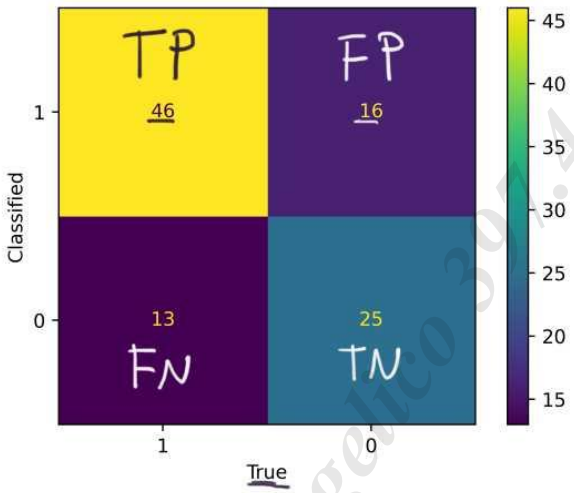
Prof. Luiz Paulo Lopes Fávero

PRINTS TIRADOS DURANTE A AULA DE 27/08/2024:

cutoff = 0,5.



59 41

Matriz de Confusão (Confusion Matrix):

Classified \ True	1	0
1	TP	FP
0	FN	TN

$\text{Precision} = \frac{TP}{TP + FP}$

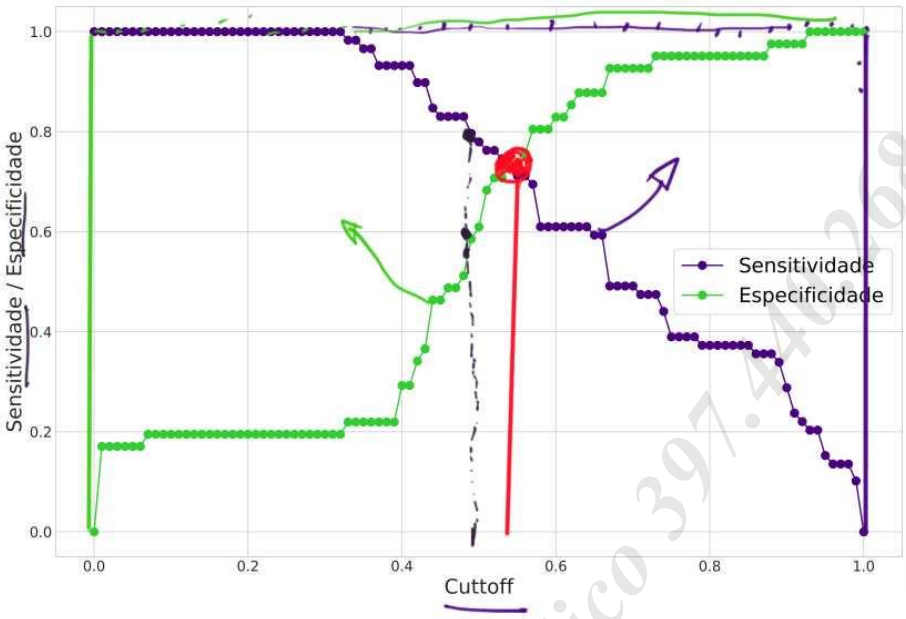
Para Determinado cutoff!

$\text{Sensitividade (Recall)} = \frac{TP}{TP + FN}$

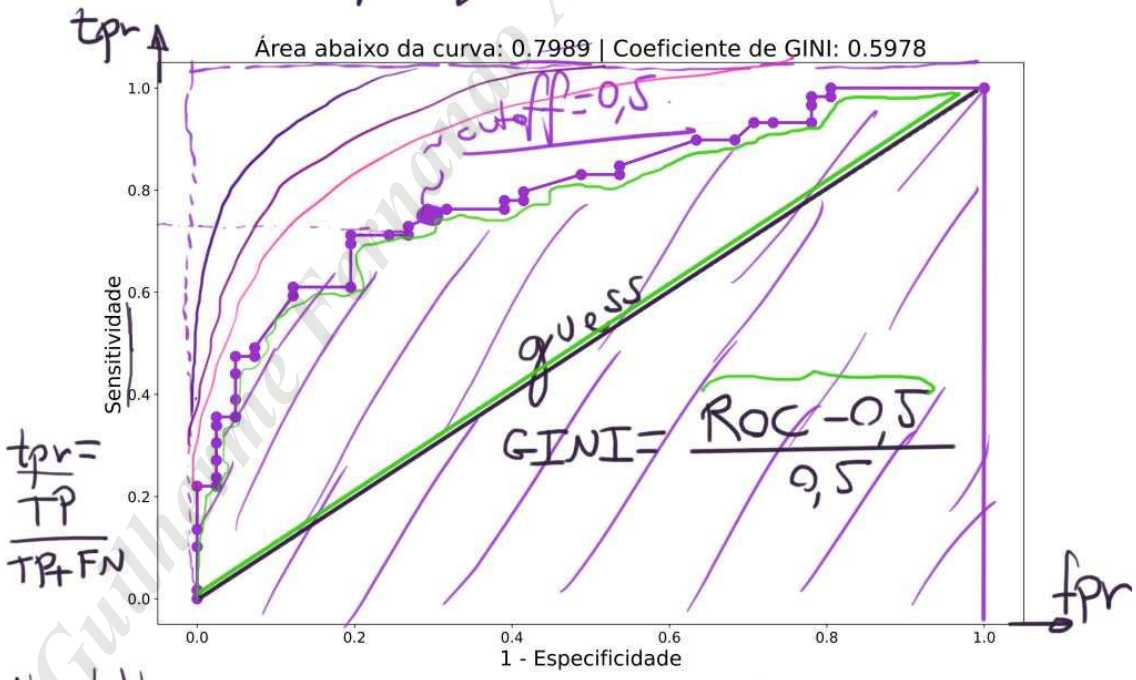
$\text{Especificidade} = \frac{TN}{FP + TN}$

$\text{Acurácia (EGM)} = \frac{TP + TN}{TP + FP + FN + TN}$

$\text{F1\_score} = \frac{2 \times \text{Recall} \times \text{Precision}}{(\text{Recall} + \text{Precision})}$ { Média Harmônica entre Recall & Precision.



ROC: Receiver Operating Characteristic



$GINI = \frac{ROC - 0,5}{0,5}$

$tpr = \frac{TP}{TP+FN}$

thresholds = cutoffs

$fpr = \frac{FP}{FP+TN} = 1 - especificidade$

Indicadores / Estatísticas :

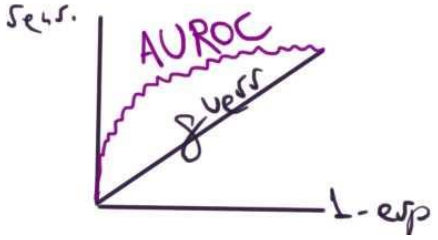
Independem do cutoff

- Log Lik (LLm → LLf) ↑
- χ^2 (p-value χ^2 → cuidado ao p-value F).

$$\chi^2 = -2(LL_0 - LL_m)$$

- $AIC = -2 \cdot LL_m + 2 \cdot (k+1)$
- $BIC = -2 LL_m + (k+1) \cdot \ln(n)$

↓



$$GINI = \frac{ROC - 0,5}{0,5}$$

Robert Clemen

Making Hard Decisions (1997)

$$P_i = \frac{1}{1 + e^{-(\alpha + \beta_1 X_{1,i} + \cdots + \beta_K X_{K,i})}}$$

$$e^{\beta} = \text{odds ratio} = \frac{p}{1-p}$$

$$\ln(\text{odds}) = \alpha + \beta_1 X_{1,i} + \cdots + \beta_K X_{K,i}$$

$$\text{odds} = e^{\beta_1}$$

odds ratio

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	23.7750	11.820	2.011	0.044	0.607	46.943
Q('temperatura')	-0.3667	0.175	-2.093	0.036	-0.710	-0.023

$$P_{falha=1} = \frac{1}{1 + e^{-(23{,}77 - 0{,}366 * temp)}}$$

Logística Binária:

$$P_0 = 1 - \frac{e^z}{1+e^z} = \frac{1}{1+e^z} \cdot \frac{e^0}{e^0}$$ Não Evento

$$P_1 = \frac{1}{1+e^{-z}} = \frac{e^z}{1+e^z} \cdot \frac{e^0}{e^0}$$ Evento

$$\text{Soma} = 1$$

$$z_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki}$$

(Dist. Binomial) Logística Multinomial: (Ex.: 3 cat. y).

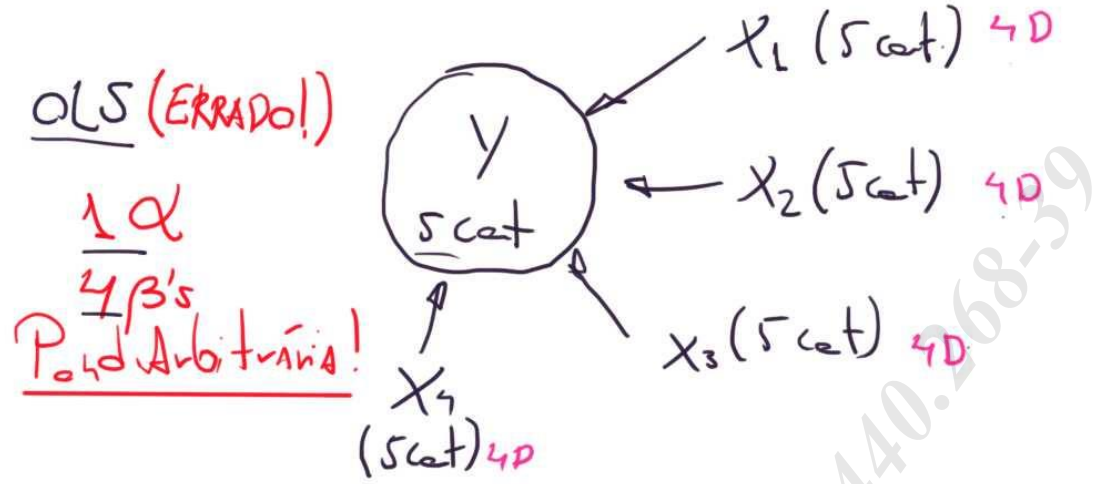
$$\begin{cases} z_1 = \alpha_1 + \beta_{11} X_{1i} + \beta_{21} X_{2i} + \cdots + \beta_{K1} X_{Ki} \\ z_2 = \alpha_2 + \beta_{12} X_{1i} + \beta_{22} X_{2i} + \cdots + \beta_{K2} X_{Ki} \end{cases}$$

ref. $$P_0 = \frac{e^0}{e^0 + e^{z_1} + e^{z_2}} = \frac{1}{1 + e^{z_1} + e^{z_2}}$$

H1 $$P_1 = \frac{e^{z_1}}{1 + e^{z_1} + e^{z_2}}$$

H2 $$P_2 = \frac{e^{z_2}}{1 + e^{z_1} + e^{z_2}}$$

$$\text{Soma} = 1$$



Logística Multinomial com Dummies (CORRETO!)

4 logitos com 1 α e 16 β's cada.

4 α's
64 β's

atrasado2=1		coef	std err	z	P>|z|	[0.025	0.975]
const	α_1	-33.1352	12.183	-2.720	0.007	-57.014	-9.256
dist	β_{11}	0.5588	0.243	2.297	0.022	0.082	1.036
sem	β_{21}	1.6699	0.577	2.895	0.004	0.539	2.801
atrasado2=2		coef	std err	z	P>|z|	[0.025	0.975]
const	α_2	-62.2922	14.675	-4.245	0.000	-91.055	-33.530
dist	β_{21}	1.0784	0.302	3.566	0.000	0.486	1.671
sem	β_{22}	2.8949	0.686	4.220	0.000	1.550	4.239

$$P_0 = \frac{1}{1 + e^{(-33,13 + 0,55 \cdot dist_i + 1,67 \cdot sem_i)} + e^{(-62,29 + 1,07 \cdot dist_i + 2,89 \cdot sem_i)}}$$

Z_1 $\qquad$ Z_2

$$P_1 = \frac{e^{(-33,13 + 0,55 \cdot dist_i + 1,67 \cdot sem_i)}}{\text{*MESMO DENOMINADOR}}$$

$$P_2 = \frac{e^{(-62,29 + 1,07 \cdot dist_i + 2,89 \cdot sem_i)}}{\text{*MESMO DENOMINADOR}}$$

SOMA = 1

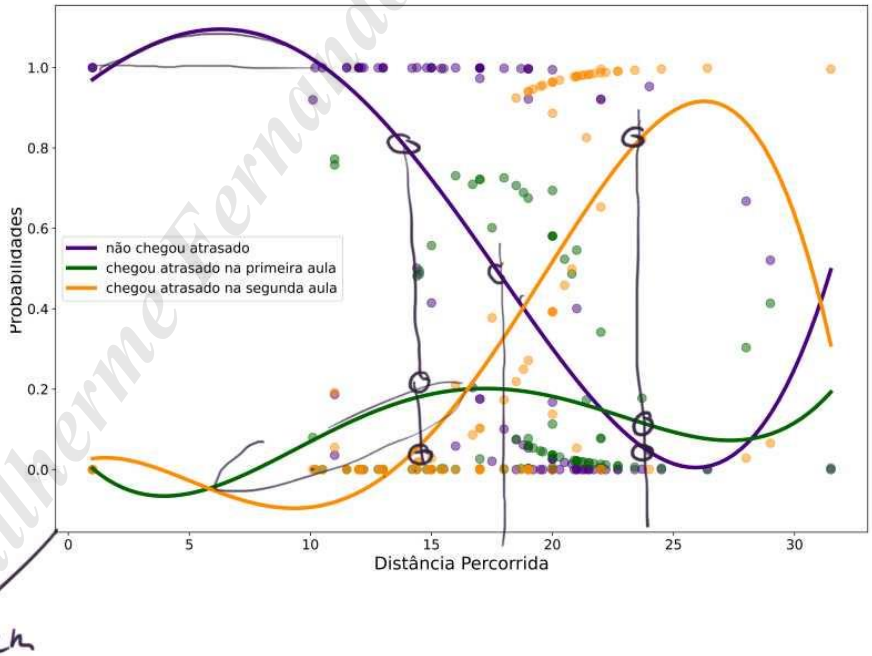
true

clas

predicao_label	chegou atrasado primeira aula	chegou atrasado segunda aula	nao chegou atrasado
chegou atrasado primeira aula	12	5	2
chegou atrasado segunda aula	3	30	0
não chegou atrasado	1	0	47

$$EGH = \frac{12 + 30 + 47}{100} = 89\%$$

ACURÁCIA.



seh