

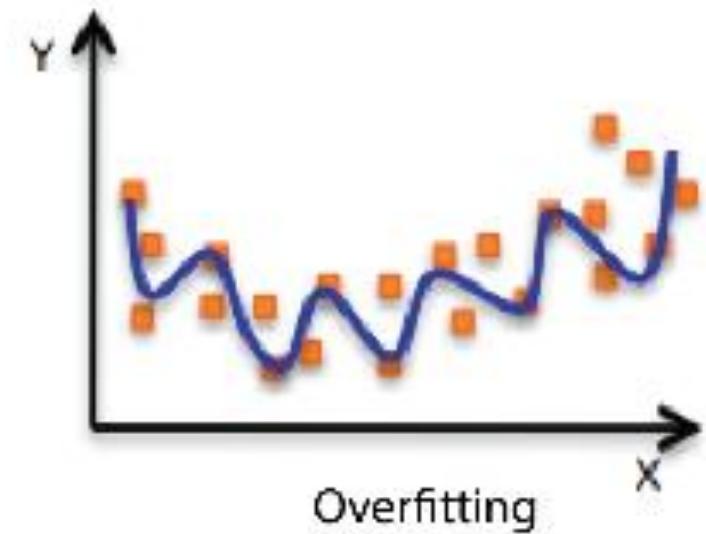
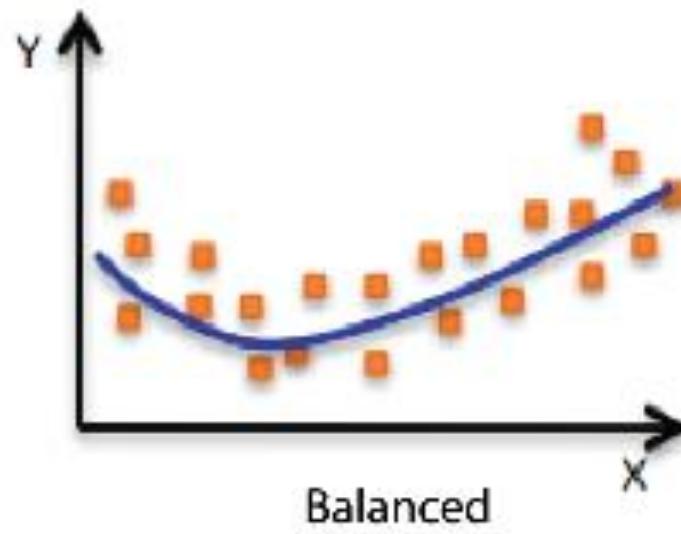
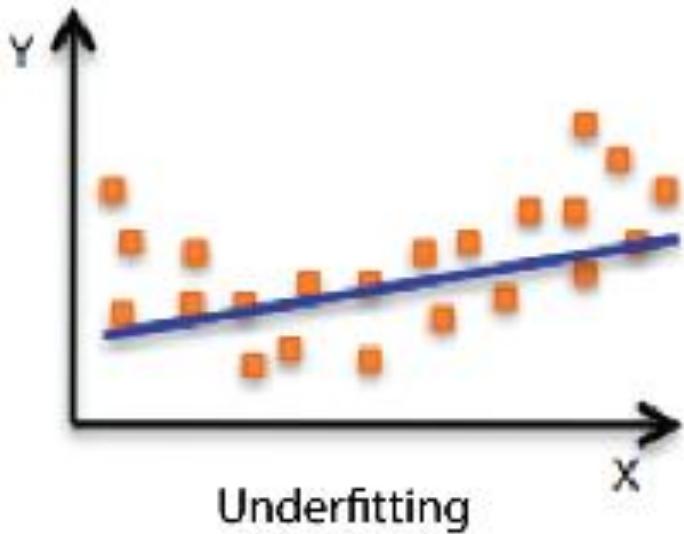


escola
britânica de
artes criativas
& tecnologia

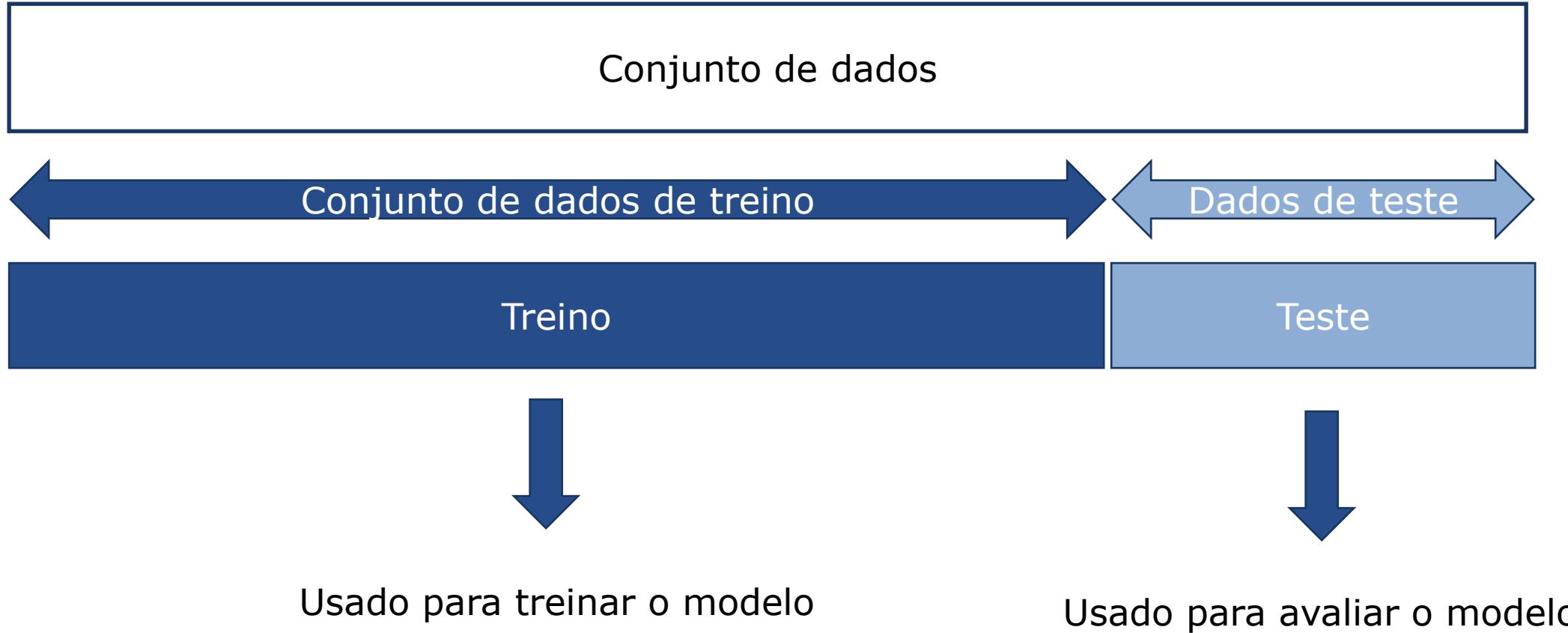
Profissão Cientista de Dados
Árvore de decisão

Overfitting, treino e teste

Overfitting



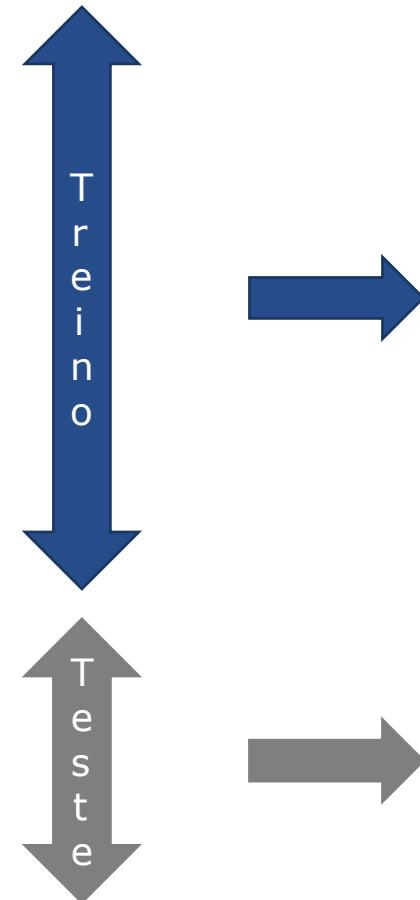
Treino e teste



Treino e teste

Conjunto de dados para treinar o modelo

	x1	x2	y
0	10	100	Não
1	22	200	Não
2	13	300	Não
3	42	400	Sim
4	25	500	Não
5	36	600	Sim
6	24	400	Não
7	15	500	Não
8	26	600	Sim
9	10	1000	Sim



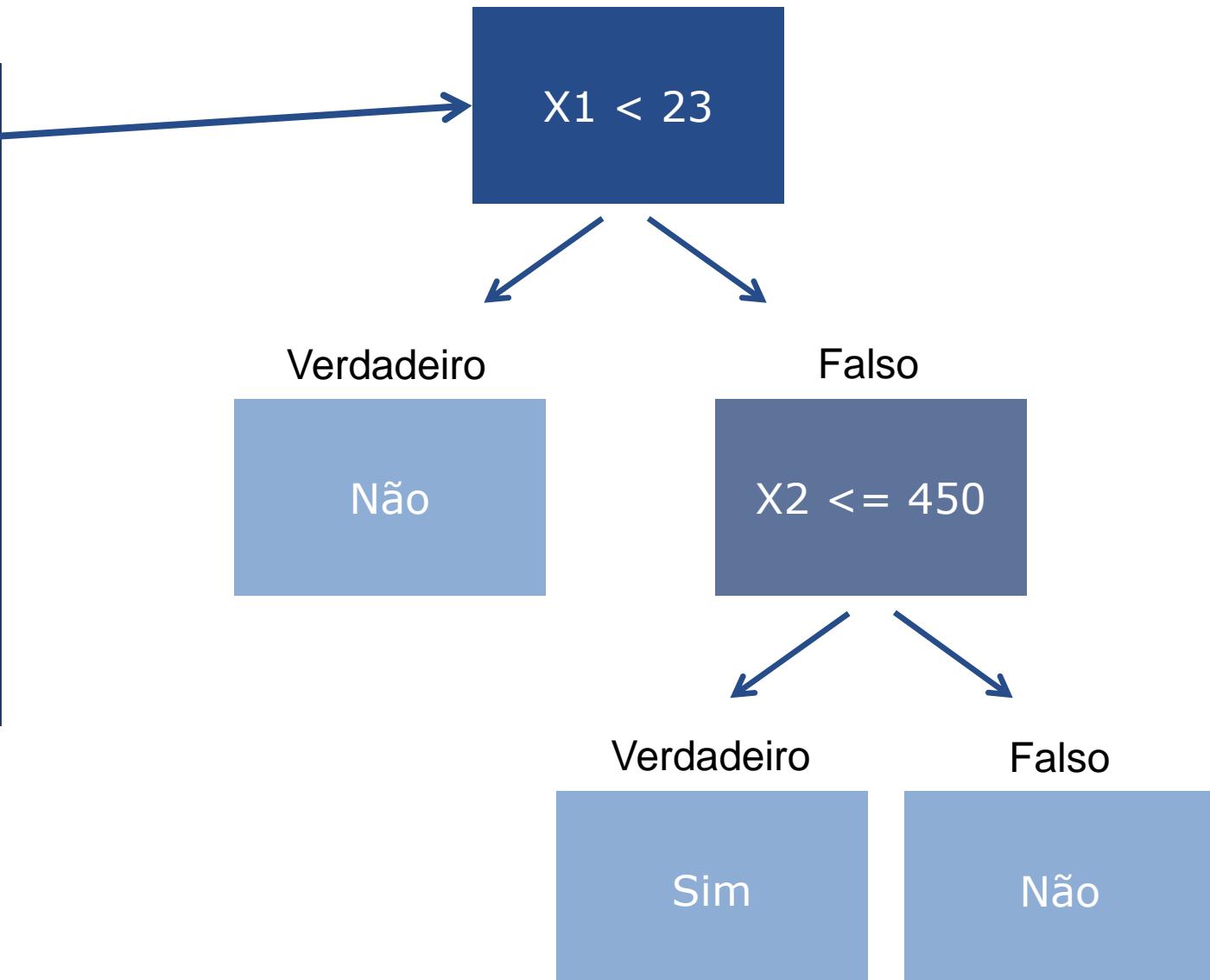
	x1	x2	y
0	10	100	Não
1	22	200	Não
2	13	300	Não
3	42	400	Sim
4	25	500	Não
5	36	600	Sim
6	24	400	Não

Conjunto de dados para avaliar o modelo

	x1	x2	y
7	15	500	Não
8	26	600	Sim
9	10	1000	Sim

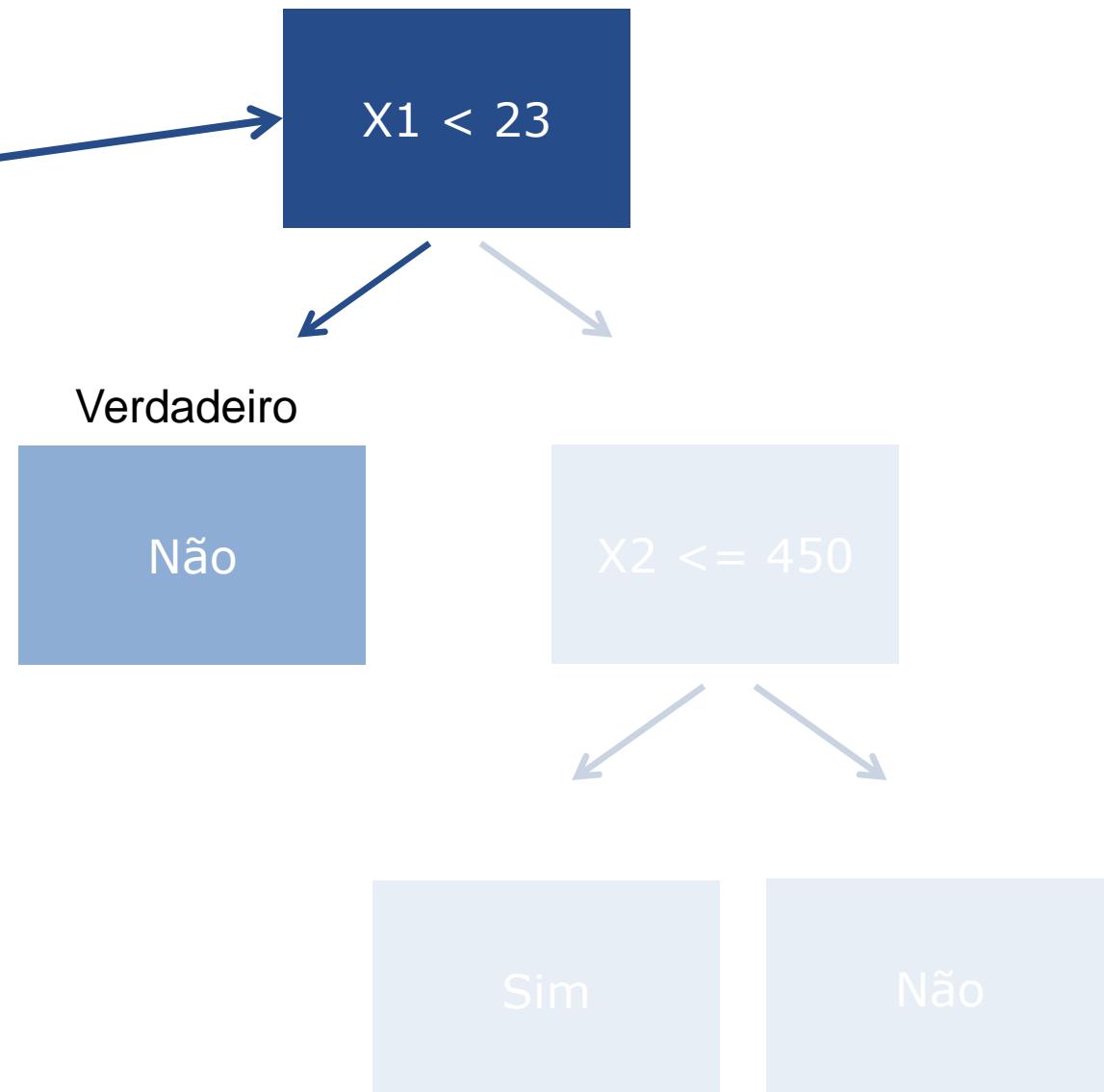
Treinando o modelo

	x1	x2	y
0	10	100	Não
1	22	200	Não
2	13	300	Não
3	42	400	Sim
4	25	500	Não
5	36	600	Sim
6	24	400	Sim



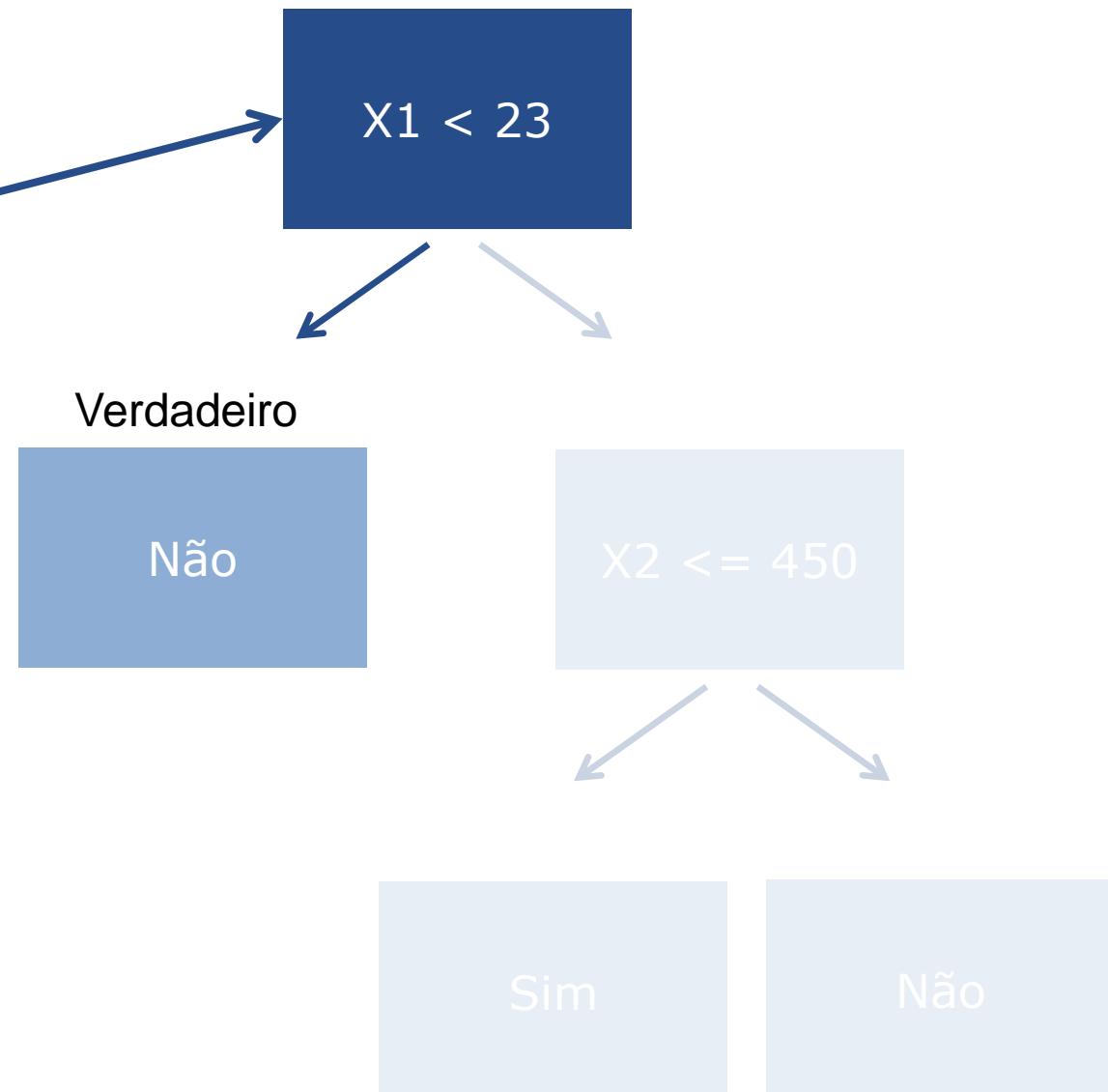
Treinando o modelo

	x1	x2	y	\hat{y}
0	10	100	Não	Não
1	22	200	Não	
2	13	300	Não	
3	42	400	Sim	
4	25	500	Não	
5	36	600	Sim	
6	24	400	Sim	



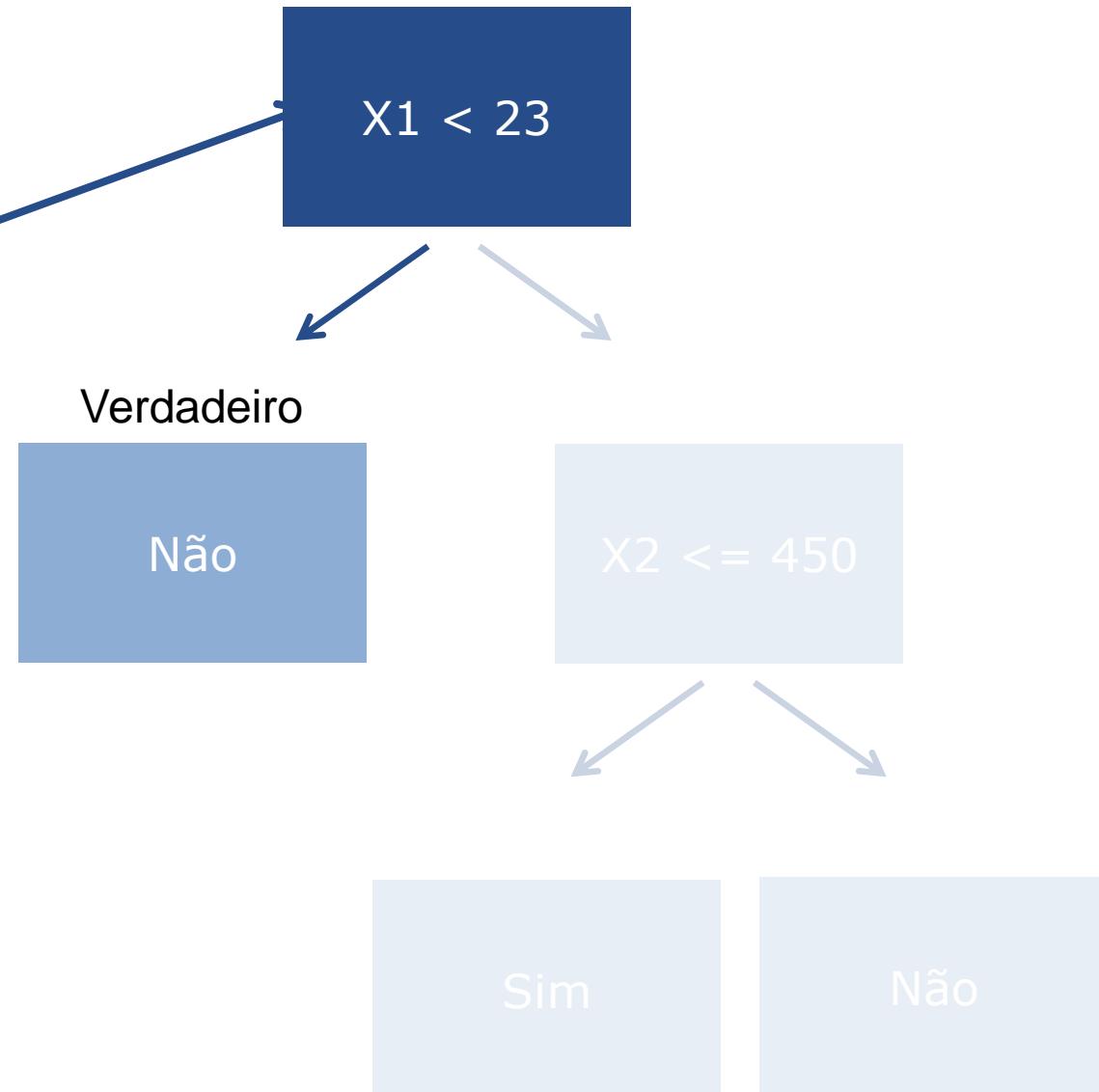
Treinando o modelo

	x1	x2	y	\hat{y}
0	10	100	Não	Não
1	22	200	Não	Não
2	13	300	Não	
3	42	400	Sim	
4	25	500	Não	
5	36	600	Sim	
6	24	400	Sim	



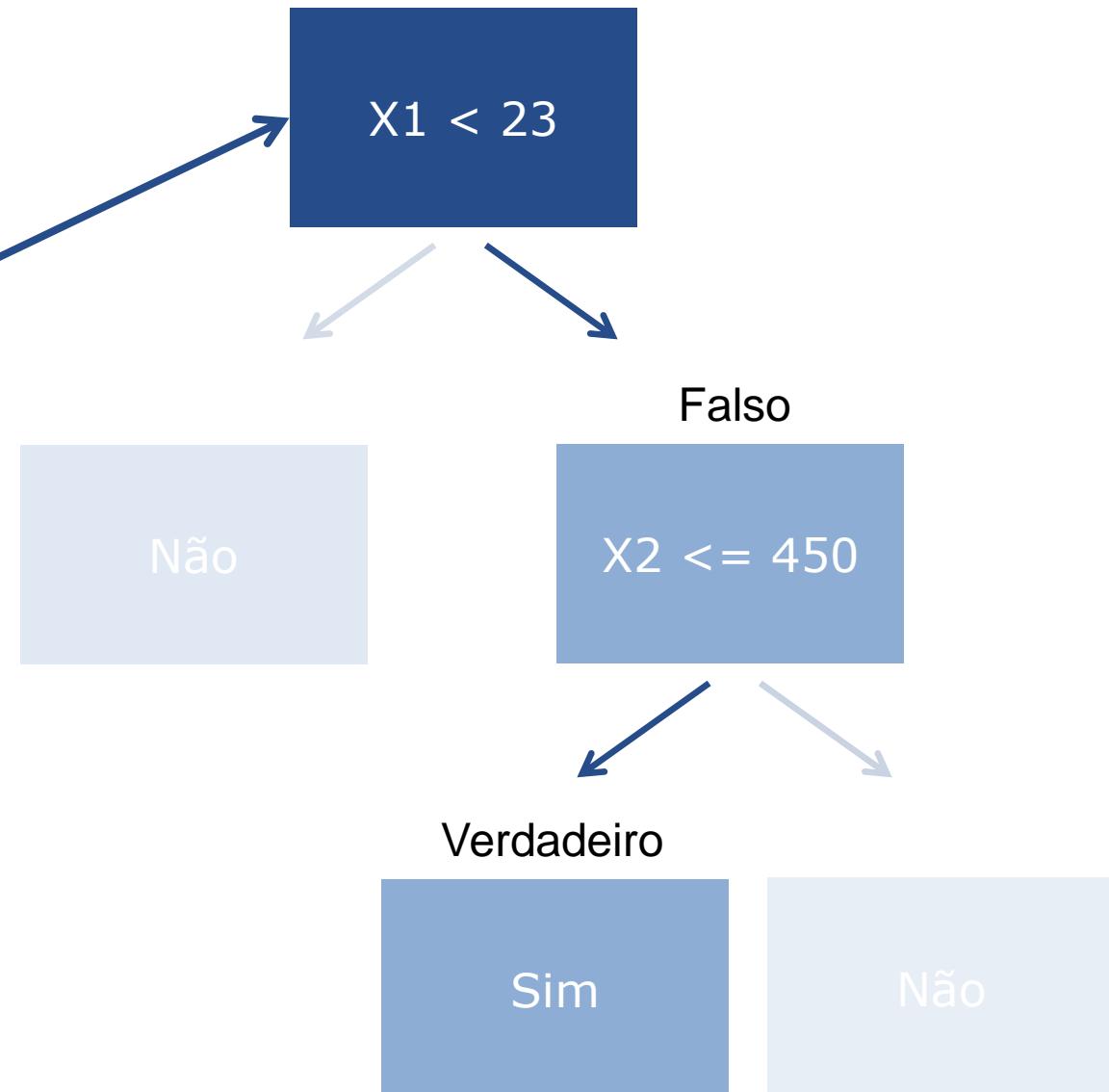
Treinando o modelo

	x1	x2	y	\hat{y}
0	10	100	Não	Não
1	22	200	Não	Não
2	13	300	Não	Não
3	42	400	Sim	
4	25	500	Não	
5	36	600	Sim	
6	24	400	Sim	



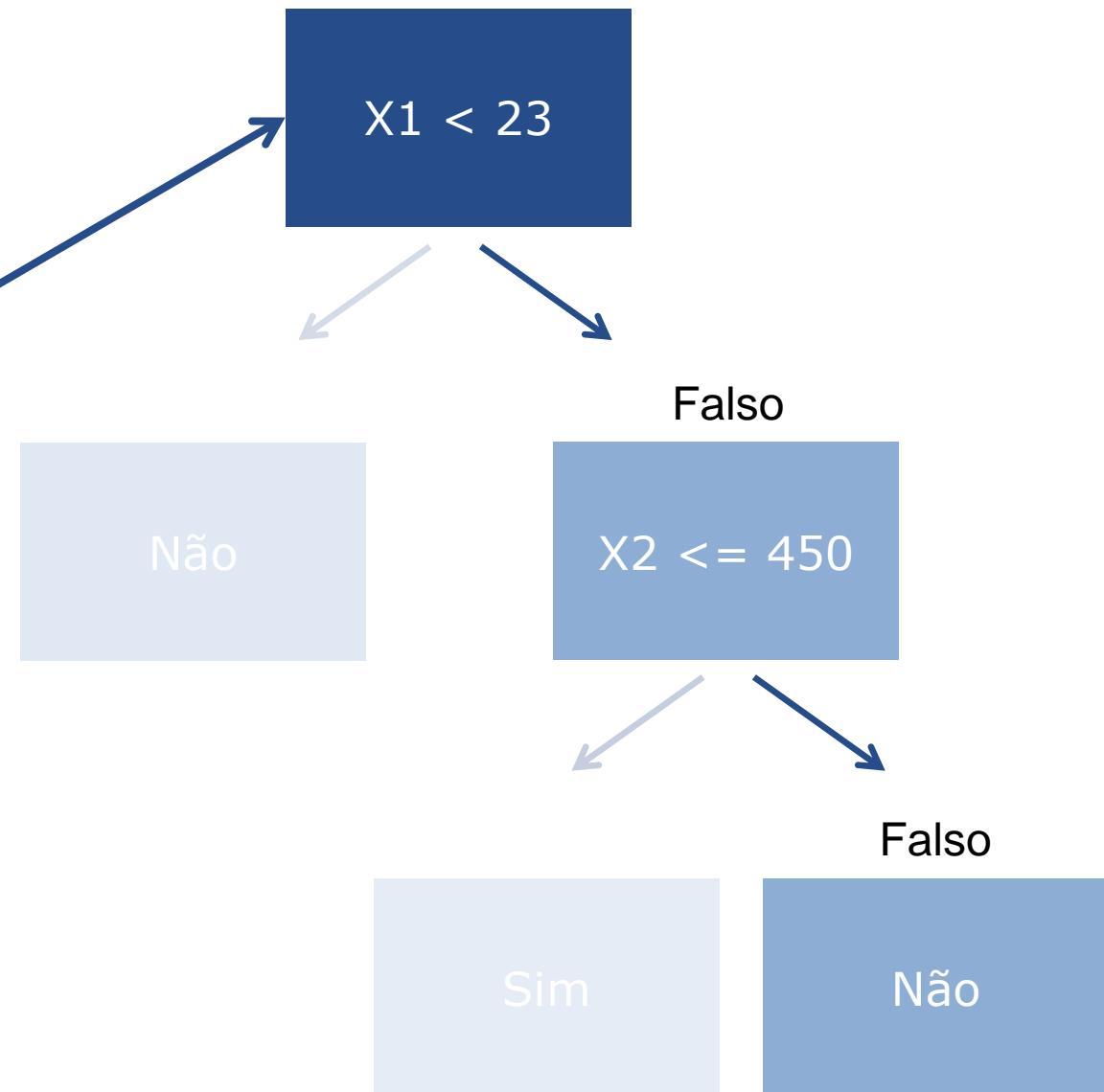
Treinando o modelo

	x1	x2	y	\hat{y}
0	10	100	Não	Não
1	22	200	Não	Não
2	13	300	Não	Não
3	42	400	Sim	Sim
4	25	500	Não	
5	36	600	Sim	
6	24	400	Sim	



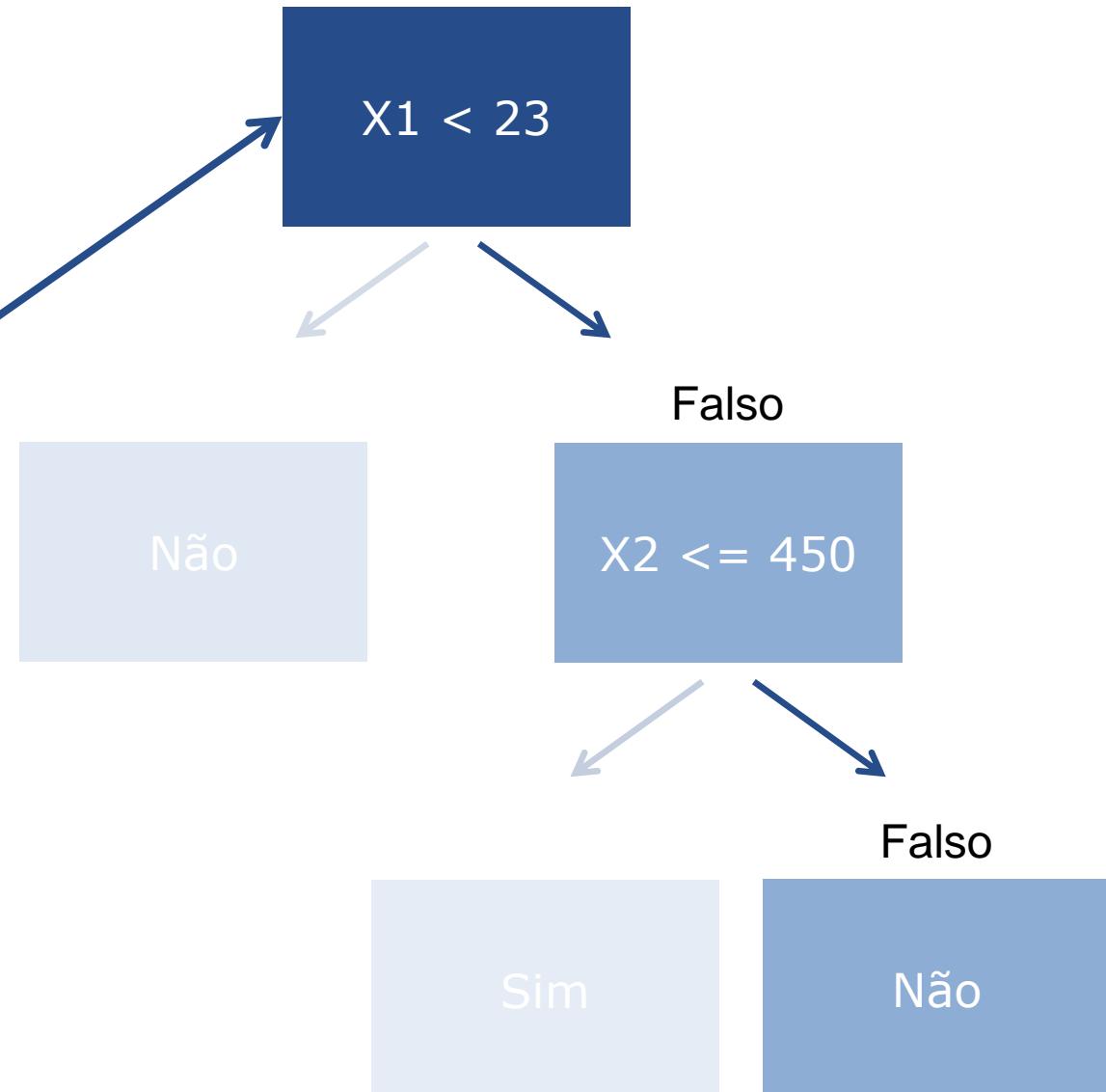
Treinando o modelo

	x1	x2	y	\hat{y}
0	10	100	Não	Não
1	22	200	Não	Não
2	13	300	Não	Não
3	42	400	Sim	Sim
4	25	500	Não	Não
5	36	600	Sim	
6	24	400	Sim	



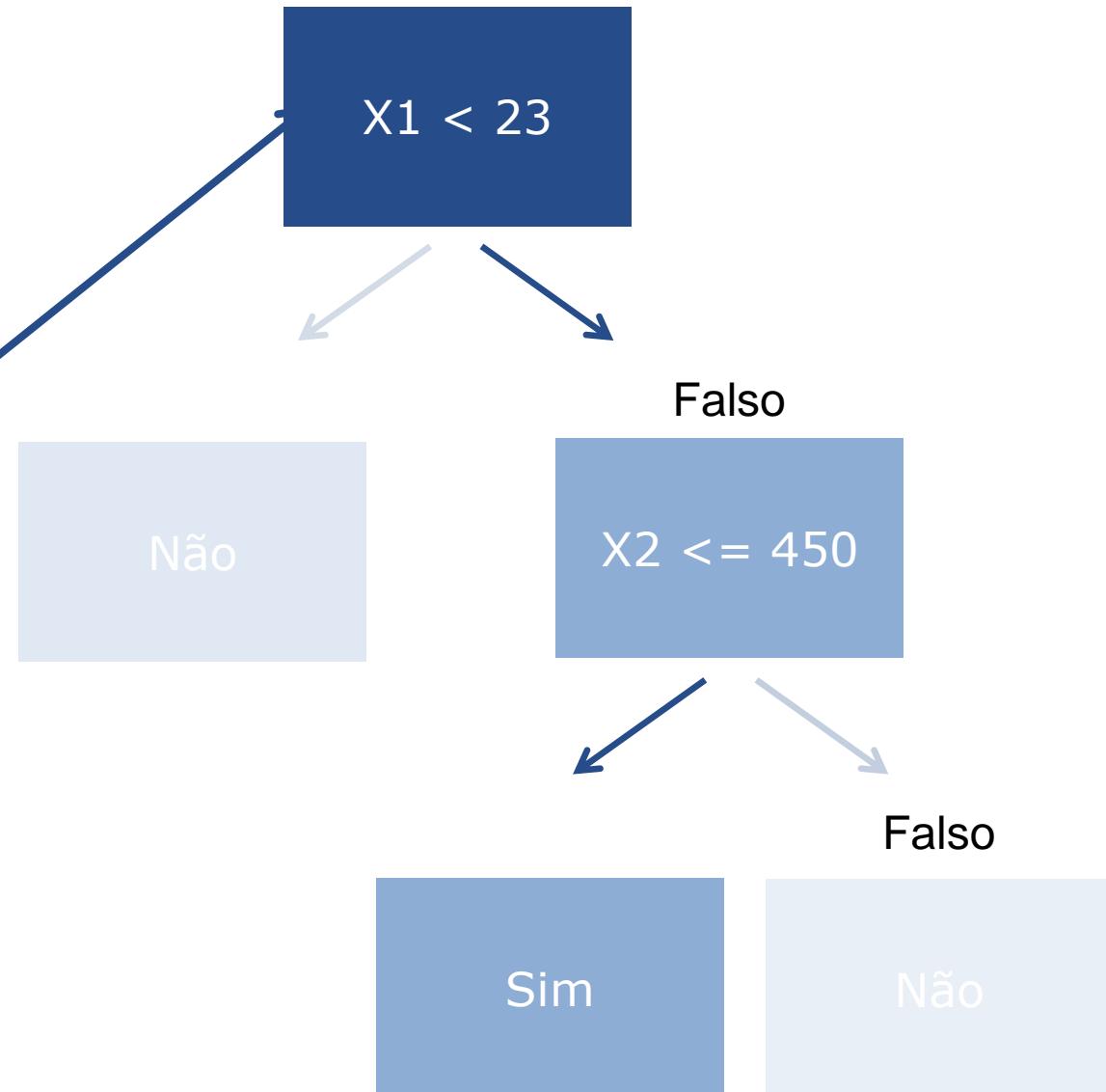
Treinando o modelo

	x1	x2	y	\hat{y}
0	10	100	Não	Não
1	22	200	Não	Não
2	13	300	Não	Não
3	42	400	Sim	Sim
4	25	500	Não	Não
5	36	600	Sim	Não
6	24	400	Sim	



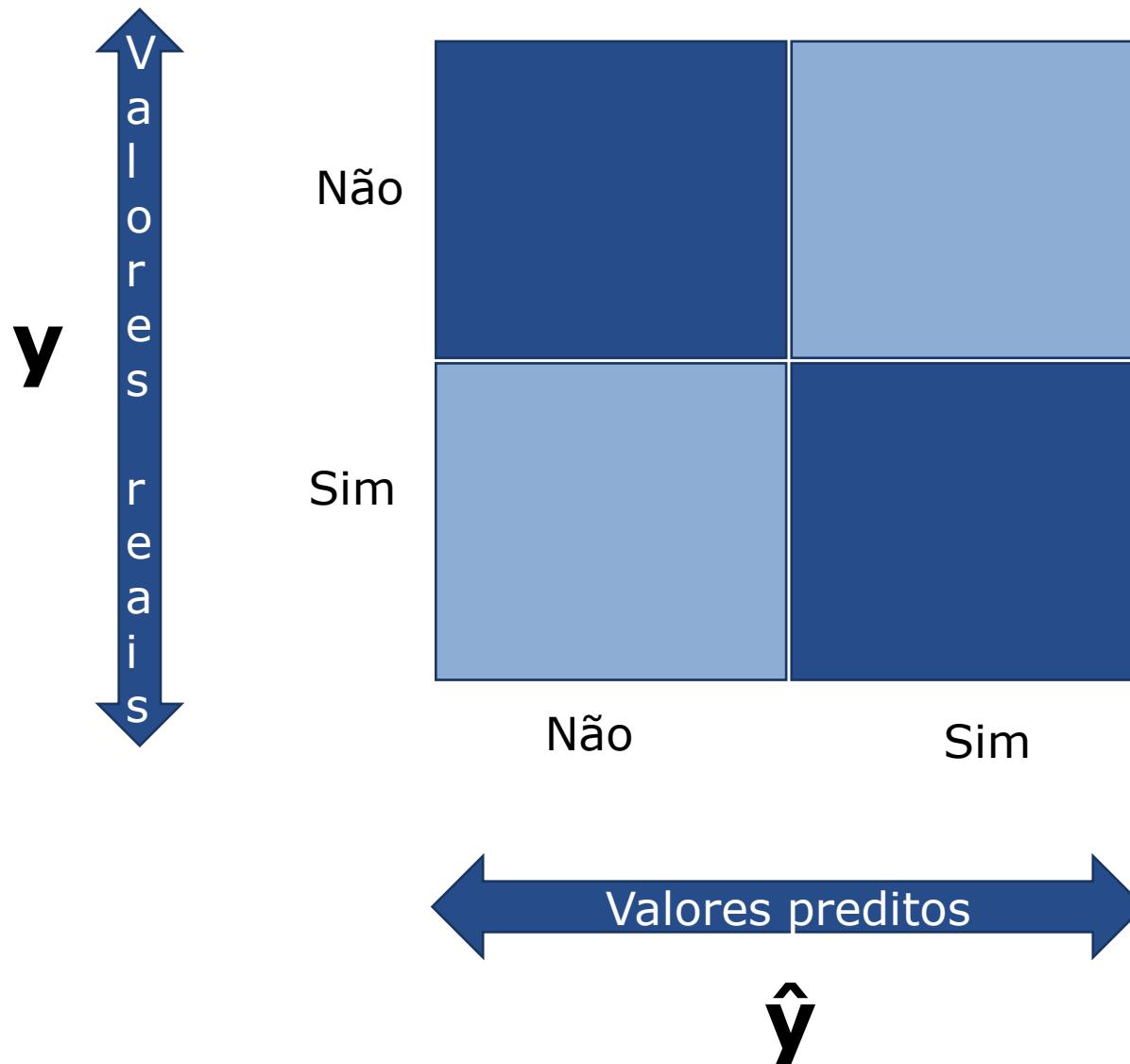
Treinando o modelo

	x1	x2	y	\hat{y}
0	10	100	Não	Não
1	22	200	Não	Não
2	13	300	Não	Não
3	42	400	Sim	Sim
4	25	500	Não	Não
5	36	600	Sim	Não
6	24	400	Sim	Sim



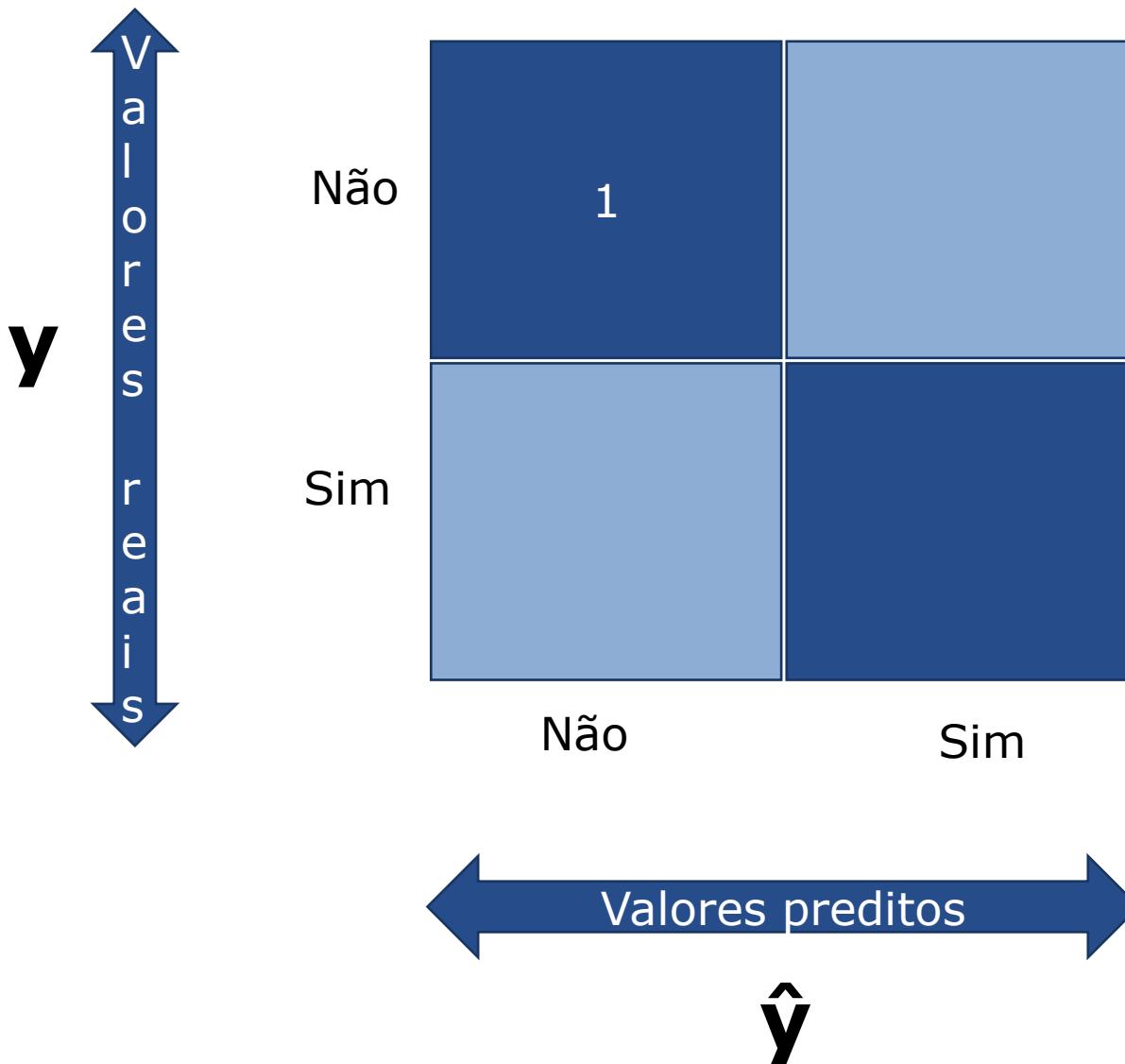
Matriz de confusão treino

	y	\hat{y}
0	Não	Não
1	Não	Não
2	Não	Não
3	Sim	Sim
4	Não	Não
5	Sim	Não
6	Sim	Sim



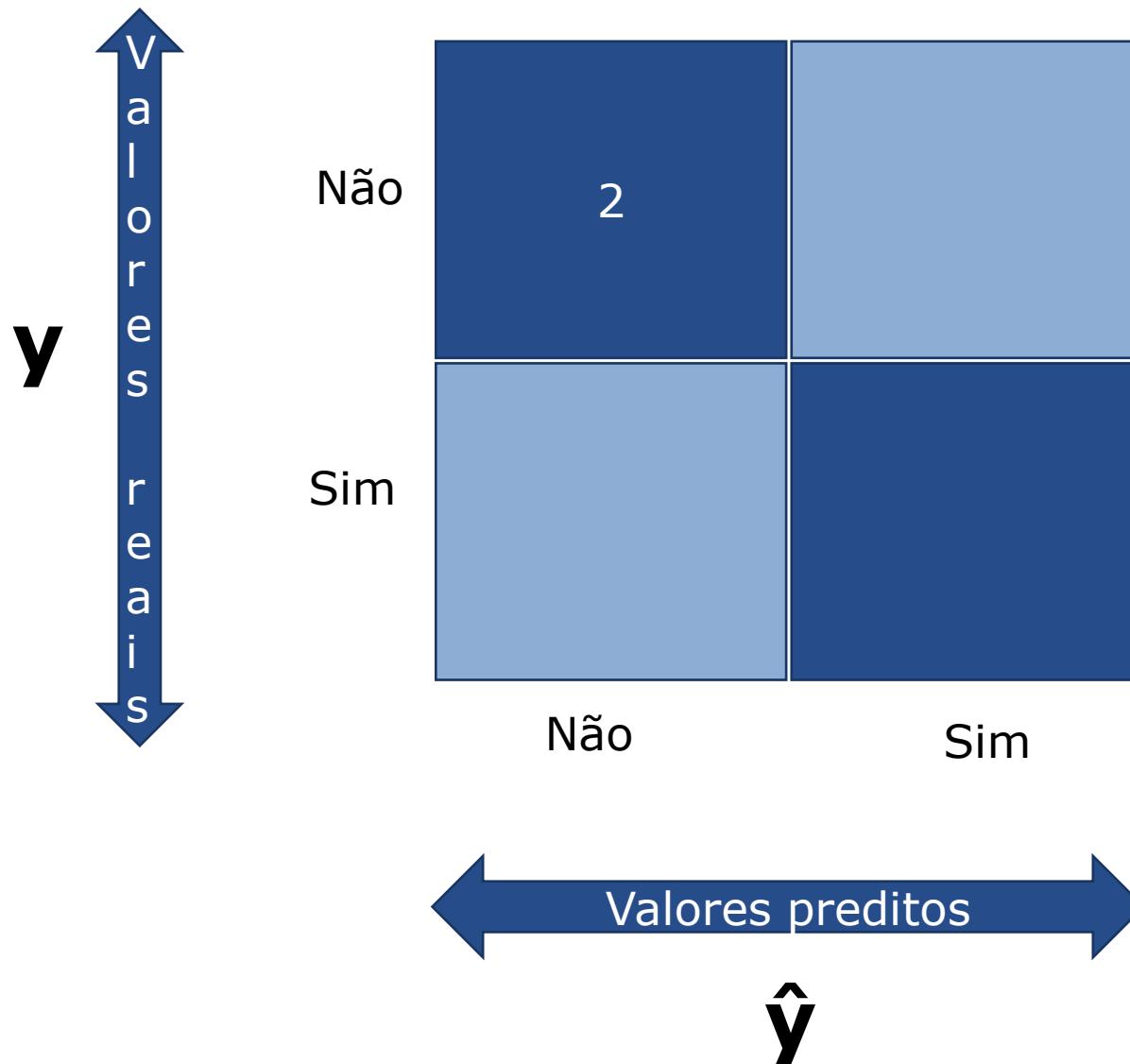
Matriz de confusão treino

	y	\hat{y}
0	Não	Não
1	Não	Não
2	Não	Não
3	Sim	Sim
4	Não	Não
5	Sim	Não
6	Sim	Sim



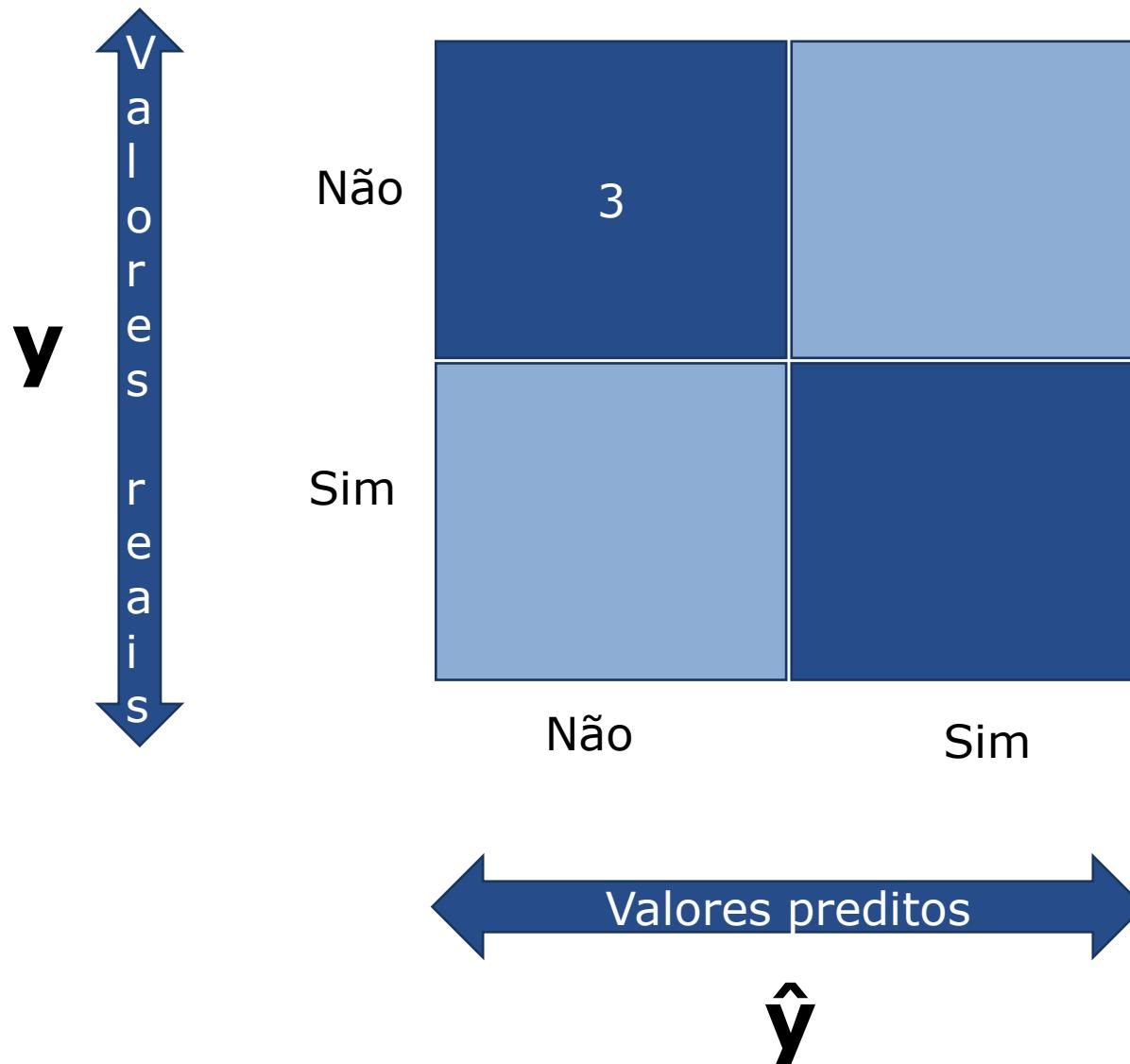
Matriz de confusão treino

	y	\hat{y}
0	Não	Não
1	Não	Não
2	Não	Não
3	Sim	Sim
4	Não	Não
5	Sim	Não
6	Sim	Sim



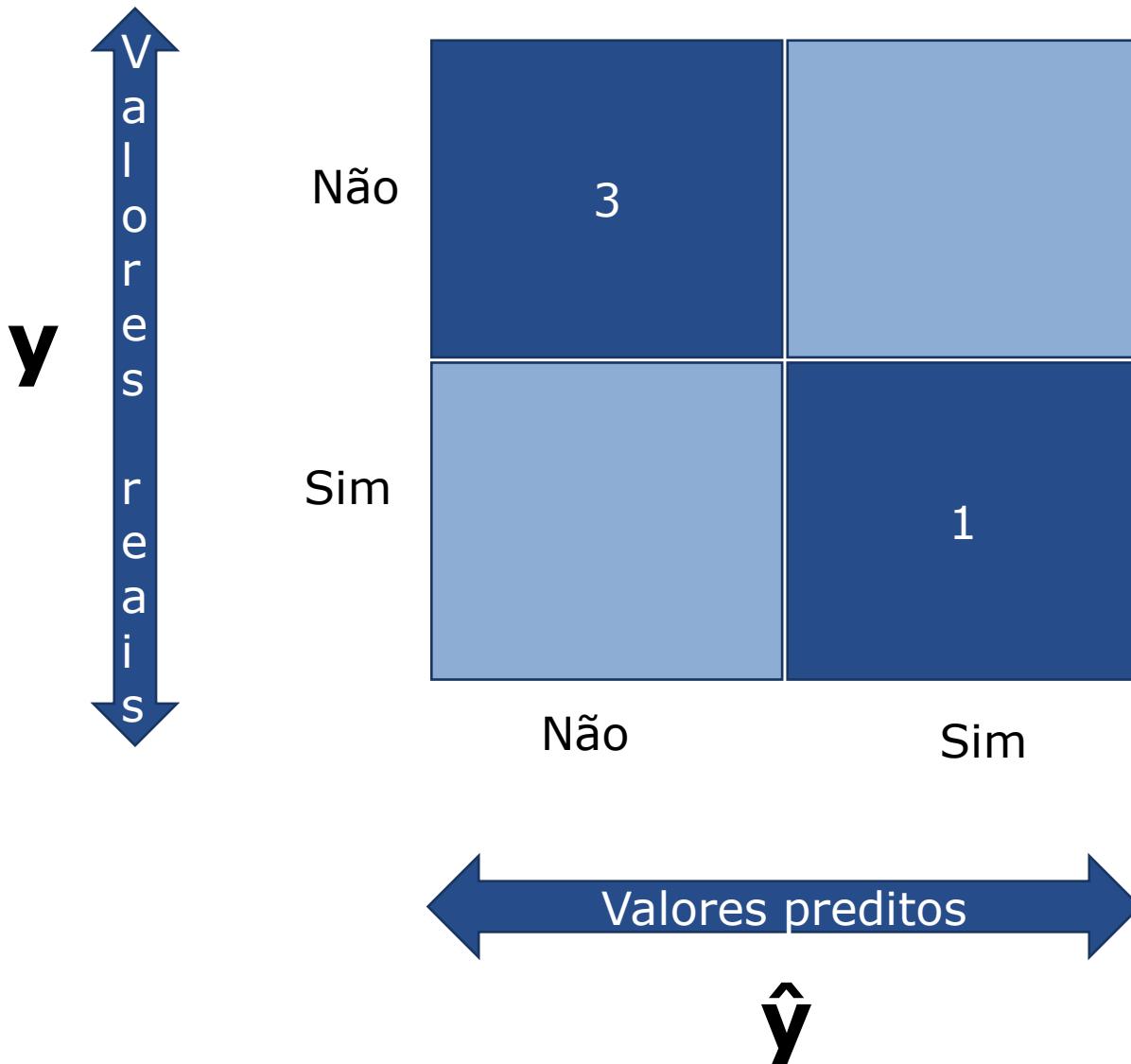
Matriz de confusão treino

	y	\hat{y}
0	Não	Não
1	Não	Não
2	Não	Não
3	Sim	Sim
4	Não	Não
5	Sim	Não
6	Sim	Sim



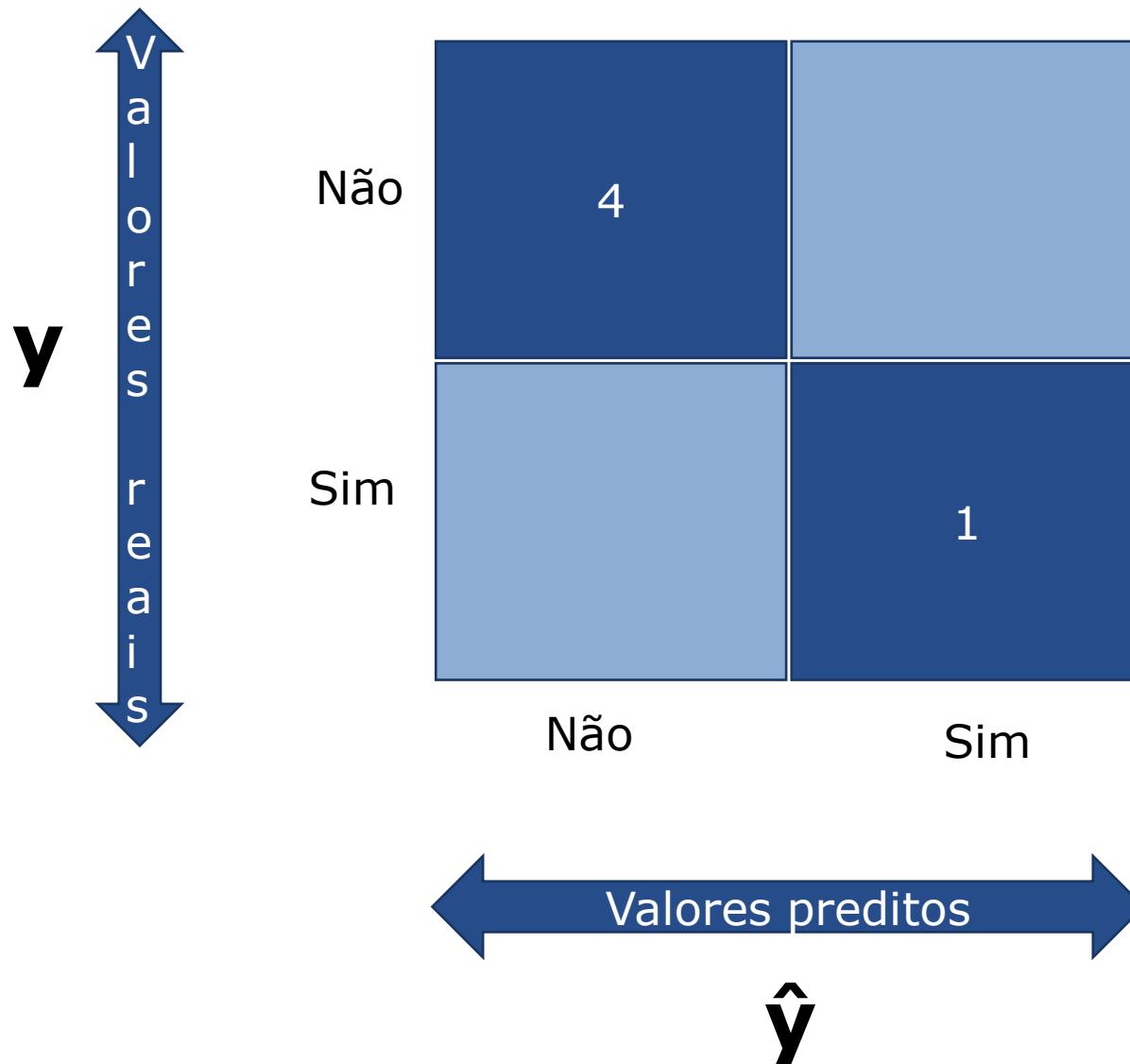
Matriz de confusão treino

	y	\hat{y}
0	Não	Não
1	Não	Não
2	Não	Não
3	Sim	Sim
4	Não	Não
5	Sim	Não
6	Sim	Sim



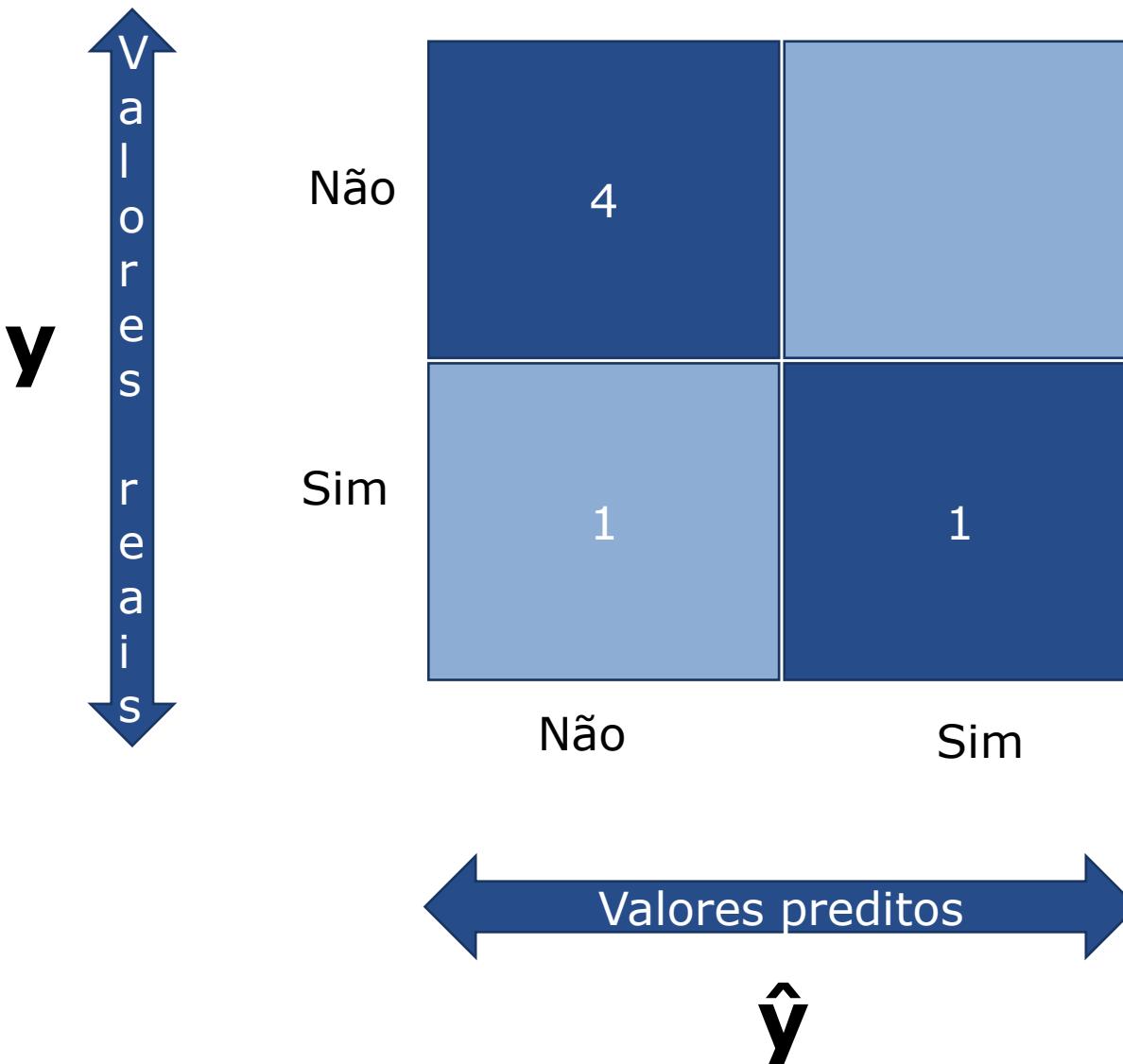
Matriz de confusão treino

	y	\hat{y}
0	Não	Não
1	Não	Não
2	Não	Não
3	Sim	Sim
4	Não	Não
5	Sim	Não
6	Sim	Sim



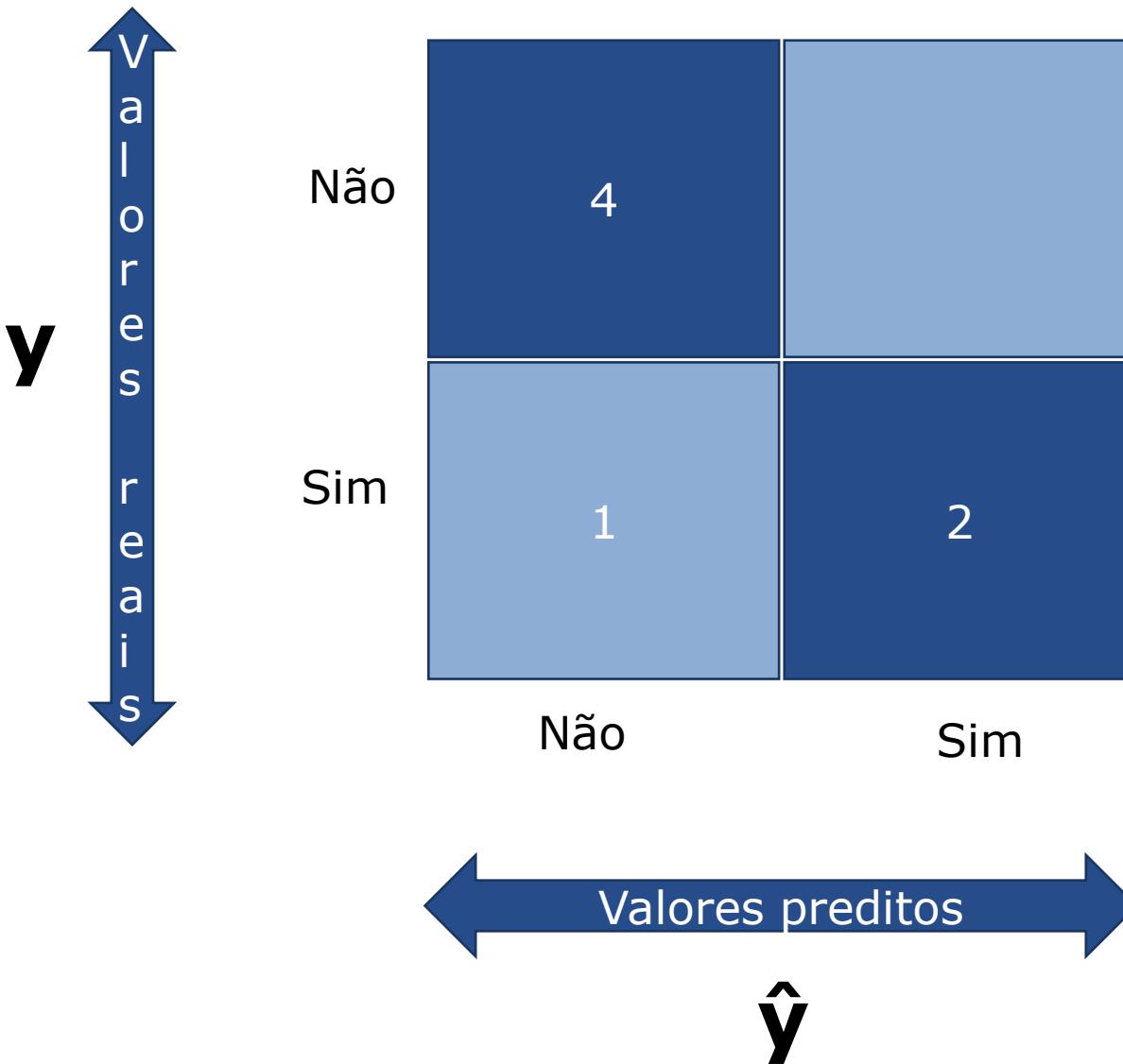
Matriz de confusão treino

	y	\hat{y}
0	Não	Não
1	Não	Não
2	Não	Não
3	Sim	Sim
4	Não	Não
5	Sim	Não
6	Sim	Sim



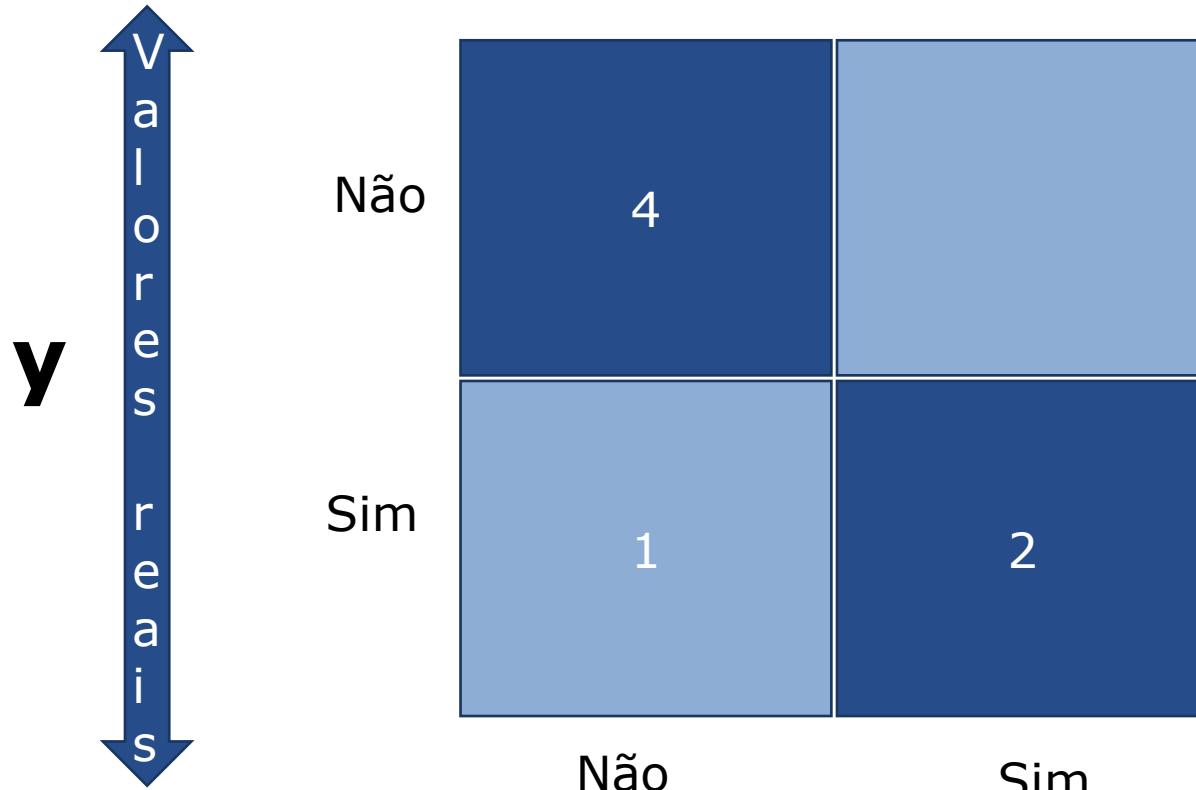
Matriz de confusão treino

	y	\hat{y}
0	Não	Não
1	Não	Não
2	Não	Não
3	Sim	Sim
4	Não	Não
5	Sim	Não
6	Sim	Sim



Acurácia conjunto de treino

	y	\hat{y}
0	Não	Não
1	Não	Não
2	Não	Não
3	Sim	Sim
4	Não	Não
5	Sim	Não
6	Sim	Sim



Acurácia = (Quantidade de nãoXnão + Quantidade de simXsim) / TOTAL

$$Acc = (4+2)/(4+2+1)$$

$$= 0,857$$

Matriz de confusão teste

	x1	x2	y	\hat{y}
7	15	500	Não	Não
8	26	600	Sim	
9	10	1000	Sim	

$X < 23$

Verdadeiro

Não

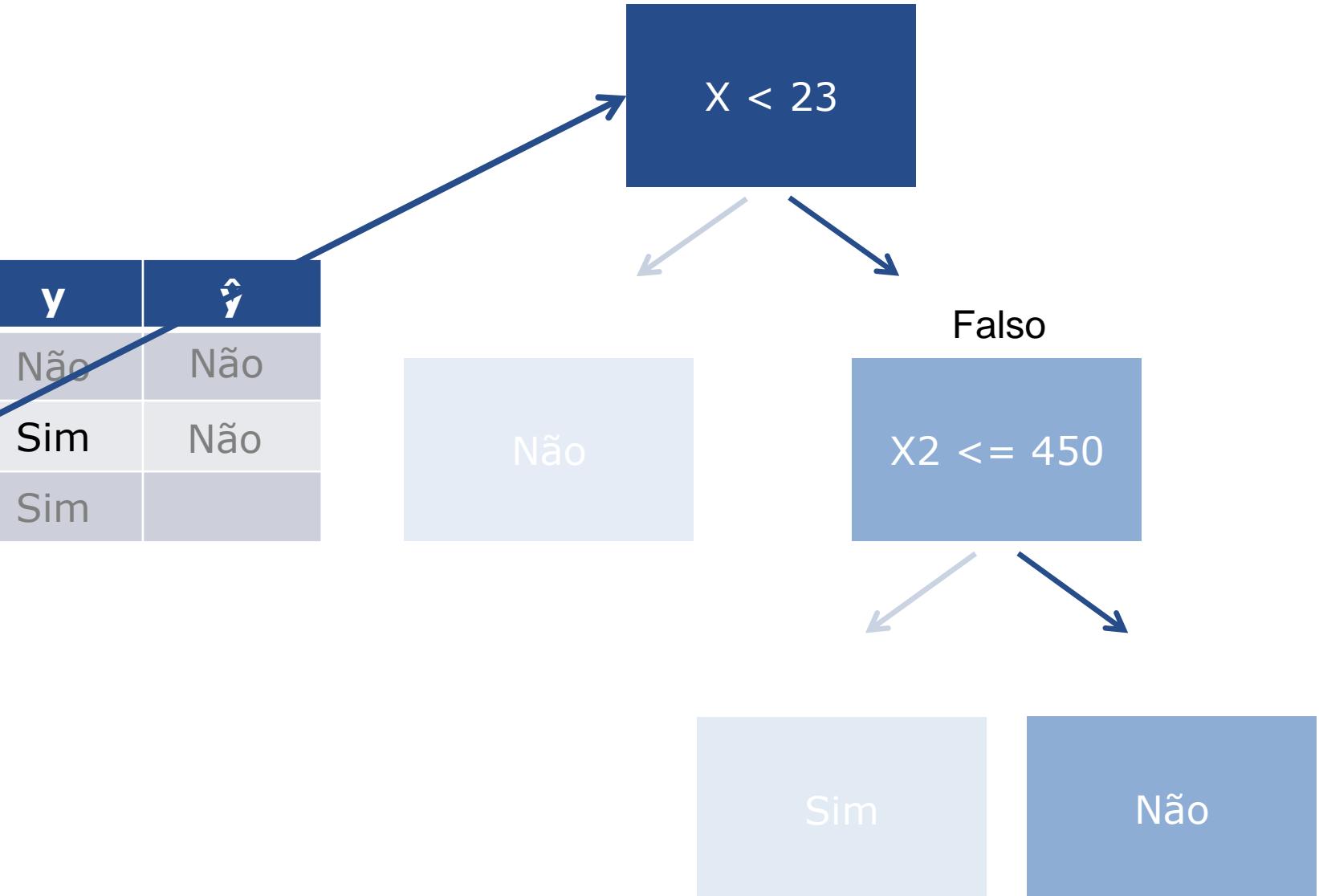
$X2 \leq 450$

Sim

Não

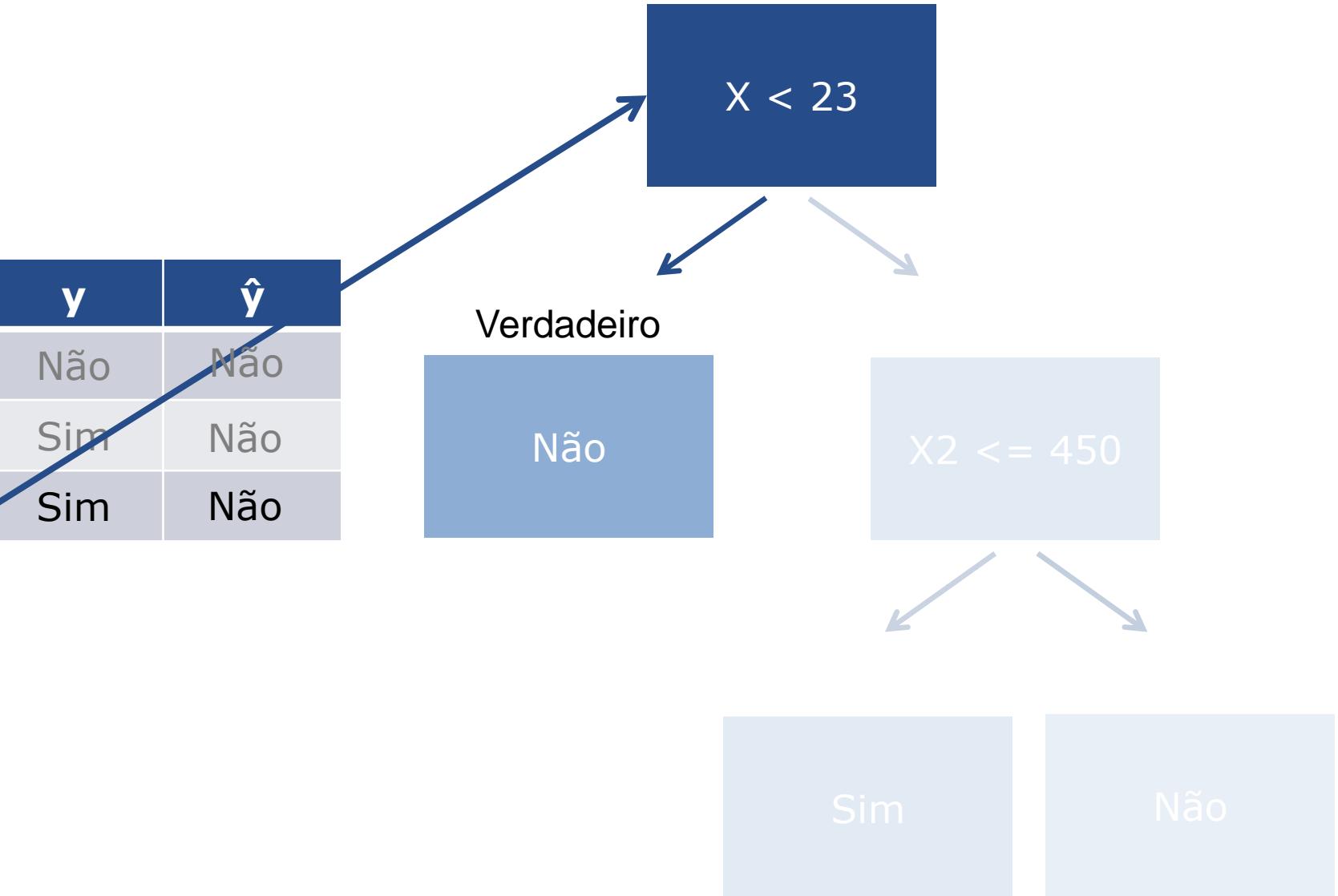
Matriz de confusão teste

	x1	x2	y	\hat{y}
7	15	500	Não	Não
8	26	600	Sim	Não
9	10	1000	Sim	



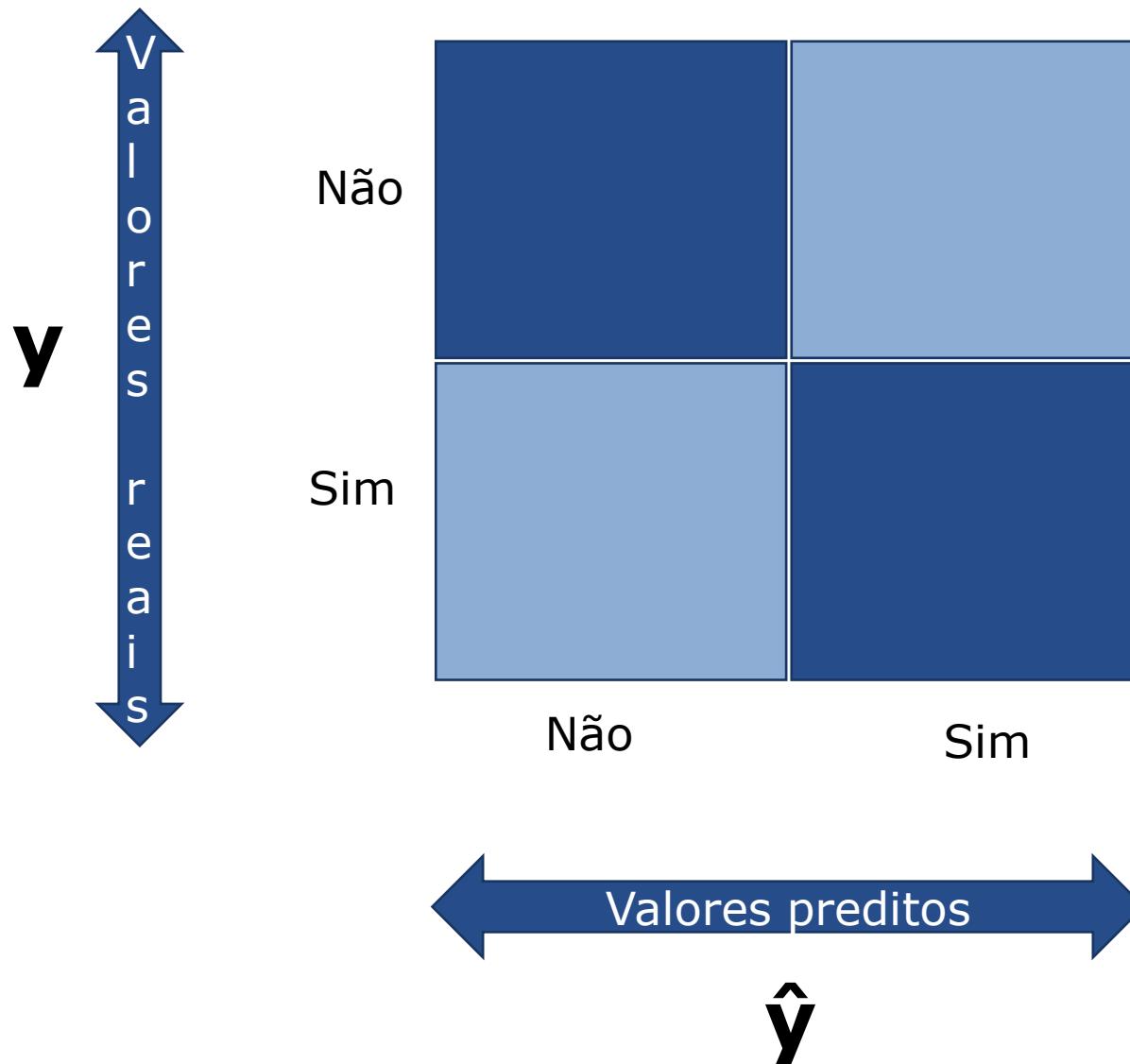
Matriz de confusão teste

	x1	x2	y	\hat{y}
7	15	500	Não	Não
8	26	600	Sim	Não
9	10	1000	Sim	Não



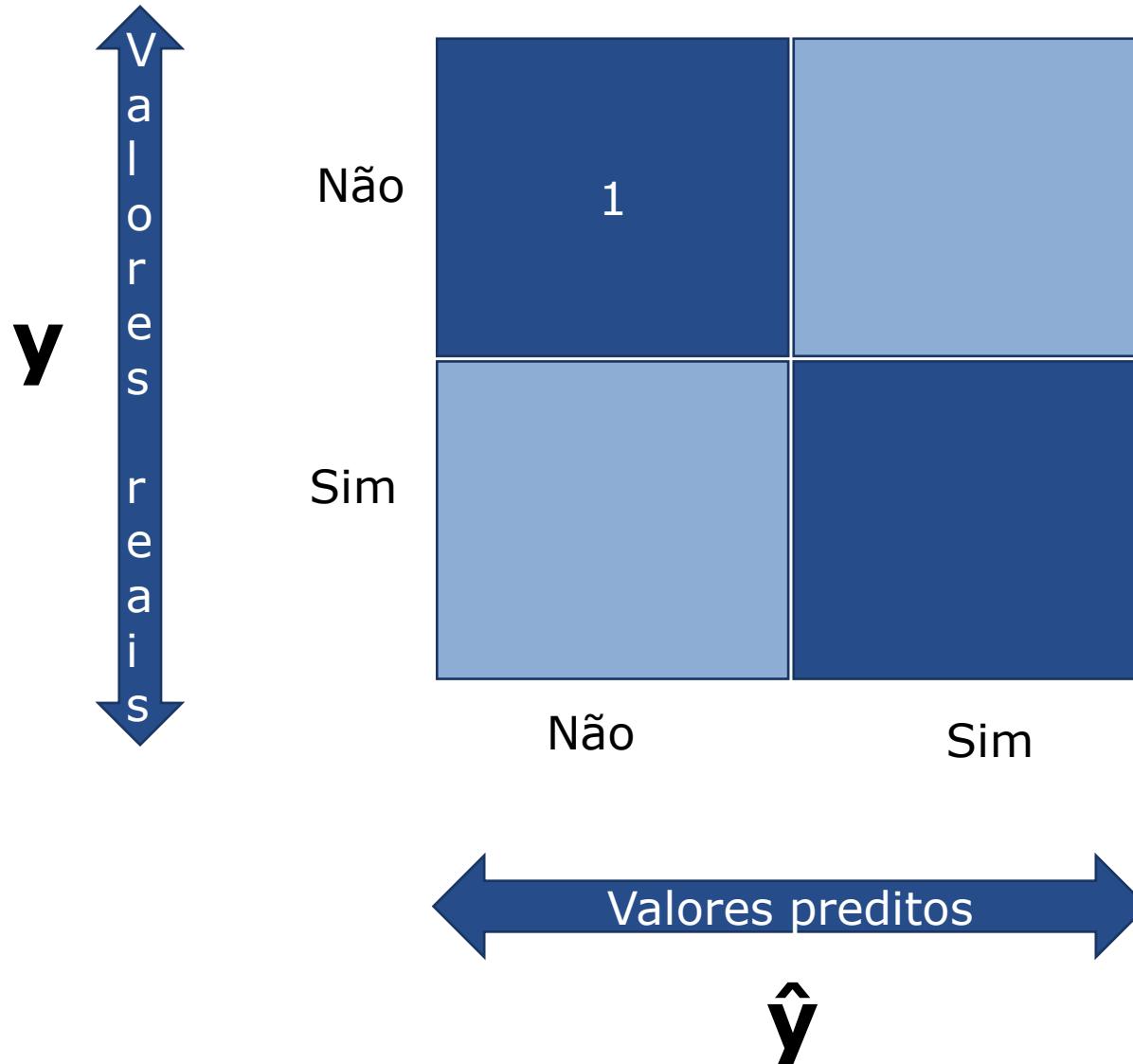
Matriz de confusão teste

	y	\hat{y}
7	Não	Não
8	Sim	Não
9	Sim	Não



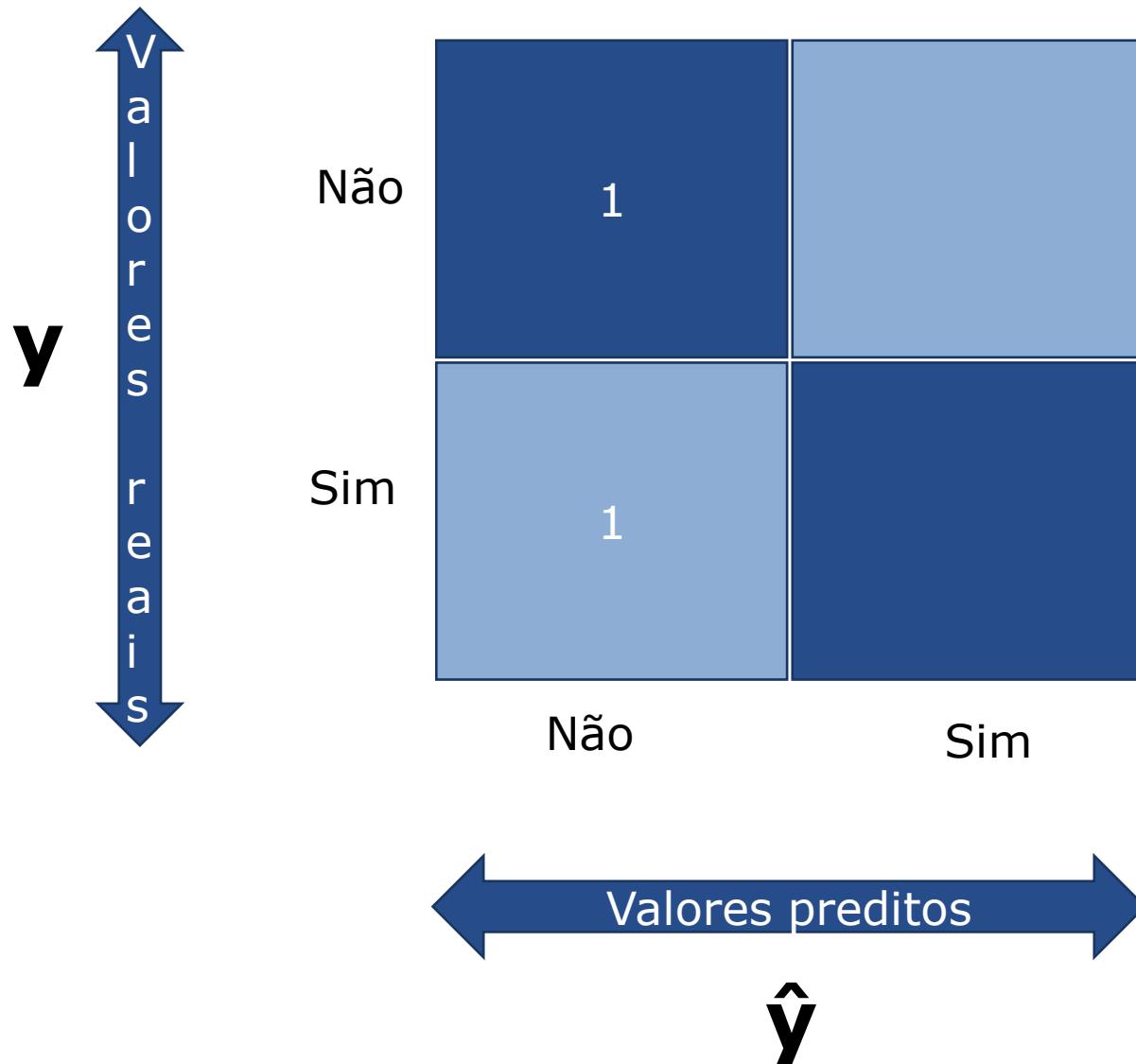
Matriz de confusão teste

	y	\hat{y}
7	Não	Não
8	Sim	Não
9	Sim	Não



Matriz de confusão teste

	y	\hat{y}
7	Não	Não
8	Sim	Não
9	Sim	Não



Acurácia conjunto de teste

	y	ŷ
7	Não	Não
8	Sim	Não
9	Sim	Não

$$\text{Acc} = 1/(2+1) \\ = 0,333$$

