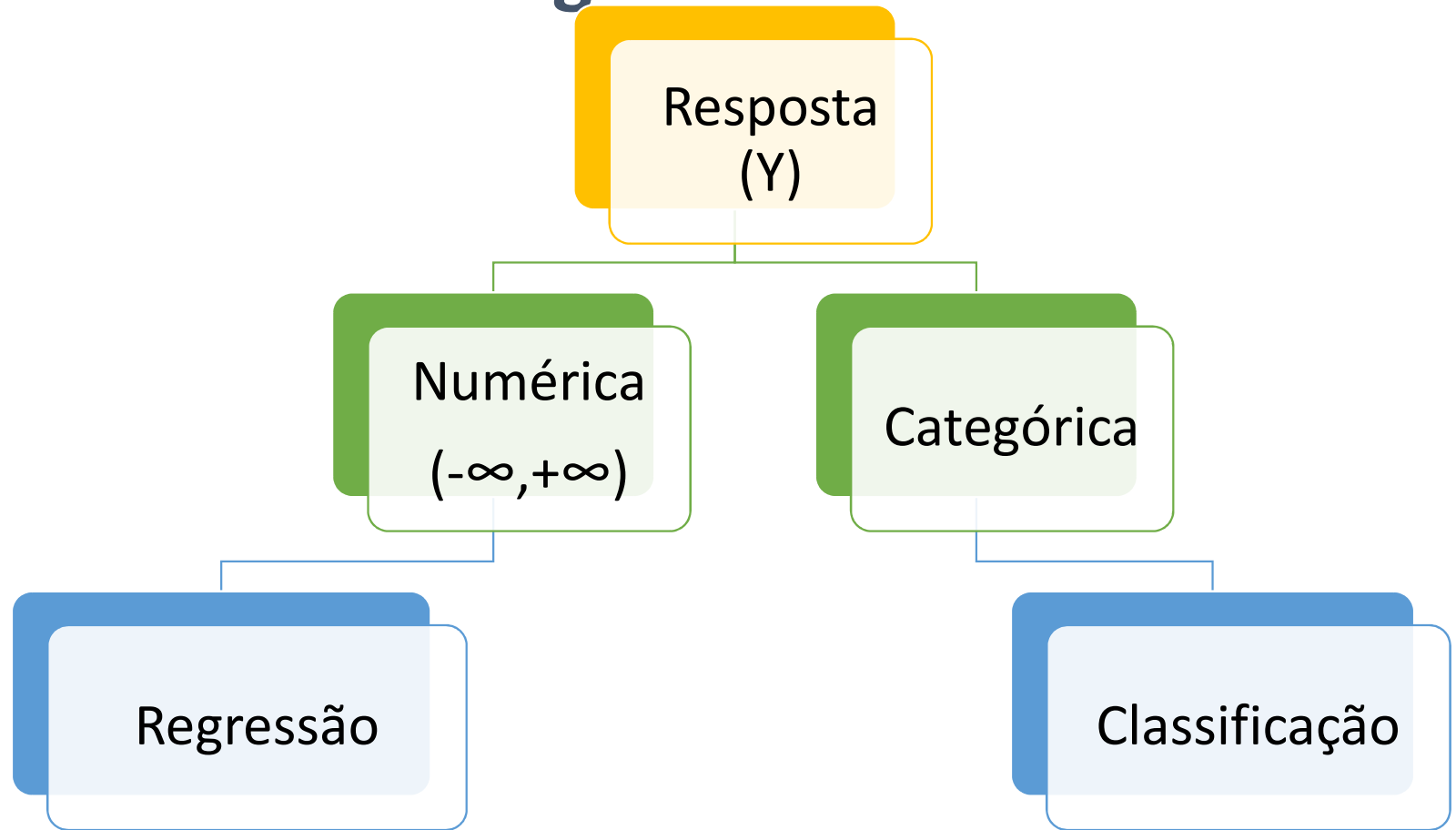


GLM para dados de binários ou proporções

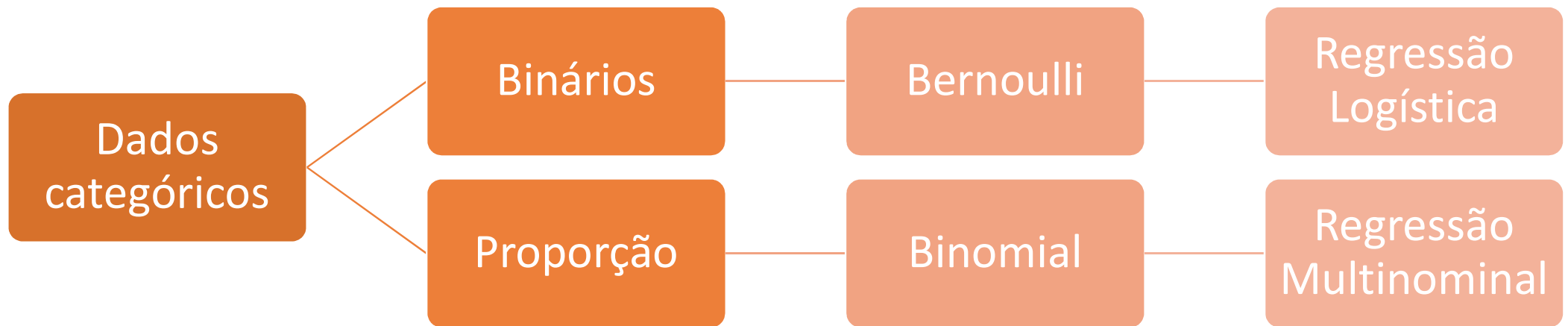
Modelos de Machine Learning



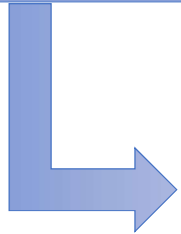
GLM para dados de binários ou proporções

Interesse em modelar fenômenos aleatórios com dois ou mais rótulos:

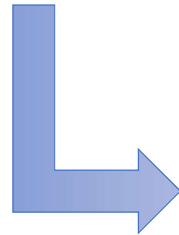
Ex: 1: Sucesso ; 0: Fracasso



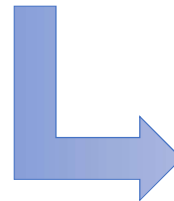
**Variáveis
Categóricas**



Binárias



**Estudar a
probabilidade de
ocorrência**



**Obtemos uma
resposta numérica
(probabilidade)**

**Probabilidade
Valores estão
entre 0 e 1**

**GLM para dados de
binários ou proporções**

Regressão logística

- Técnica de classificação usada para prever uma resposta qualitativa

Regressão Logística

Exemplos

- Fraudes – Probabilidade da transação ser fraudulenta ou não. **Classificação** usada para prever uma resposta qualitativa (binária)
- Marketplace - probabilidade do cliente comprar ou não a mercadoria

Por que não uma regressão linear?

Não podemos utilizar uma regressão linear para prever eventos categóricos.

Ex: Condição médica dos pacientes:

- AVC
- Parada Cardíaca
- Overdose

Podemos simplesmente ordenar e realizar a regressão linear?

1 = AVC

2 = Parada Cardíaca

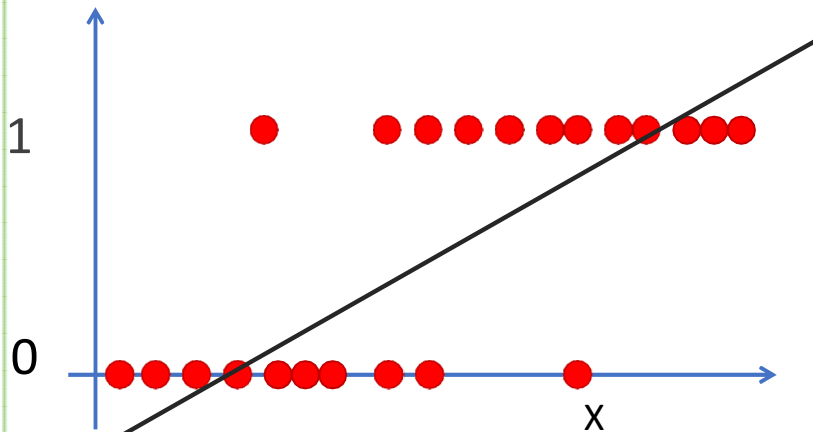
3 = Overdose

Devemos utilizar as variáveis dummy e transformar as variáveis.

Regressão Logística - Definição Teórica

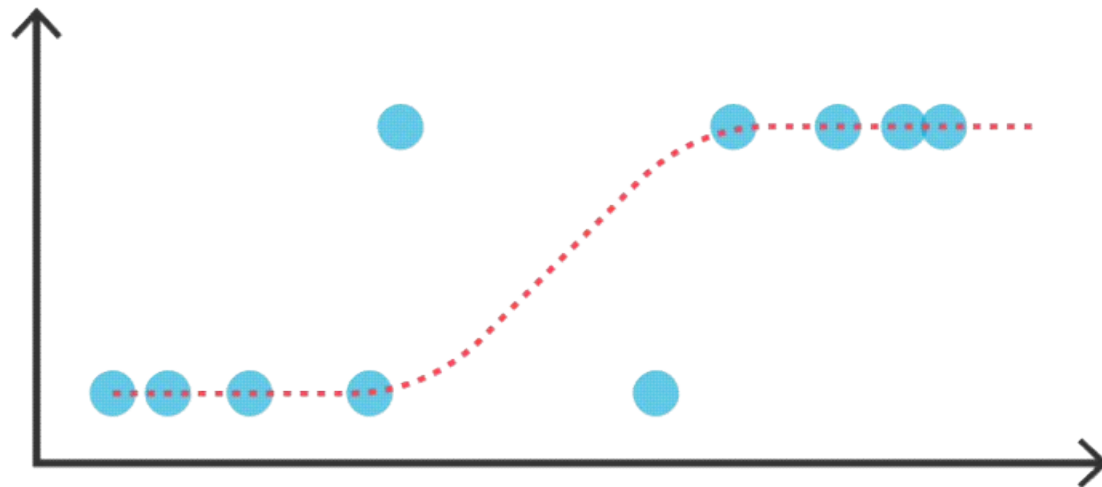
Permite **estimar a probabilidade** associada à **ocorrência de determinado evento** em vista de um conjunto de variáveis preditoras.

- Probabilidade de sucesso (1)
- Probabilidade de fracasso (0)



Ao interpretar Y como probabilidade, temos que realizar transformações para que a resposta de nossa regressão esteja entre 0 e 1.

Regressão para dados binários



Função Logística

- Retorna os valores entre 0 e 1
- Tem formato de “S”

Regressão para dados binários

Considerando $Y_1, Y_2, Y_3, \dots, Y_n$ variáveis aleatórias independentes tais que

$$Y_i | x_i \sim \text{bin}(m_i, \pi_i), i = 1, \dots, n$$

Em que $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ os vetores de variáveis explicativas associadas a cada observação

Logo podemos especificar o GLM da seguinte maneira:

$$g(\pi_i) = \eta_i = X_i' B$$

Assumimos que a variável resposta é uma variável aleatória de Bernoulli, com função de Probabilidade:

y_i	Probabilidade
1	$P(y_i = 1) = p_i$
0	$P(y_i = 0) = 1 - p_i$

- Uma vez que:

$$E(y_i) = 1(p_i) + 0(1 - p_i) = p_i$$

- Temos que:

$$E(y_i) = x'_i \beta = p_i$$

Logo, a **resposta encontrada na regressão logística** sempre será a **probabilidade de sucesso (1)**



PUC Minas
Virtual