

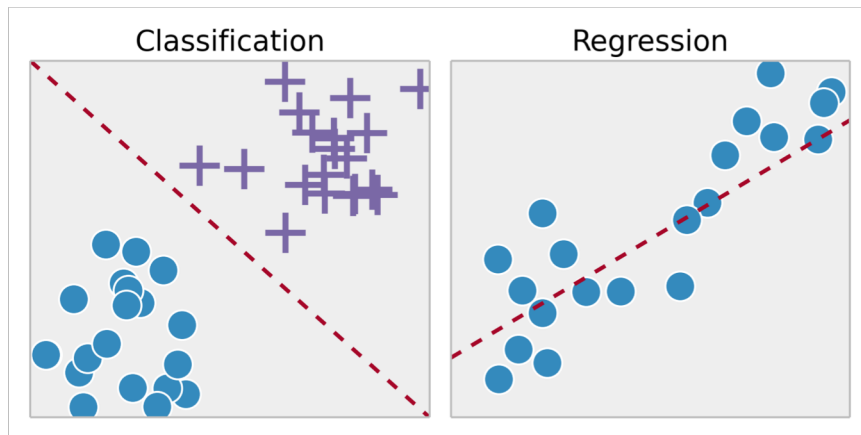
Regressão linear

Jones Granatyr



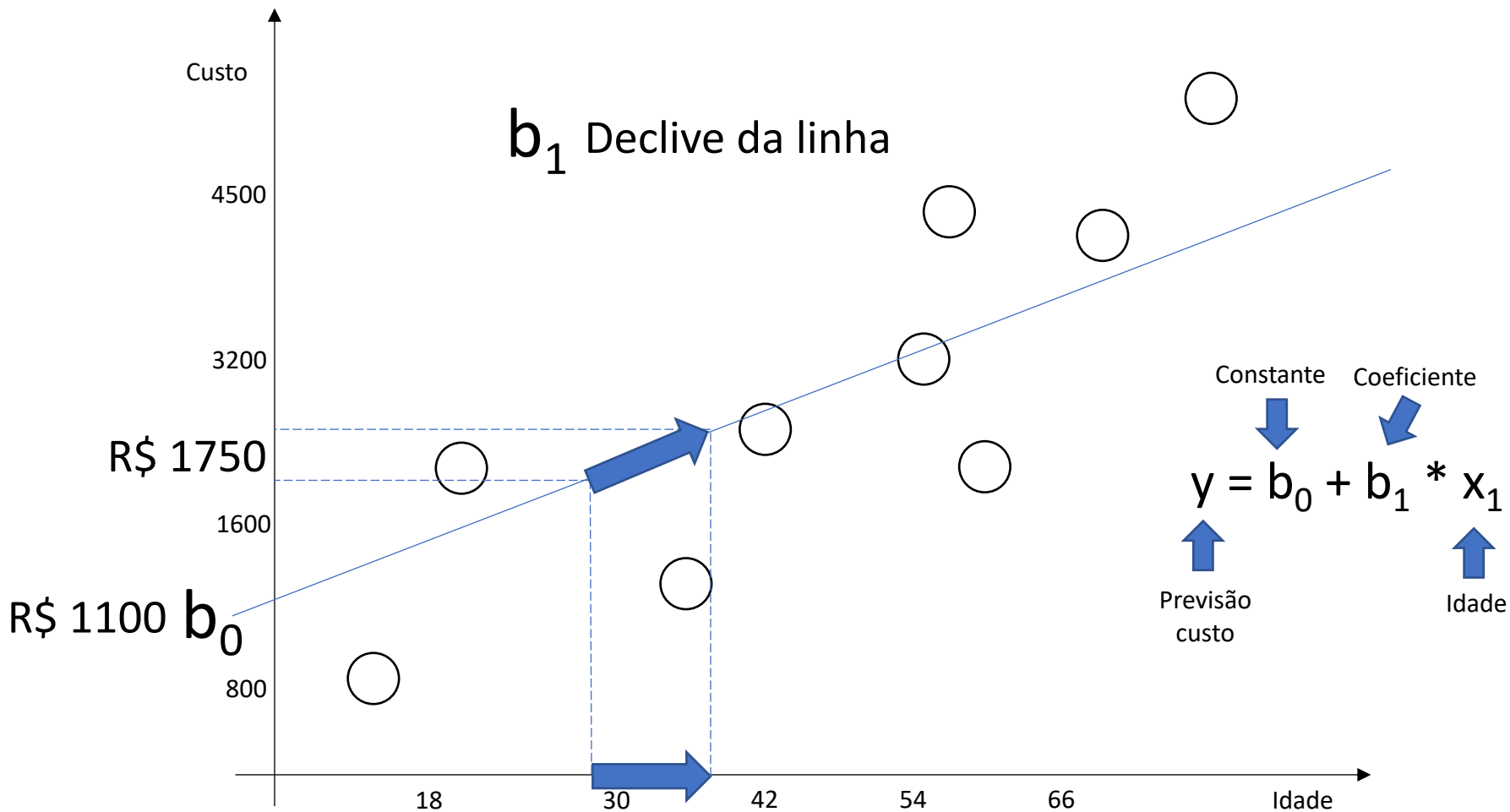
Regressão linear

- Modelagem da relação entre variáveis numéricas (variável dependente y e variáveis explanatórias x)
- Temperatura, umidade e pressão do ar (x) \rightarrow velocidade do vento (y)
- Gastos no cartão de crédito, histórico (x) \rightarrow limite do cartão (y)
- Idade (x) \rightarrow custo plano de saúde (y)
- Tamanho da casa (x) \rightarrow preço da casa (y)

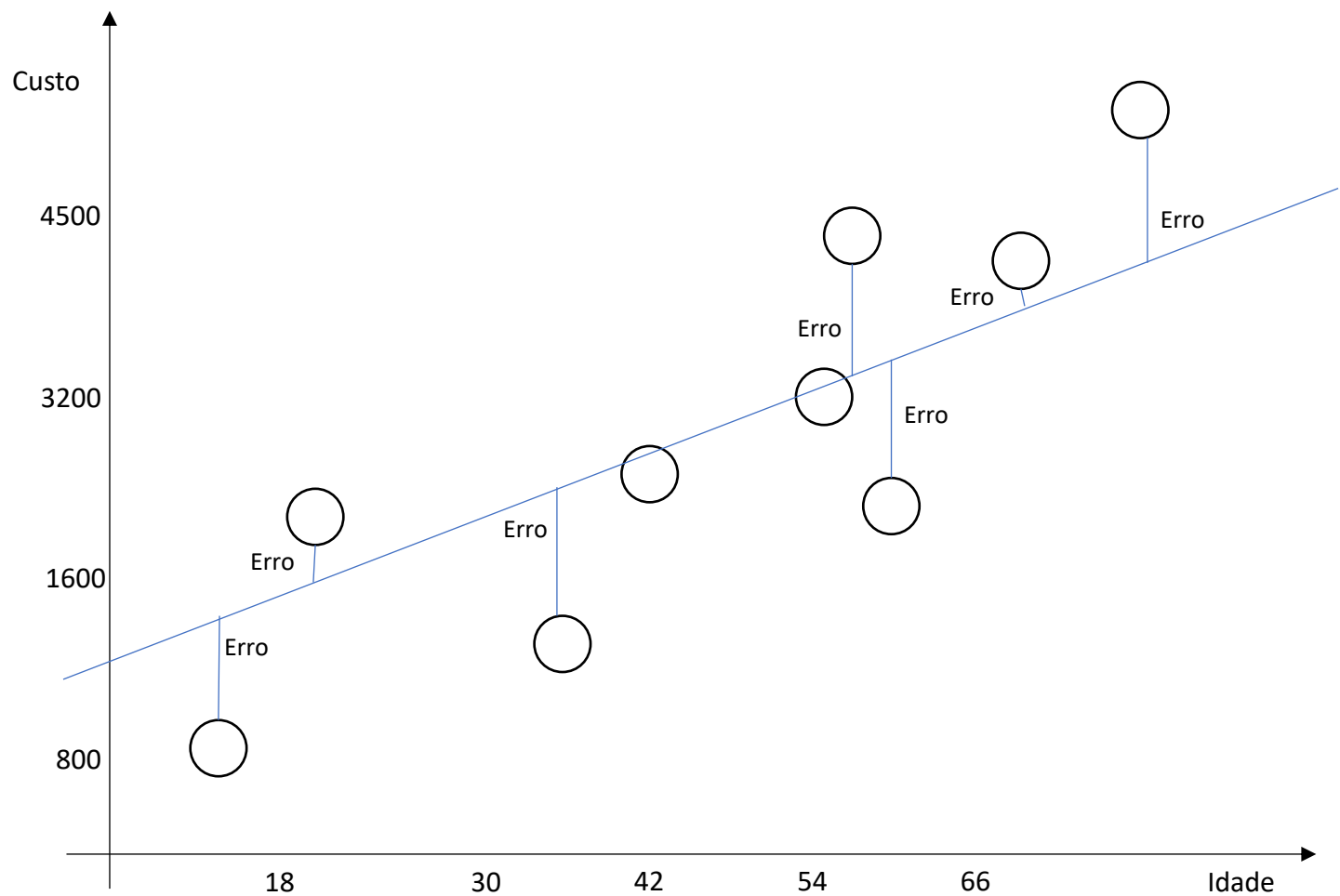


Regressão linear

Relação linear entre os atributos: quanto maior a idade, maior o custo
 b_0 e b_1 definem a localização da linha (treinamento)



Regressão linear



Mean square error (MSE)

$$MSE = \frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2$$

| Preço real | Preço calculado | Erro |
|------------|-----------------|-----------------------|
| 150 | 180 | $(150 - 180)^2 = 900$ |
| 60 | 55 | $(60 - 55)^2 = 25$ |
| 220 | 230 | $(220 - 230)^2 = 100$ |
| 45 | 67 | $(45 - 67)^2 = 484$ |

Soma = 1.509

$MSE = 1.509 / 4 = 377,25$

$y = b_0 + b_1 * x_1$ Objetivo: ajustar os parâmetros b_0 e b_1 para ter o menor erro!

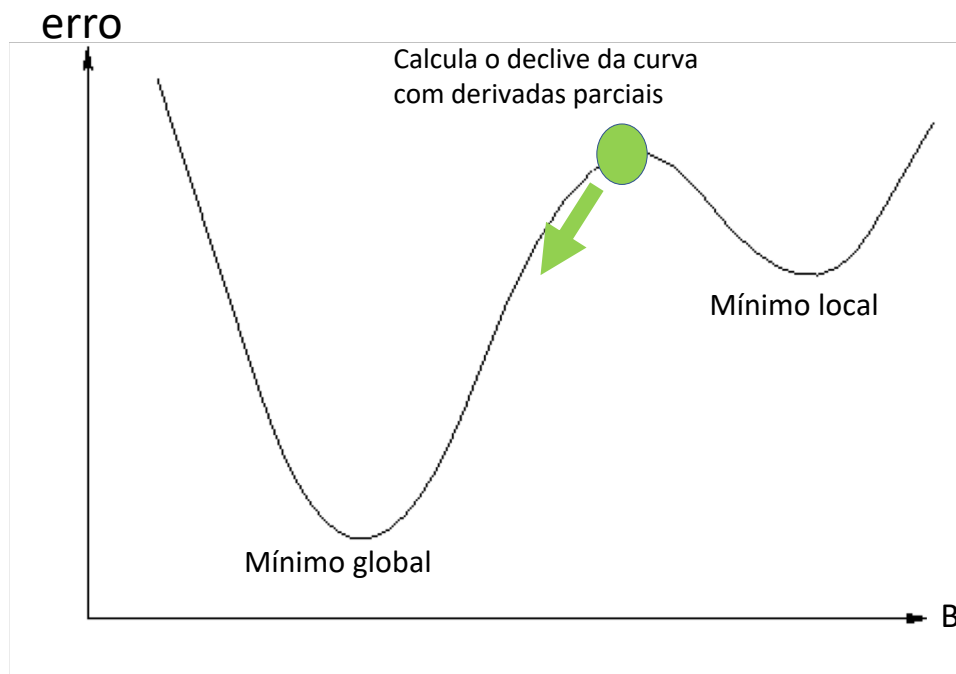
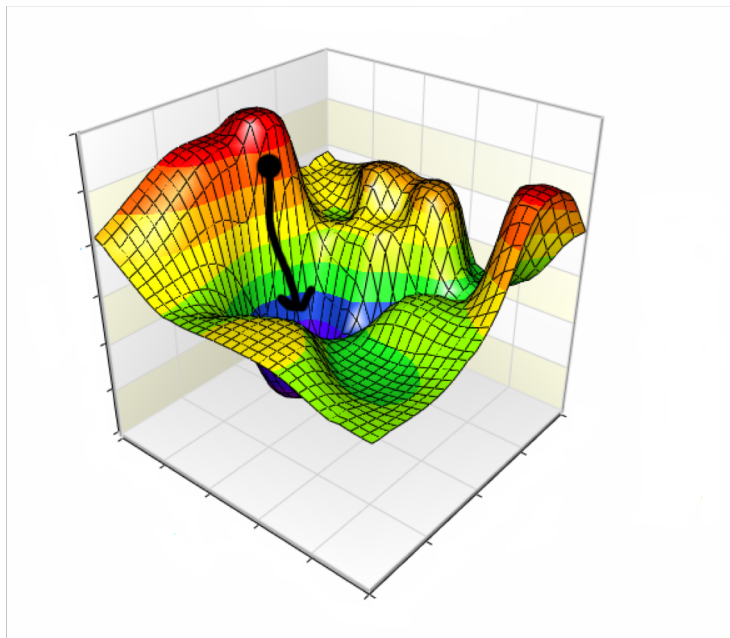
Regressão linear – ajuste dos parâmetros

- Design matrix (Álgebra Linear)
 - Bases de dados com poucos atributos
 - Inversão de matrizes que tem um custo computacional alto
- Gradient descent (descida do gradiente)
 - Desempenho melhor com muitos atributos

Descida do gradiente

$$\min C(B_1, B_2 \dots B_n)$$

Taxa de aprendizagem



Regressão linear múltipla

$$y = b_0 + b_1 * x_1$$

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$$

Conclusão

