

Tópicos Especiais em Computação I

Regressão Logística

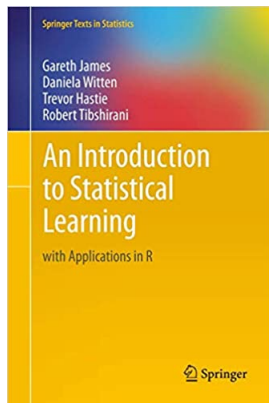
Patrícia Lucas

Bacharelado em Sistemas de Informação
IFNMG - Campus Salinas

Salinas
Abril 2021

Referência

Regressão Logística

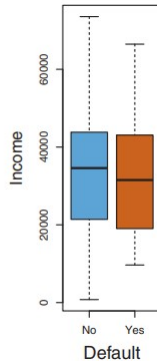
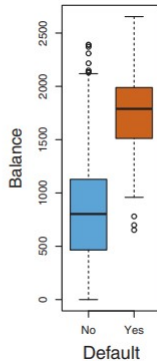
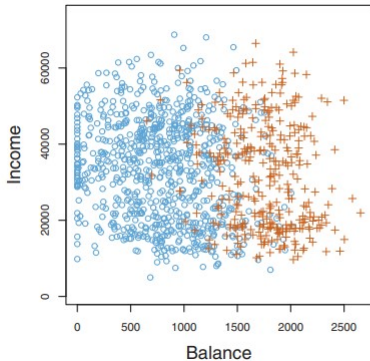


Capítulo 4: Logistic Regression

An Introduction to Statistical Learning: with Applications in R. G. James, D. Witten, T. Hastie, and R. Tibshirani. Springer, 2013.

Exemplo

Regressão Logística



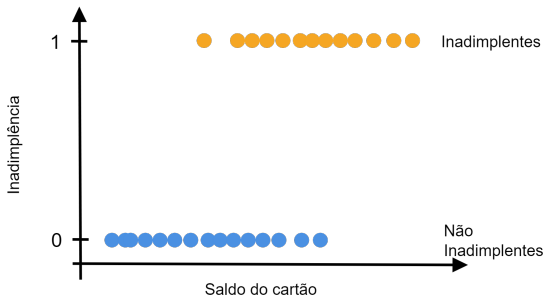
Exemplo

Regressão Logística

Vamos supor que queremos classificar os clientes como inadimplentes ou não inadimplentes de acordo com o saldo do cartão.

Ou seja, queremos encontrar uma curva que nos dê a probabilidade de um cliente ser inadimplente com base no seu saldo do cartão.

$$P(\text{inadimplência} = \text{Sim} | \text{Saldo}) \quad (1)$$



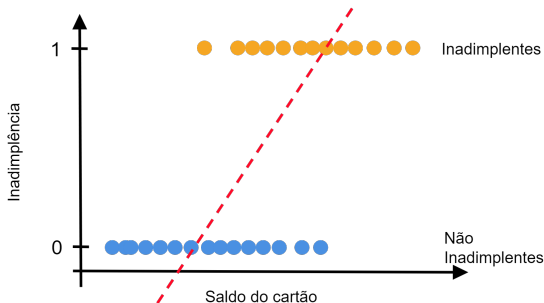
Regressão linear??

Regressão Logística

Podemos usar regressão linear para resolver esse problema?

- Para saldos próximos a zero, prevemos uma probabilidade negativa de inadimplência.
- Com saldos muito grandes, obteríamos valores maiores que 1.

Essas previsões não são sensatas, pois é claro que a verdadeira probabilidade de inadimplência, independentemente do saldo do cartão de crédito, deve cair entre 0 e 1.



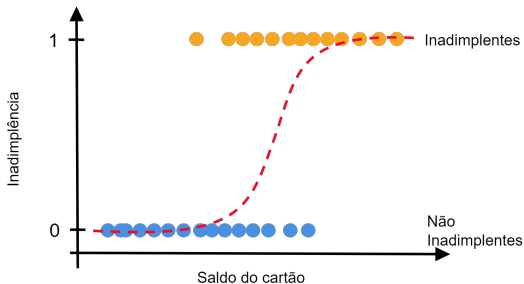
Função logística

Regressão logística

Para evitar esse problema, precisamos modelar $P(X)$ usando uma função que fornece saídas entre 0 e 1 para todos os valores de X .

Na regressão logística, usamos a **função logística**, que ajusta uma curva em “S” que varia entre 0 e 1.

$$P(X) = \frac{\exp \beta_0 + \beta_1 X}{1 + \exp \beta_0 + \beta_1 X} \quad (2)$$



Estimativa dos coeficientes

Regressão logística

Os coeficientes β_0 e β_1 são desconhecidos e devem ser estimados com base nos dados de treinamento disponíveis.

Usamos a abordagem de mínimos quadrados para estimar os coeficientes de regressão linear. Para a regressão logística, usa-se geralmente o método de máxima verossimilhança.

A intuição básica por trás do método de máxima verossimilhança é tentar encontrar β_0 e β_1 de modo que, ao incluir essas estimativas no modelo para $P(X)$, produzam um número próximo de 1 para todos os indivíduos que são inadimplentes e um número próximo de 0 para todos os indivíduos que não são.

Estimativa dos coeficientes

Regressão logística

$$l(\beta_0, \beta_1) = \prod_{i: y_i=1} p(x_i) \prod_{i': y_{i'}=0} p(x_{i'}) \quad (3)$$

