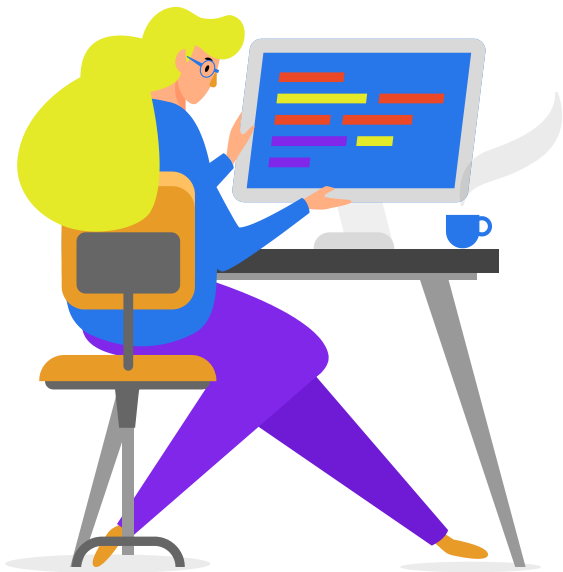


Regressão Logística

Professor: Gabriel
Oliveira Assunção

Modelos de classificação



Objetivo **categorizar** uma nova observação em classes pré determinadas.

Modelos podem retornar a classe ou a probabilidade de se estar em uma classe.

Modelo de Regressão

Inputs

Features

Conjunto de dado conhecido



Equação

$$Y = a + bX$$



Outputs

Target

Predição do target,
dado desconhecido



Modelo de Regressão

Inputs

Features

Conjunto de dado conhecido



Equação

$$\text{Prob} = f(x)$$



Outputs

Target

Predição da probabilidade ou rótulo



Modelos de regressão

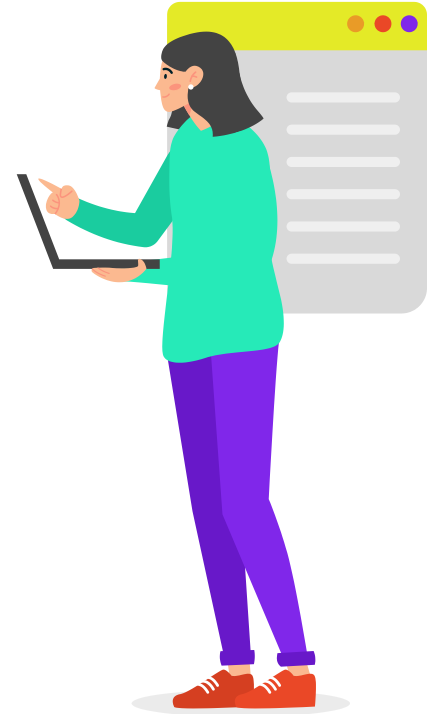
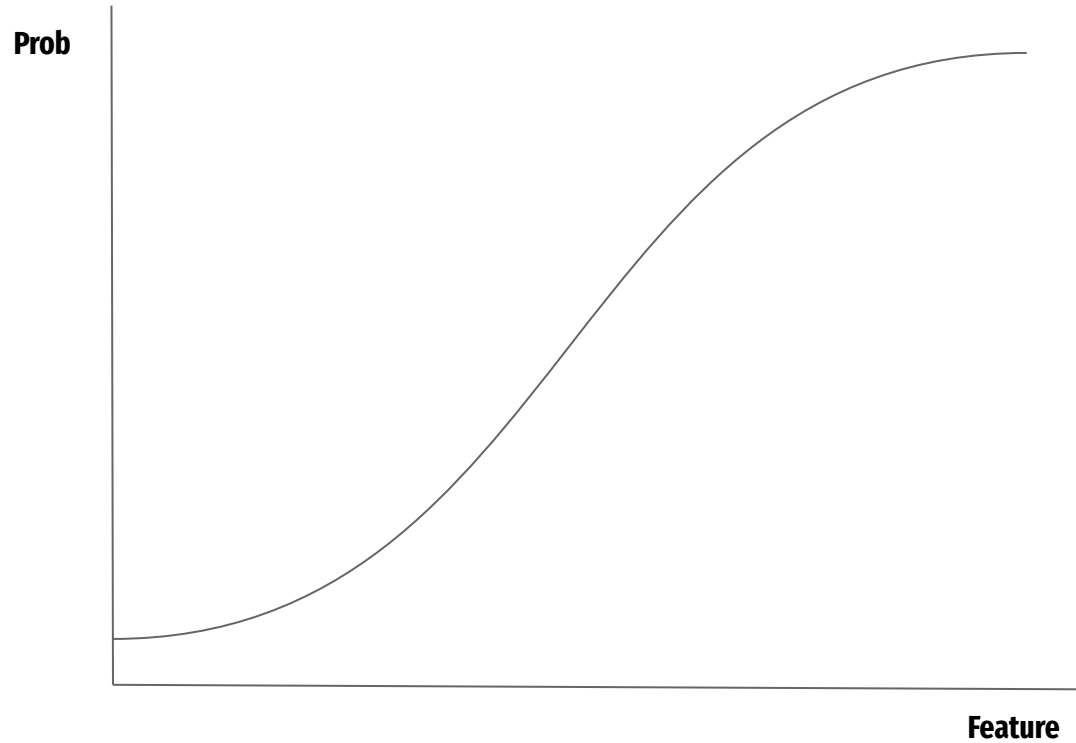
Regressão Linear

- Consideramos linearidade entre features e variáveis respostas
- Suposição de normalidade

Regressão Logística

- Consideramos linearidade sobre as probabilidades
- Não há suposição de normalidade

Análise da curva



Alguns conceitos

Função Sigmoid

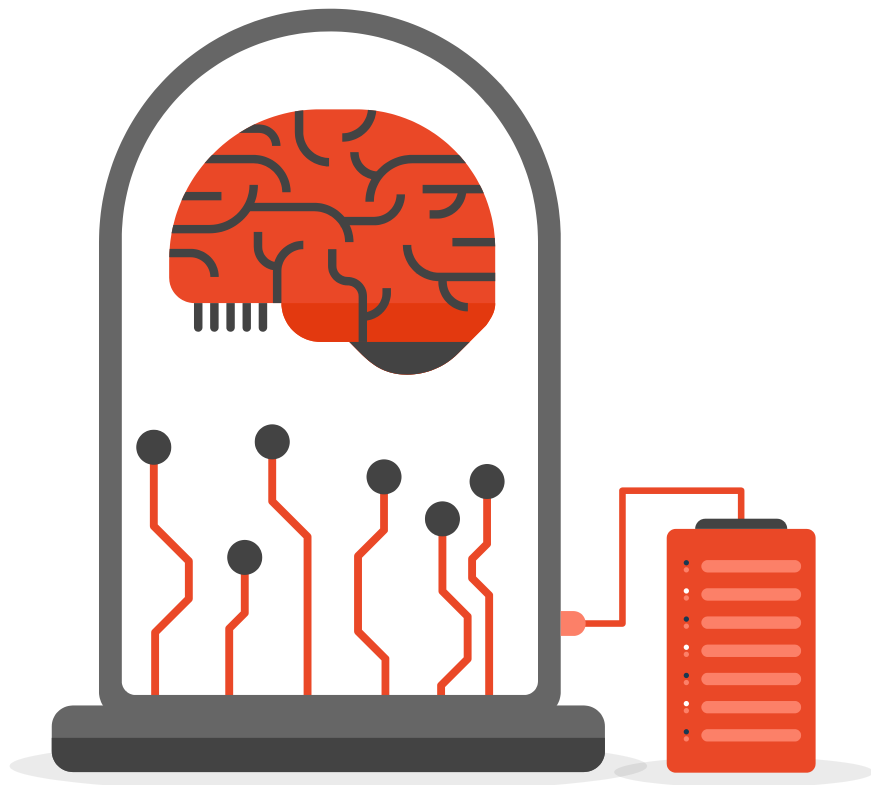
$$Prob(Y = 1) = \frac{1}{1 + \exp(-(a + bx + cz + \dots))}$$

01

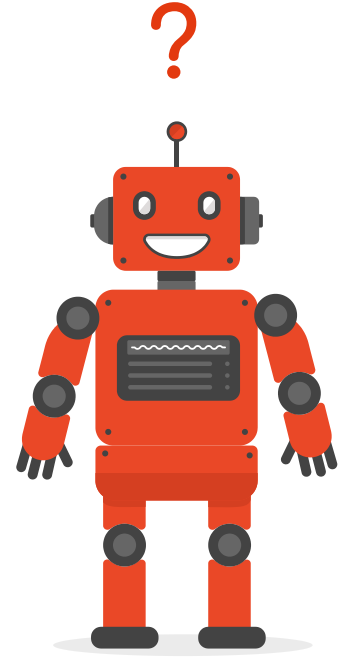
Função de perda log-loss

$$\frac{1}{n} \sum y \times \log(p) + (1 - y) \times \log(1 - p)$$

02



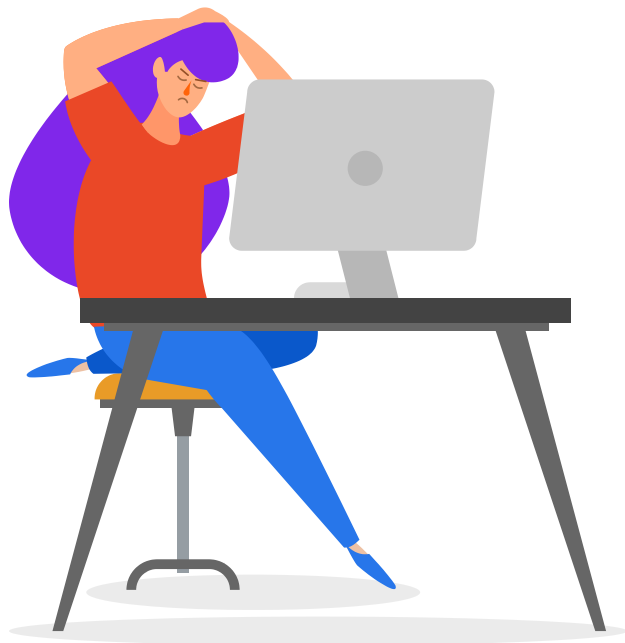
Como interpretar os coeficientes?



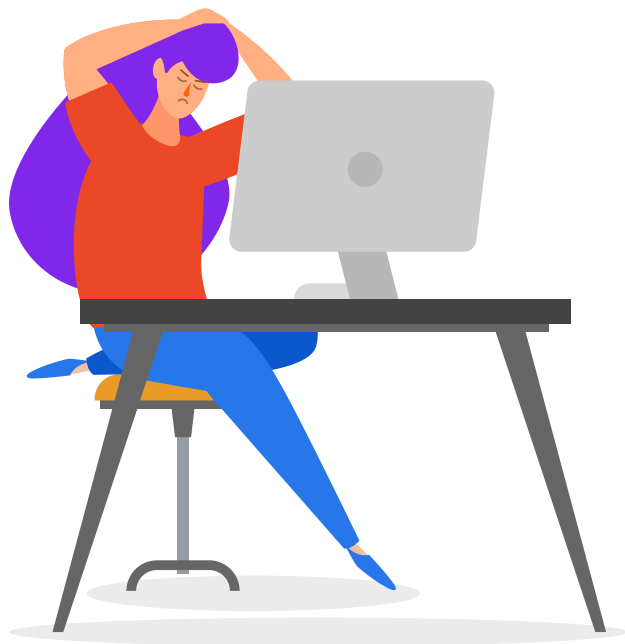
Alguns conceitos

Chance: Razão entre a probabilidade do rótulo positivo ($Y=1$) sobre a probabilidade rótulo negativo

$$O(x) = \frac{Prob(Y=1|x)}{Prob(Y=0|x)} = \exp(a + bx + cz + \dots)$$



Alguns conceitos



Razão de chances: divisão de duas razões onde se altera o valor de uma unica feature

$$\frac{o(x+1)}{o(x)} = \exp(b)$$

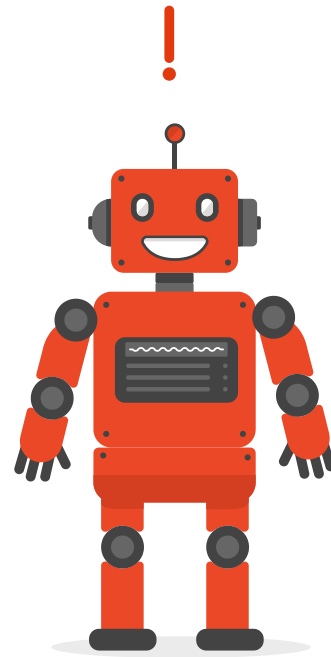
Interpretação do coeficiente:

Se aumentar uma unidade da nossa variável x a chance é multiplicada por $\exp(b)$

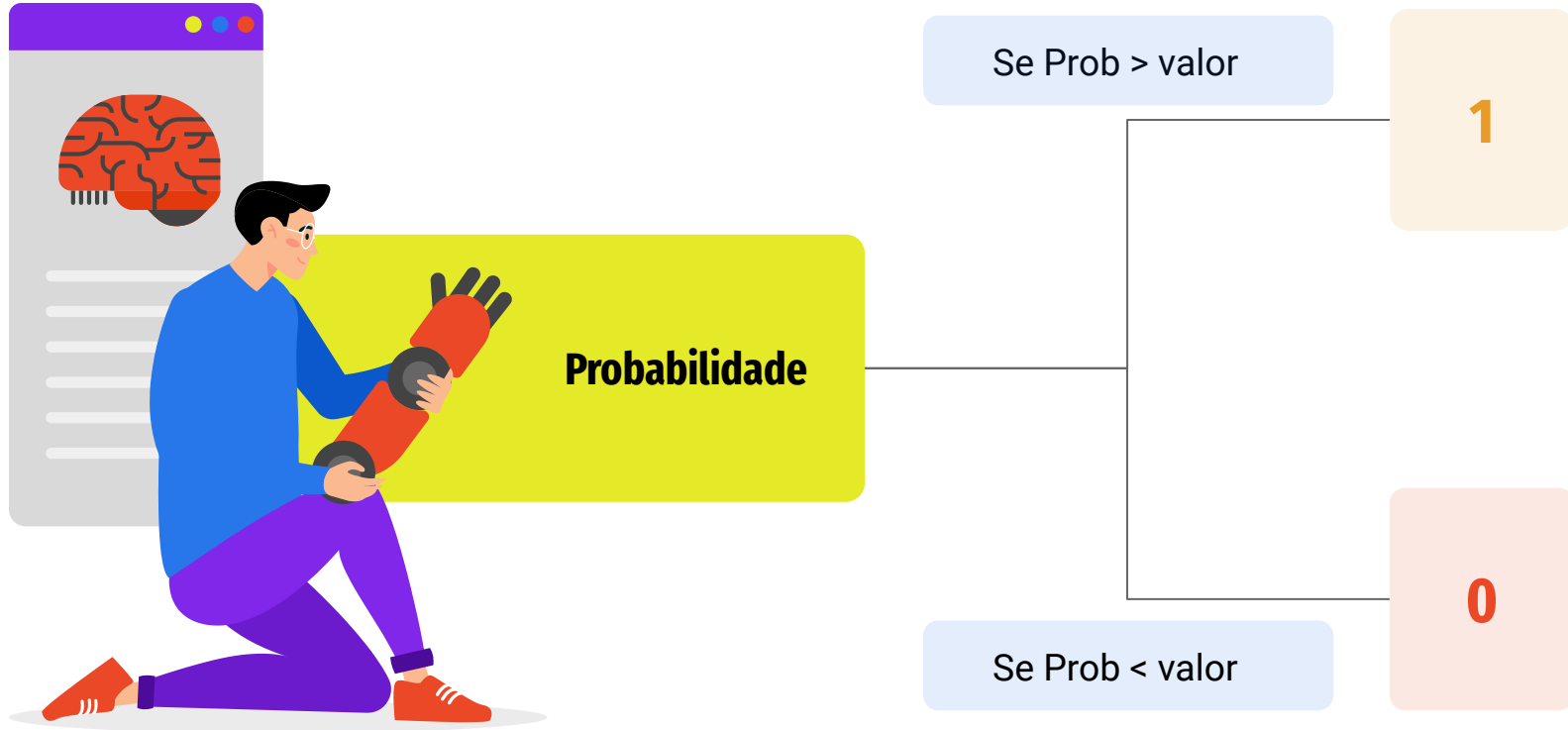
O que isso quer dizer?

Se $b > 0$: ao aumentar uma unidade da variável a chance aumenta

Se $b < 0$: ao aumentar uma unidade da variável a chance diminui

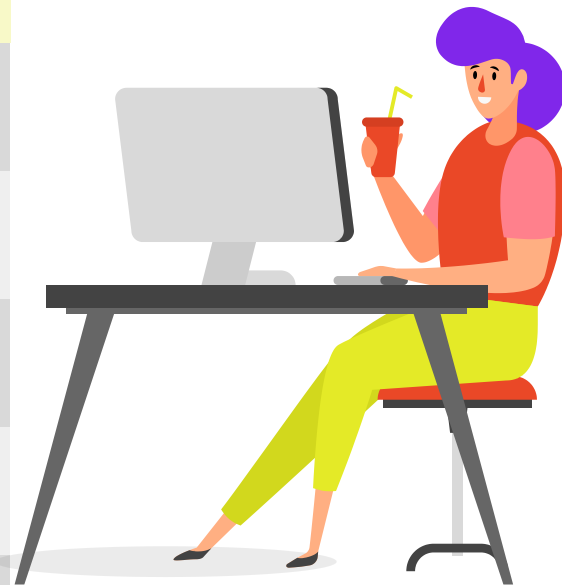


Classificador



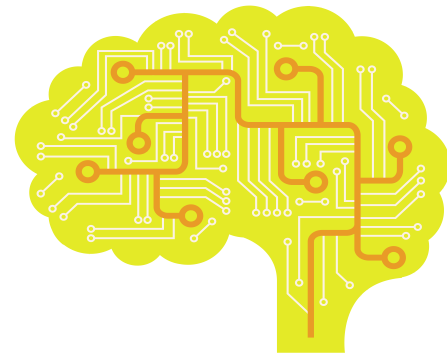
Métricas de classificação

Métricas	
Acurácia	Proporção de classificados corretamente
Precisão	Proporção dos preditos para um rótulo que de fato são do rótulo
Recall	Proporção dos dados de um rótulo que foram preditos corretamente
F1-score	Média Harmônica entre recall e precisão
AUC	Qualidade do modelo independente do critério de classificação



Matriz de confusão

		Rótulo Real	
		Neg	Pos
Classificado	Neg	Verdadeiro Negativo	Falso Negativo
	Pos	Falso Positivo	Verdadeiro Positivo



Matriz de confusão

Rótulo Real			
		Neg	Pos
Classificado	Neg	Verdadeiro Negativo	Falso Negativo
	Pos	Falso Positivo	Verdadeiro Positivo

Recall*, especificidade, razão de verdadeiro negativo

Recall, sensibilidade, razão de verdadeiro positivo

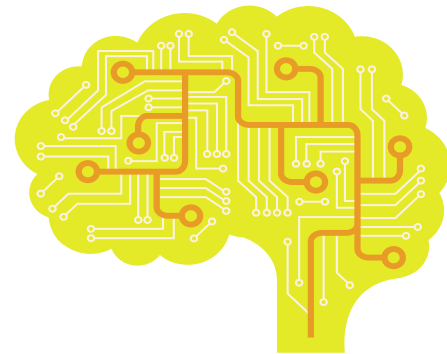
Matriz de confusão

Rótulo Real			
		Neg	Pos
Classificado	Neg	Verdadeiro Negativo	Falso Negativo
	Pos	Falso Positivo	Verdadeiro Positivo
		Precisão	Precisão

Matriz de confusão

		Rótulo Real	
		Neg	Pos
Classificado	Neg	10	5
	Pos	5	10

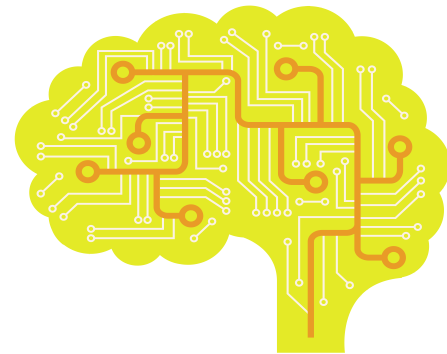
Limiar: 0.5; Precisão: 0.66; Recall: 0.66



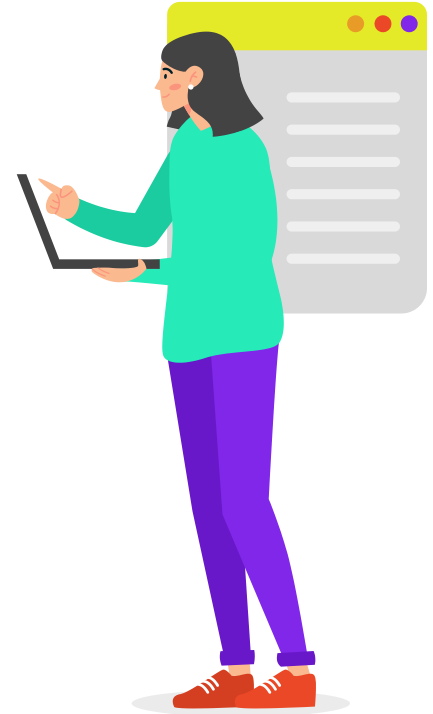
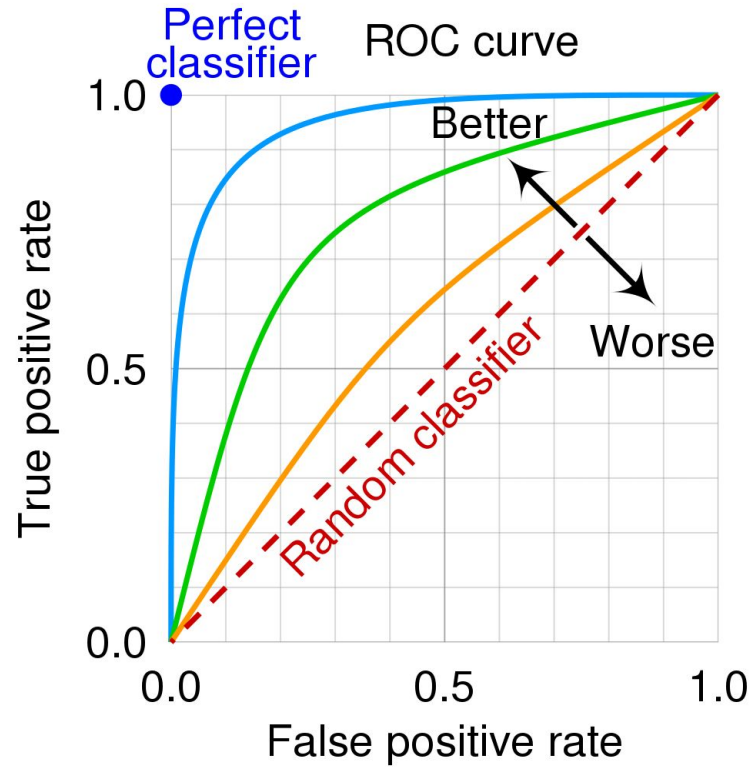
Matriz de confusão

		Rótulo Real	
		Neg	Pos
Classificado	Neg	12	6
	Pos	3	9

Limiar: 0.7; Precisão: 0.75; Recall: 0.6



Curva ROC





Hora da prática!!!!