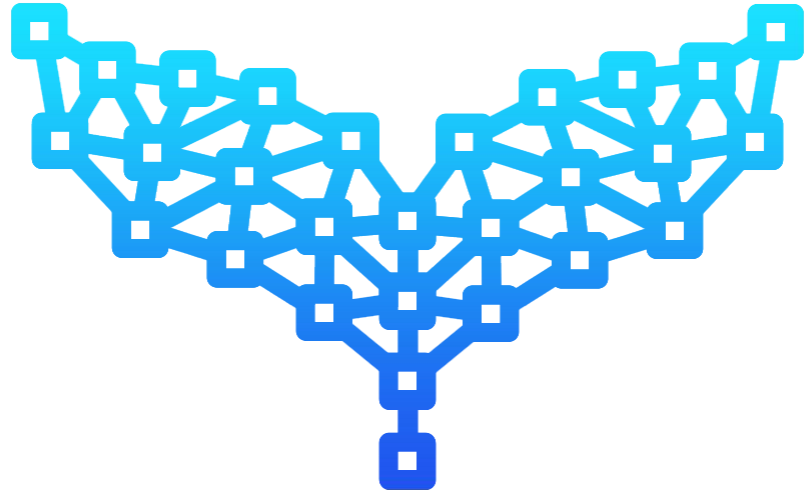
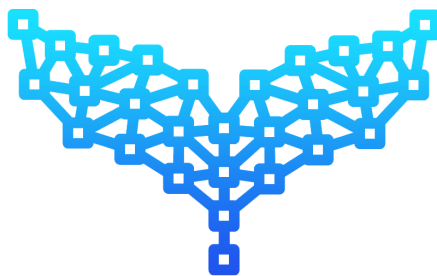


TAIL

Technology and Artificial
Intelligence League



REGRESSÃO LOGÍSTICA



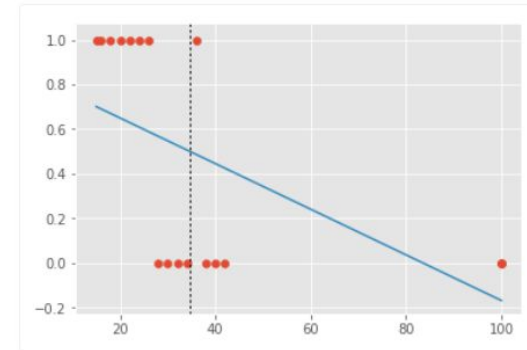
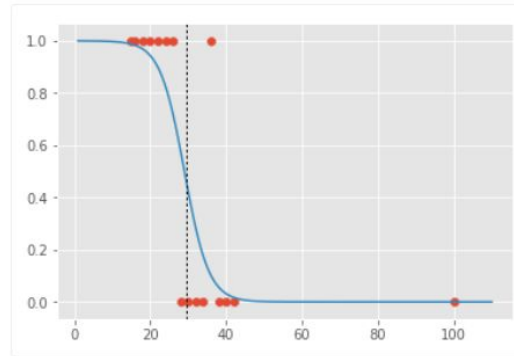


1. Introdução

A regressão logística é uma técnica utilizada para determinar a probabilidade de um objeto pertencer ou não a uma determinada classe a partir de um ou mais valores apresentados, ou seja, o objetivo é atribuir um rótulo a um objeto.

Alguns exemplos em que pode ser aplicada:

- Em medicina, permite determinar os fatores que caracterizam um grupo de indivíduos doentes em relação a indivíduos saudáveis.
- Em instituições financeiras, pode detectar os grupos de risco para a subscrição de um crédito.



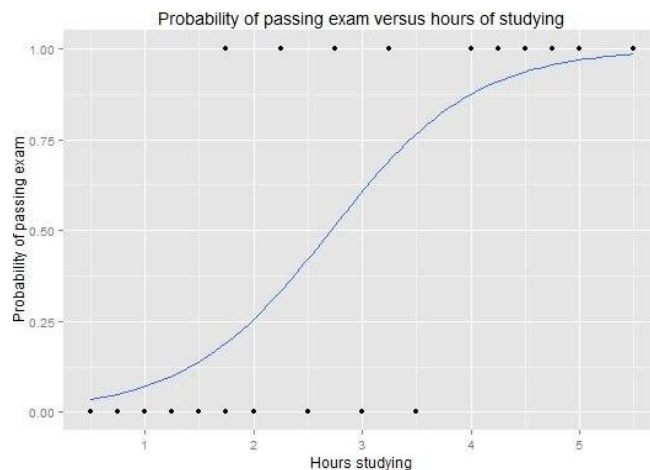
Exemplos gráficos de classificação por regressão logística (esquerda) e regressão linear (direita)

2. Exemplo



Pensemos na seguinte situação: Determinado estudante irá fazer uma prova, com isso, este mesmo aluno quer saber qual a probabilidade de passar na prova, baseado em suas horas de estudo

Nesse caso podemos observar que:
A probabilidade de passar na prova é o rótulo;
Quanto mais horas de estudo, a probabilidade de passar aumenta.



Probabilidade de se passar numa prova
baseada nas horas de estudo

3. Função Logística



A função logística, ou função sigmoid, é um tipo de função de ativação a qual limita a saída em um range binário (0 e 1), o que faz dessa função ideal para o cálculo de probabilidades.

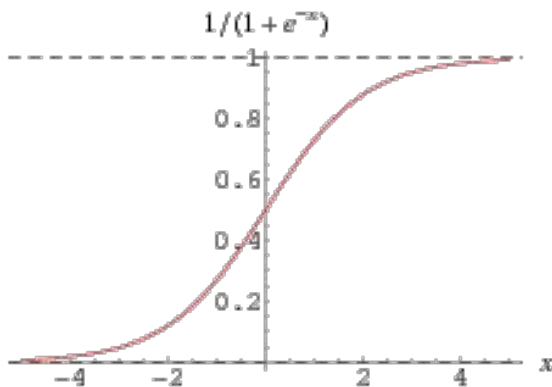


Gráfico da função logística



4. Cálculo da Regressão Logística

Para estimar os valores na regressão, utilizamos a seguinte equação:

$$y = b_0 + b_1 * x_1$$

Onde:

- y é o atributo que desejamos prever;
- b_0 é uma constante;
- b_1 é um coeficiente;
- x_1 é a variável independente.



5. Exemplo

Suponhamos que haja uma base de crédito com os seguintes atributos previsores:

- Histórico de pagamento = Bom (0)
- Dívida = Média (0)
- Garantias = Nenhuma (1)
- Renda anual => 30 (2)

Em seguida, iremos elencar coeficientes b_n para cada atributo.

- $b_0 = -0,12$
- $b_1 = -0,71$
- $b_2 = 0,24$
- $b_3 = -0,54$
- $b_4 = 1,07$



6. Exemplo

Calculando o valor de y , temos que

$$\begin{aligned}y &= -0,12 + (b_1 * \text{Histórico de pagamento}) + (b_2 * \text{Dívida}) + (b_3 * \text{Garantias}) + (b_4 * \text{Renda anual}) \\y &= -0,12 + (-0,71 * 0) + (0,24 * 0) + (-0,54 * 1) + (1,07 * 2) \\y &= 1,48\end{aligned}$$

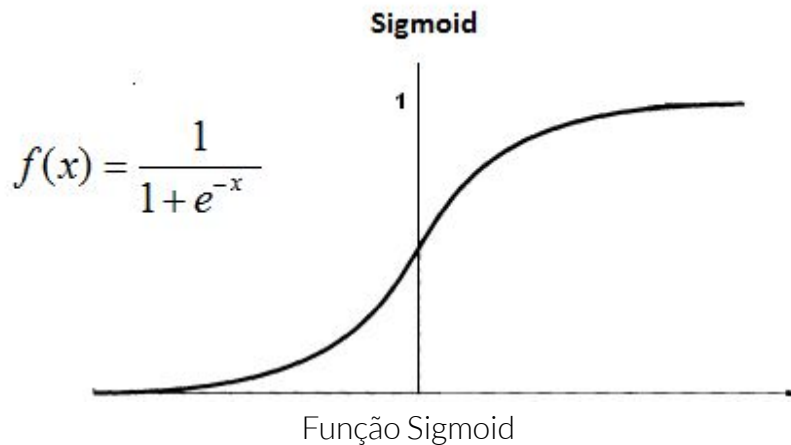
Em seguida, utilizaremos o valor de y como entrada para a função sigmoid. Logo,

$$\begin{aligned}f(y) &= \frac{1}{1 + e^{-y}} \\f(1,48) &= \frac{1}{1 + e^{-1,48}} \\f(1,48) &= 0,8145\end{aligned}$$

7. Exemplo



Concluimos assim que, o usuário com esse perfil teria 81% de probabilidade de pagar determinado empréstimo.



Referências



FACURE, Matheus. Regressão Logística. Disponível em:
<<https://matheusfacure.github.io/2017/02/25/regr-log/>>. Acesso em 18 de Setembro de 2020;

WOLFRAM MATHWORLD. Sigmoid Function. Disponível em:
<<https://mathworld.wolfram.com/SigmoidFunction.html>>. Acesso em 18 de Setembro de 2020;

DEEPAI. Sigmoid Function. Disponível em:
<<https://deepai.org/machine-learning-glossary-and-terms/sigmoid-function>>. Acesso em 18 de Setembro de 2020.