not difficult to show that the $(k+1) \times (k+1)$ matrix of second derivatives of the log-likelihood ℓ is given by

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} = \mathbf{X}^t \mathbf{W} \mathbf{X}$$

where $\mathbf{W} = \text{diag}(p_1(1-p_1), \ldots, p_n(1-p_n))$. Suppose that the iterative process converges with the MLE resulting in $\hat{\boldsymbol{\theta}} = (-12.78, 1.48, 3.79, -2.3)$ and

$$\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t}\right)^{-1} = \left(\mathbf{X}^t \mathbf{W} \mathbf{X}\right)^{-1} = \begin{bmatrix} 1 & -1 & 2 & 0 \\ -1 & 4 & -1 & 1 \\ 2 & -1 & 6 & -2 \\ 0 & 1 & -2 & 4 \end{bmatrix}$$

4. Obtain the 95% confidence interval for θ_1. We know that, in a standard Gaussian Z, we have $\mathbb{P}(Z > 1.96) = 0.025$.