

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Regressão Logística: Uma Abordagem Bayesiana

Herlisson M. Bezerra¹

Instituto de Ciência e Tecnologia, Unifesp

Juliana G. Cespedes²

Instituto de Ciência e Tecnologia, Unifesp

Camila B. Martins³

Departamento de Medicina Preventiva, EPM, Unifesp

1 Introdução

As decisões de políticas de crédito das instituições financeiras, a partir de um passado recente, deixaram de ser guiadas apenas por critérios julgamentais e, atualmente, são fortemente norteadas por modelos quantitativos. Tais modelos são fundamentais para a mensuração e previsão do risco de crédito. Mensurar o risco de crédito é o processo de quantificar a possibilidade da instituição financeira incorrer em perdas, caso os fluxos de caixa esperados com a operação de crédito não se confirmem [1]. Para mensurar o risco de crédito, neste trabalho, propõe-se um modelo de *credit scoring* baseado em Regressão Logística utilizando Inferência Bayesiana.

2 *Credit Scoring*

Os modelos de *credit scoring* são utilizados no início do relacionamento com o cliente, período em que a instituição mensura previamente o risco do proponente e atribui a ele, ou não, linhas diferenciadas em função do seu perfil, para estimar a probabilidade dele se tornar inadimplente, em determinado período, dadas suas informações pessoais e financeiras que possam influenciar na capacidade do cliente em pagar a dívida [2].

Na modelagem de *credit scoring*, a variável resposta representa o desempenho de crédito dos indivíduos durante um determinado período tempo, normalmente doze meses, e tem o objetivo de identificar se o proponente é bom ou mau pagador. A variável resposta é dicotômica (binária), assumindo valor 1 se o indivíduo é mau pagador e 0, caso contrário. Nesse caso, a variável resposta tem distribuição de Bernoulli [3]. Dentre as várias funções de distribuição propostas para modelar resposta binária disponíveis na literatura, escolheu-se a Regressão Logística por sua facilidade de aplicação do ponto de vista matemático.

¹herlissonmaciel@hotmail.com²cespedes.juliana@gmail.com³cabertini@gmail.com

3 A Regressão Logística

A Regressão Logística modela a variável resposta binária através da relação entre a função de ligação logito, as variáveis explicativas $\tilde{x} = (1, x_1, \dots, x_{n-1})^T$ e os coeficientes a serem estimados $\tilde{\beta} = (\beta_0, \dots, \beta_n)^T$, conforme mostrado em (1).

$$\ln \left(\frac{\pi(\tilde{x})}{1 - \pi(\tilde{x})} \right) = \tilde{x}^T \tilde{\beta}. \quad (1)$$

A equação 1 pode ser manipulada para que $\pi(\tilde{x})$ seja isolado. Uma vez isolado e conhecendo-se as características do pagador \tilde{x} , $\pi(\tilde{x})$ pode ser interpretado como a probabilidade de o pagador não saldar sua dívida e, conseqüentemente, ser classificado como mau. Finalmente, ao relacionar a variável resposta com $\pi(\tilde{x})$, obtém-se:

$$Y = \pi(\tilde{x}) + \varepsilon,$$

sendo que Y é a variável resposta e ε é o erro do modelo para cada variação possível \tilde{x} .

Na Regressão Logística Bayesiana, os coeficientes são estimados através da inferência realizada a partir da posteriori, obtida da relação de proporção, derivada do teorema de Bayes:

$$h(\theta|x) \propto f(x|\theta)h(\theta),$$

sendo que $h(\theta)$ é a distribuição a priori, $f(x|\theta)$ é a função de verossimilhança, que está intimamente relacionada a distribuição dos dados e a sua dimensionalidade, $h(\theta|x)$ é a distribuição a posteriori, θ é o parâmetro de interesse e x os dados utilizados para ajuste.

4 Conclusões

O modelo proposto foi ajustado utilizando dados reais de uma instituição financeira. Para sua validação, observou-se a proporção de pagadores classificados corretamente no grupo ao qual de fato pertencem. Calculou-se ainda o índice de Kolmogorov-Smirnov e a área sob a curva ROC do modelo ajustado. O modelo teve desempenho compatível com o que consta na literatura [2] [3] ao classificar corretamente 63,73% dos maus pagadores.

Referências

- [1] G. S. Brito e A. A. Neto. Modelo de classificação de risco de crédito de empresas, Revista Contabilidade e Finanças, vol. 19, 18-29, (2008).
- [2] C. Karcher, Redes Bayesianas aplicadas à análise do risco de crédito, Dissertação de Mestrado em Engenharia, Universidade de São Paulo, (2008).
- [3] T. S. Mendonça, Modelos de Regressão Logística Clássica, Bayesiana e Redes Neurais para Credit Scoring, Dissertação de Mestrado em Estatística, Universidade Federal de São Carlos, (2008).