



Regressão logística

Jones Granatyr

Aprendizagem de Máquina

Métodos Preditivos	Métodos Descritivos
Classificação	Associação
Regressão	Agrupamento
	Detecção de desvios
	Padrões sequenciais
	Sumarização

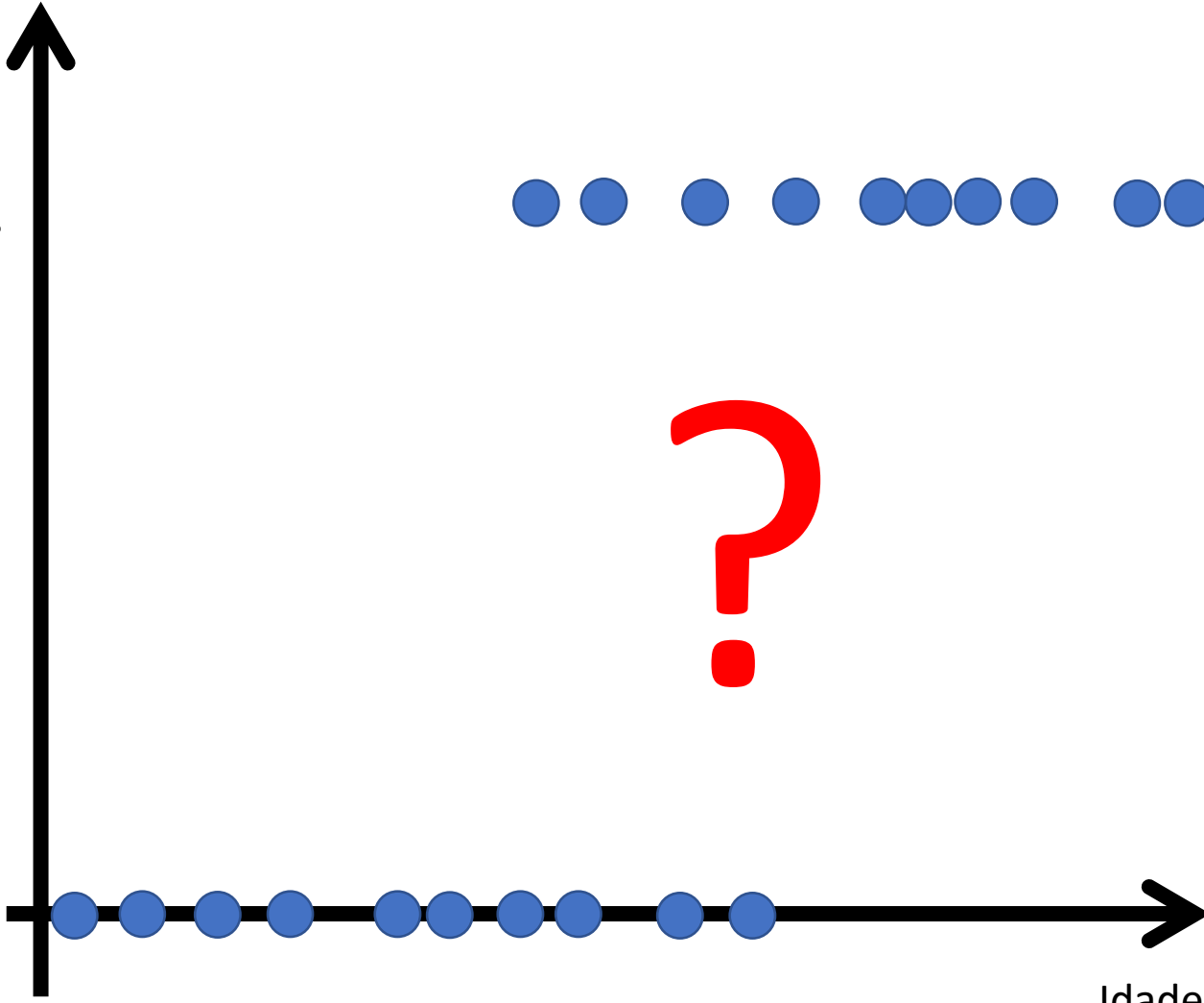
Pagar (S/N)

1

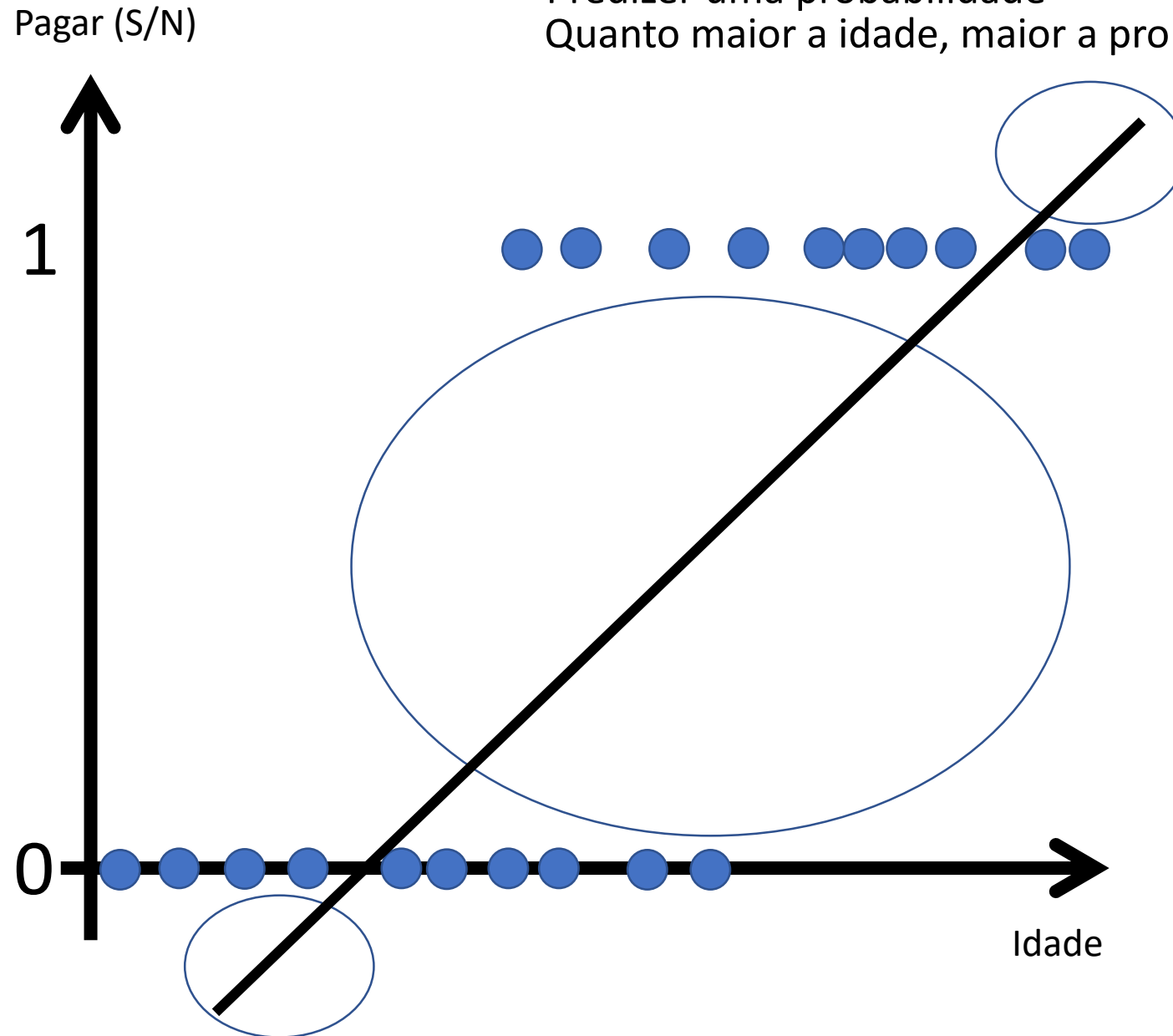
0

Idade

?



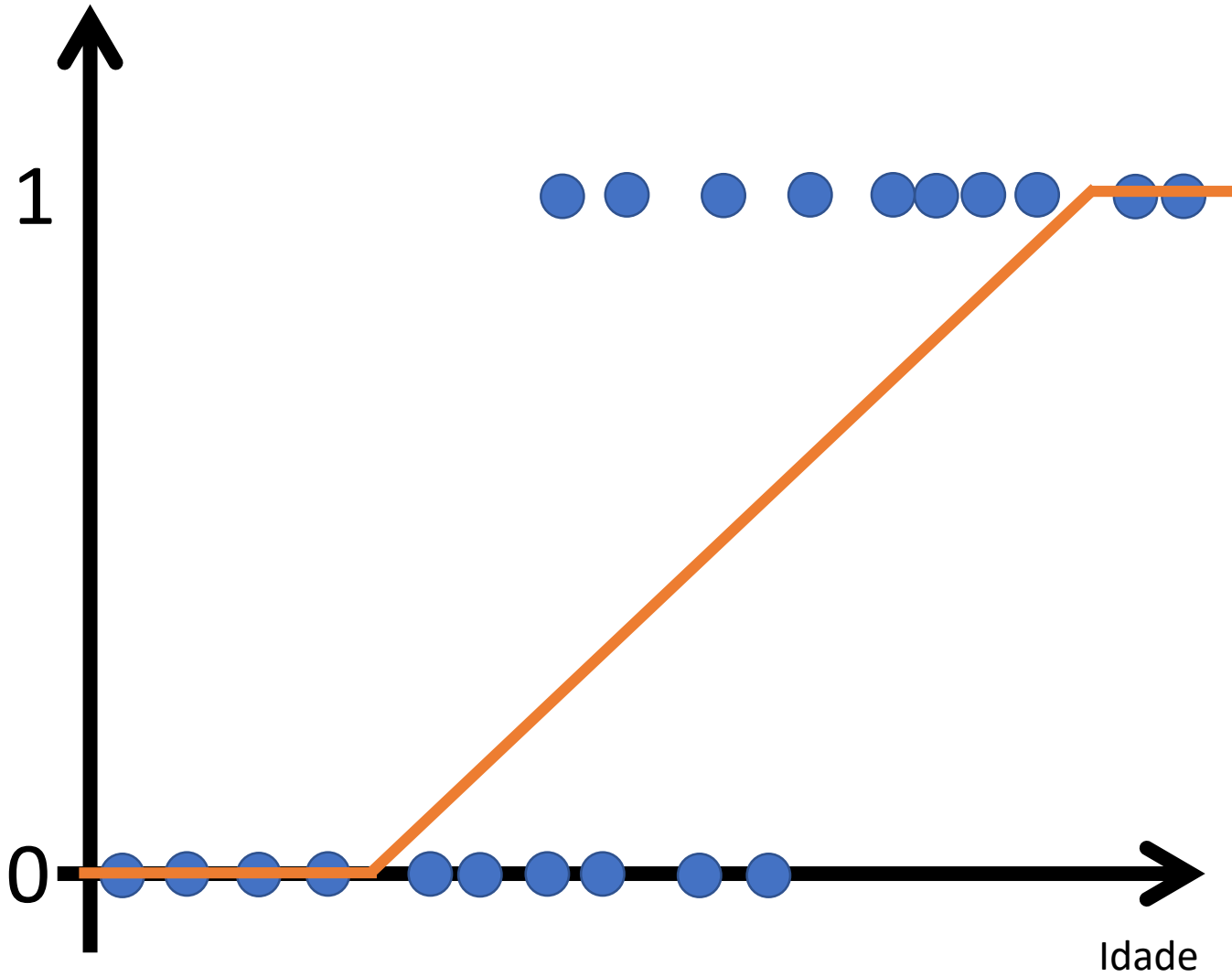
Existe algum tipo de correlação entre a idade e o pagamento?
Predizer uma probabilidade
Quanto maior a idade, maior a probabilidade de pagar



Função sigmoide

Encontrar a melhor linha para encaixar nos dados

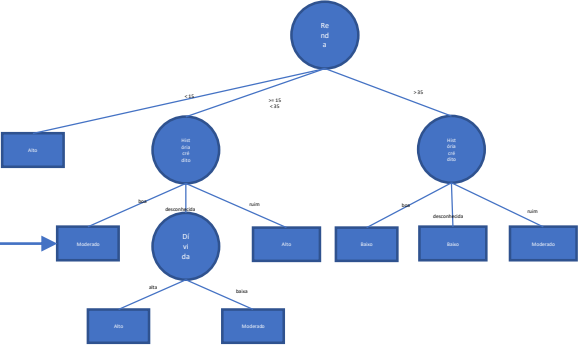
Pagar (S/N)





Estado do cliente	História do crédito			Divida		Garantias		Renda anual		
	Boa 5	Desconhecida 4 5	Ruim 6	Muito 7	Bom 7	Nenhuma 11	Aluguéis 3	< 10000 3	10000 até 20000 6	> 20000 7
Muito boa 1/5	1/5	2/5	3/4	4/7	2/7	6/11	0	3/3	2/4	1/7
Moderada boa 1/5	1/5	1/5	1/4	1/7	2/7	2/11	1/3	0	2/4	1/7
Baixa boa 3/5	3/5	2/5	0	2/7	3/7	3/11	2/3	0	0	5/7

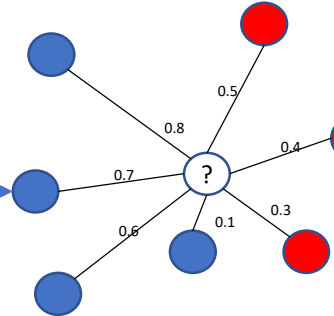
Naive bayes



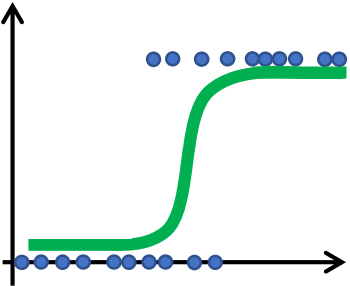
Árvore de decisão

Regra	Resultado
Se renda = >35.000 E história_crédito = BOA	Risco = BAIXO
Se renda = >35.000 e história_crédito = DESCONHECIDA	Risco = BAIXO
Default (padrão)	Risco = ALTO

Regras



kNN



Regressão logística

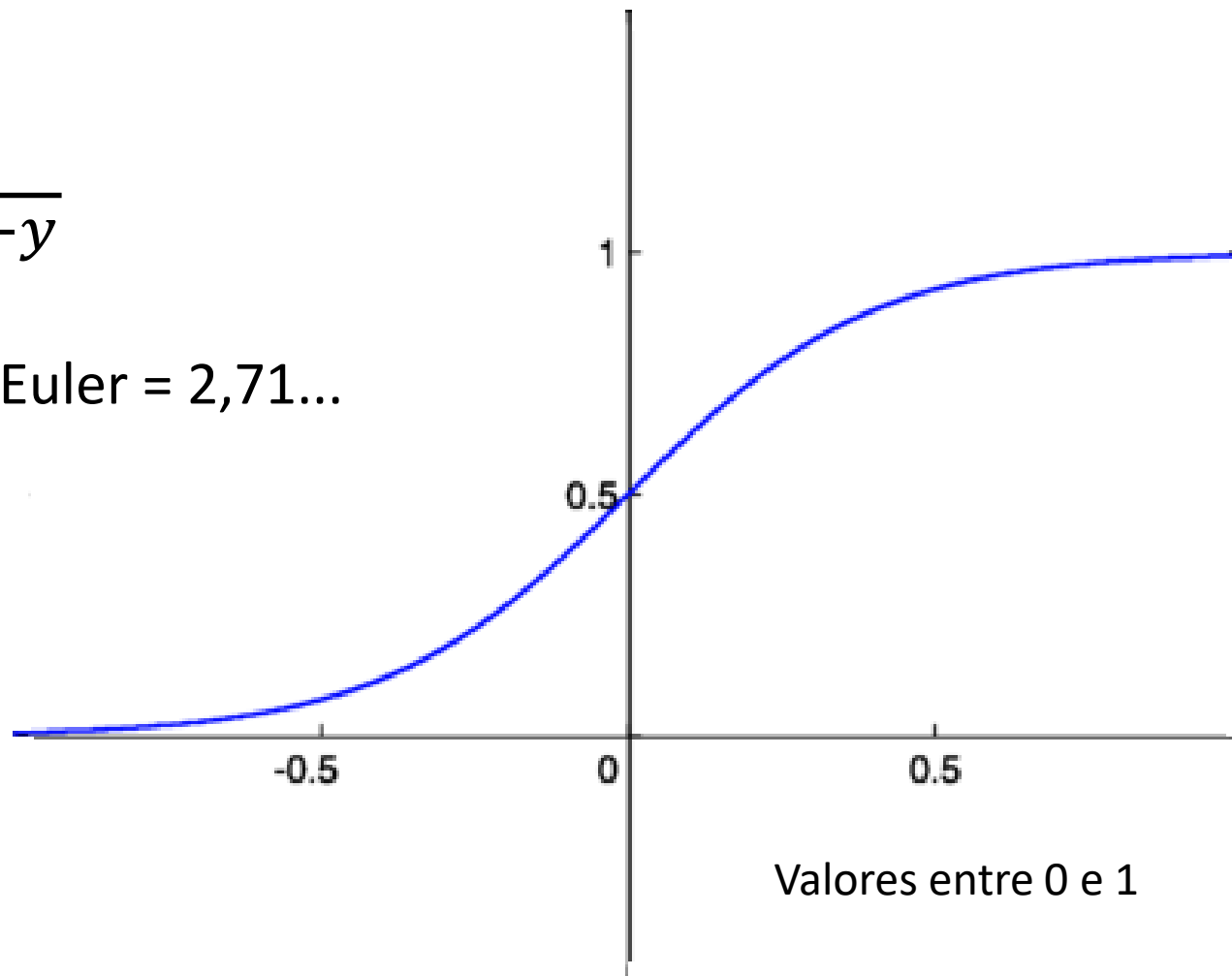
Registros
% acerto

Base teste

Sigmoid (função sigmoide)

$$p = \frac{1}{1 + e^{-y}}$$

e = número de Euler = 2,71...



Valores entre 0 e 1

Se p for alto o valor será aproximadamente 1

Se p for pequeno o valor será aproximadamente 0

Não retorna valores negativos

$$y = b_0 + b_1 * x$$

x = idade

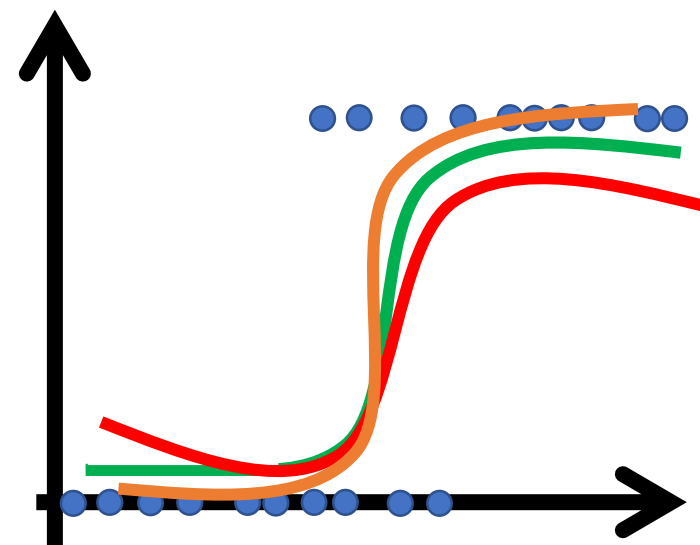
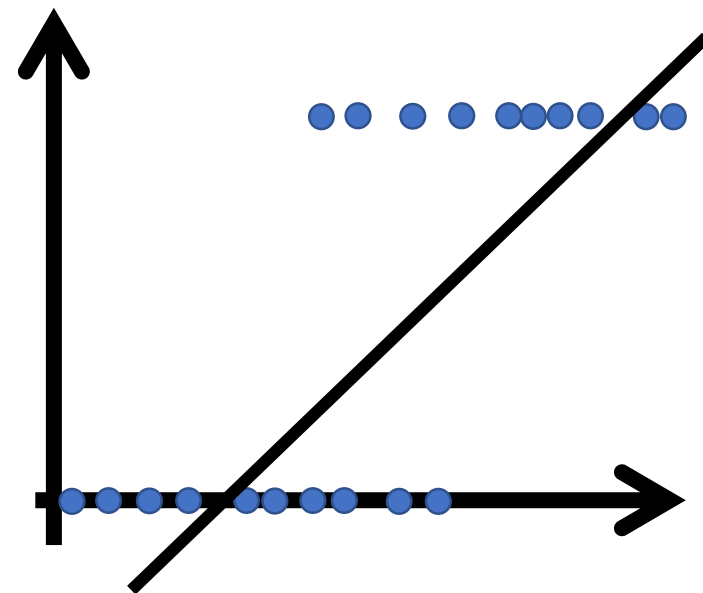
Equação da reta

$$p = \frac{1}{1 + e^{-y}}$$

Função sigmoide

$$\log\left(\frac{p}{1-p}\right) = b_0 + b_1 * x$$

Transformação "logit"

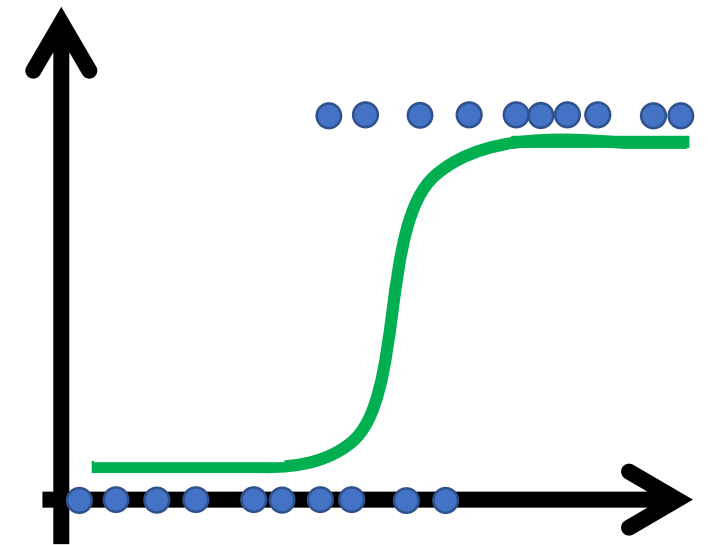


História do crédito	Dívida	Garantias	Renda anual	Risco
Ruim	Alta	Nenhuma	< 15.000	Alto
Desconhecida	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto
Desconhecida	Baixa	Nenhuma	> 35.000	Alto
Desconhecida	Baixa	Nenhuma	> 35.000	Baixo
Desconhecida	Baixa	Adequada	> 35.000	Baixo
Ruim	Baixa	Nenhuma	< 15.000	Alto
Boa	Baixa	Nenhuma	> 35.000	Baixo
Boa	Alta	Adequada	> 35.000	Baixo
Boa	Alta	Nenhuma	< 15.000	Alto
Boa	Alta	Nenhuma	> 35.0000	Baixo
Ruim	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto



História do crédito	Dívida	Garantias	Renda anual	Risco
2	0	0	0	Alto
1	0	0	1	Alto
1	1	0	2	Alto
1	1	0	2	Baixo
1	1	1	2	Baixo
2	1	0	0	Alto
0	1	0	2	Baixo
0	0	1	2	Baixo
0	0	0	0	Alto
0	0	0	2	Baixo
2	0	0	1	Alto

História do crédito	Dívida	Garantias	Renda anual	Risco
2	0	0	0	Alto
1	0	0	1	Alto
1	1	0	2	Alto
1	1	0	2	Baixo
1	1	1	2	Baixo
2	1	0	0	Alto
0	1	0	2	Baixo
0	0	1	2	Baixo
0	0	0	0	Alto
0	0	0	2	Baixo
2	0	0	1	Alto



Encontrar os parâmetros $B_0, B_1, B_2 \dots B_N$

$$B_0 = -0,12$$

$$B_1 (\text{história do crédito}) = -0,71$$

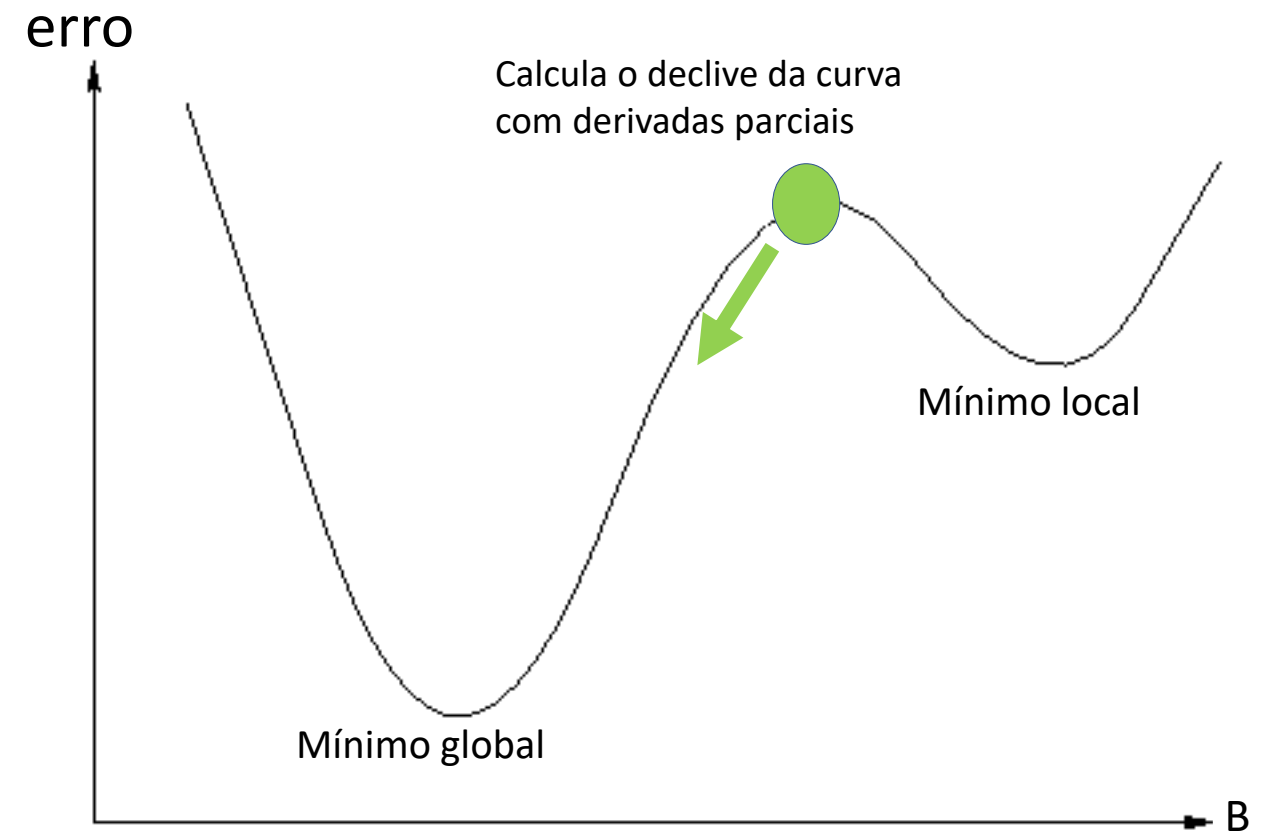
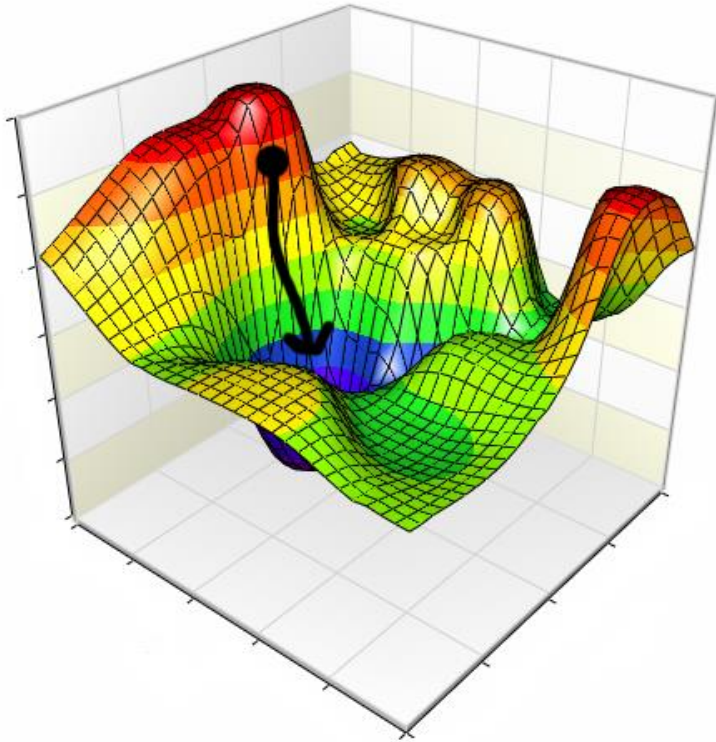
$$B_2 (\text{dívida}) = 0,24$$

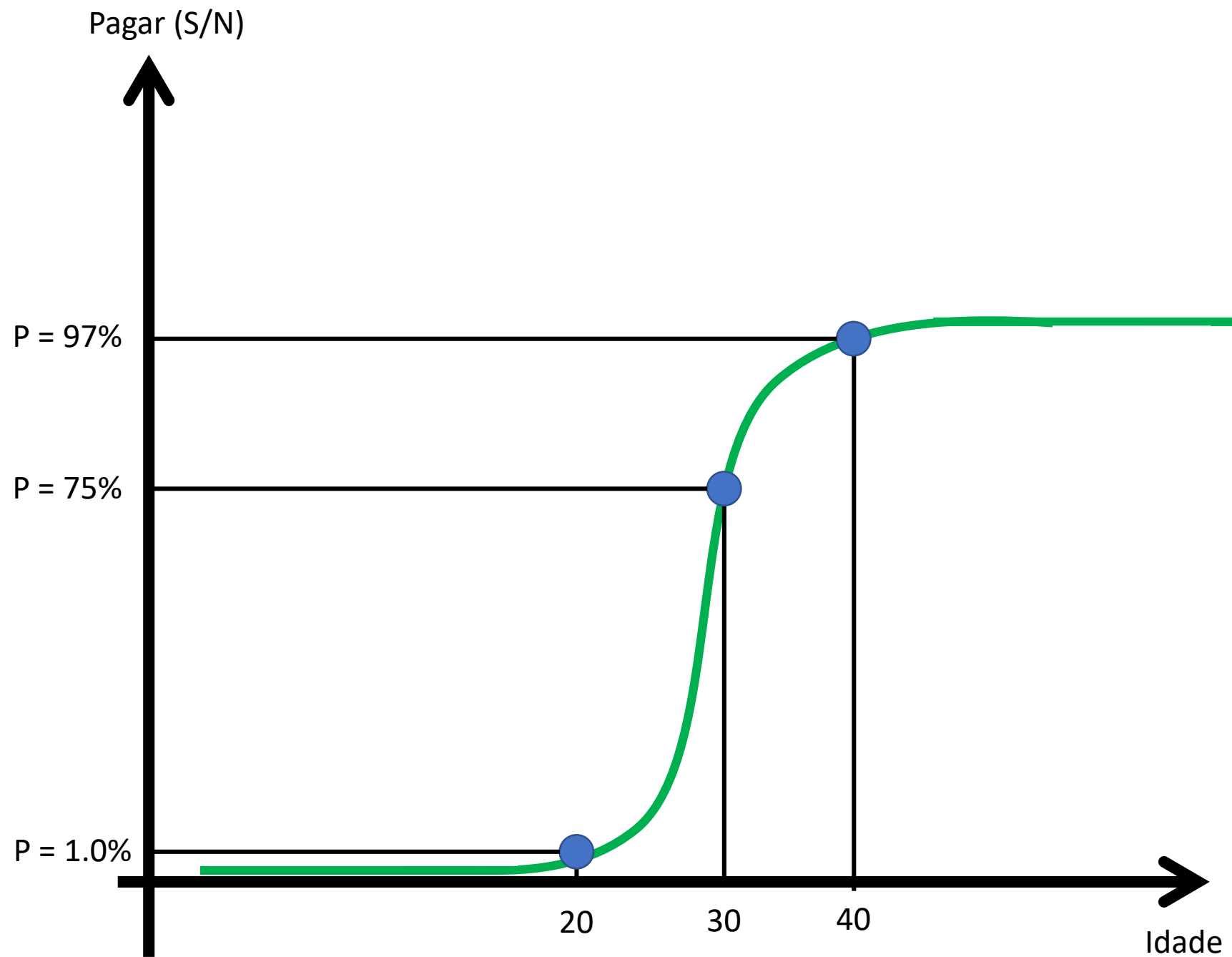
$$B_3 (\text{garantias}) = -0,54$$

$$B_4 (\text{renda anual}) = 1.07$$

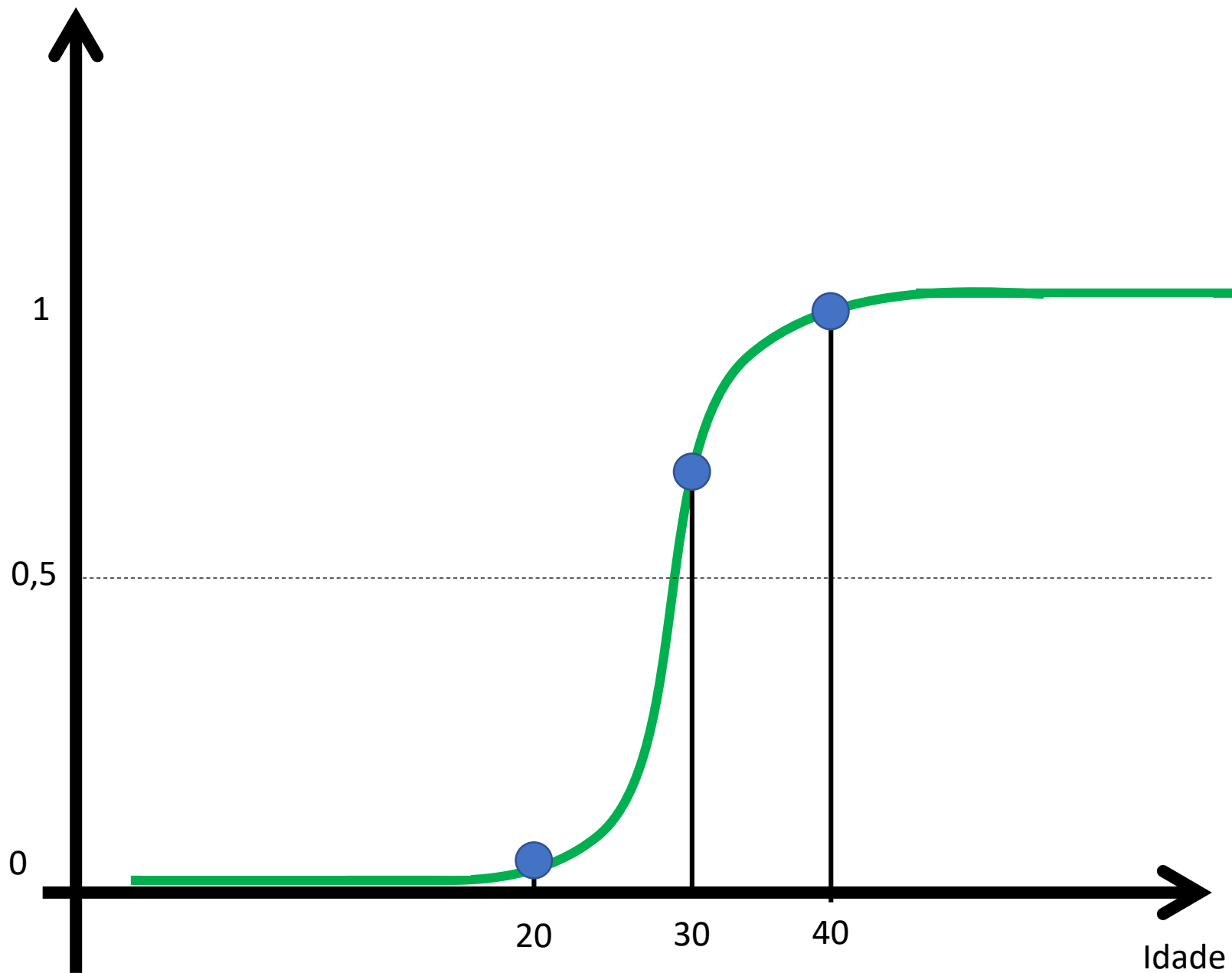
Descida do gradiente

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





Pagar (S/N)



$$y = b_0 + b_1 * x$$

Equação da reta

$$p = \frac{1}{1 + e^{-y}}$$

Função sigmoide

$$B_0 = -0,12$$

$$B_1 \text{ (história do crédito)} = -0,71$$

$$B_2 \text{ (dívida)} = 0,24$$

$$B_3 \text{ (garantias)} = -0,54$$

$$B_4 \text{ (renda anual)} = 1,07$$

História = Boa (0)

Dívida = Alta (0)

Garantias = Nenhuma (1)

Renda = > 35 (2)

$$y = b_0 + b_1 * \text{história} + b_2 * \text{dívida} + b_3 * \text{garantias} + b_4 * \text{renda}$$

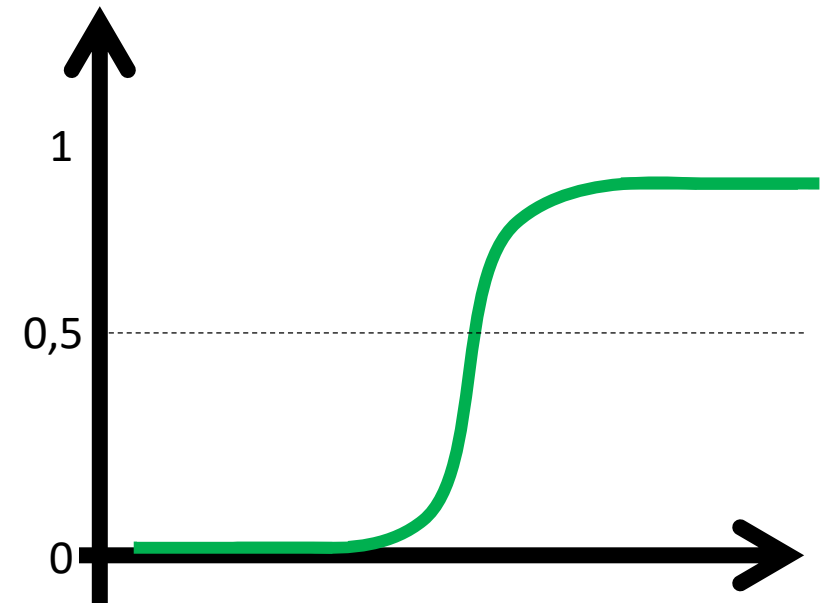
$$y = -0,12 + (-0,71 * 0) + 0,24 * 0 + (-0,54 * 1) + 1,07 * 2$$

$$y = 1,48$$

$$p = \frac{1}{1 + e^{-(1,48)}}$$

$$p = 0,81$$

Risco alto = 0
Risco baixo = 1



$$y = b_0 + b_1 * x$$

Equação da reta

$$p = \frac{1}{1 + e^{-y}}$$

Função sigmoide

$$B_0 = -0,12$$

$$B_1 \text{ (história do crédito)} = -0,71$$

$$B_2 \text{ (dívida)} = 0,24$$

$$B_3 \text{ (garantias)} = -0,54$$

$$B_4 \text{ (renda anual)} = 1,07$$

História = Ruim (3)

Dívida = Alta (0)

Garantias = Adequada (0)

Renda < 15 (0)

$$y = b_0 + b_1 * \text{história} + b_2 * \text{dívida} + b_3 * \text{garantias} + b_4 * \text{renda}$$

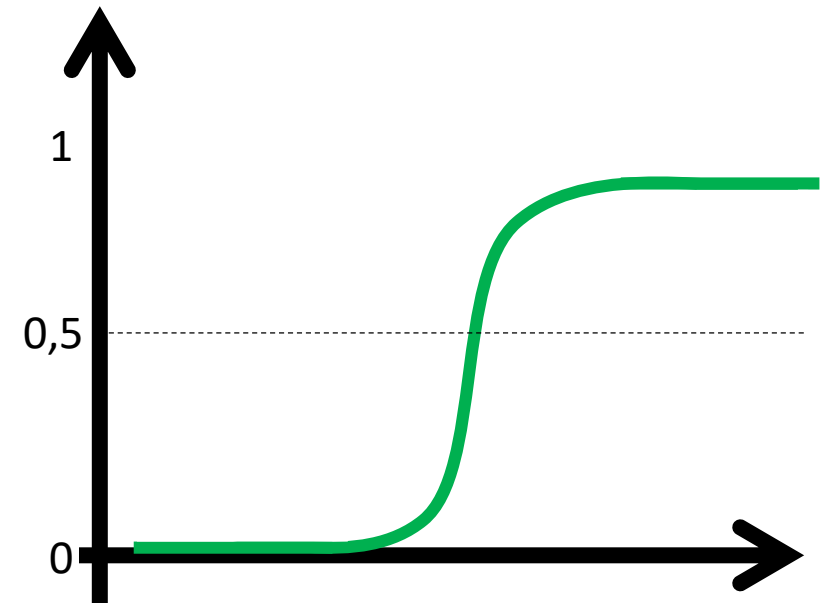
$$y = -0,12 + (-0,71 * 3) + 0,24 * 0 + (-0,54 * 0) + 1,07 * 0$$

$$y = -2,25$$

$$p = \frac{1}{1 + e^{-(-2,25)}}$$

$$p = 0,09$$

Risco alto = 0
Risco baixo = 1



Conclusão

