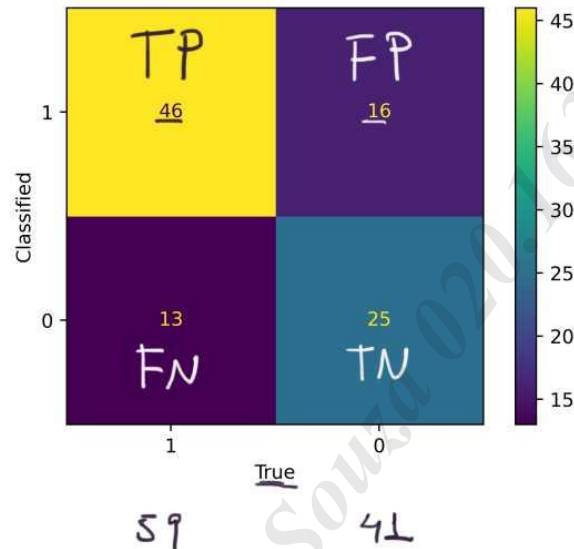Prof. Luiz Paulo Lopes Fávero

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

cutoff = 0,5.

Matriz de Confusão (Confusion Matrix):

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

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

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

Para Determinado cutoff!

$$Acurácia\ (EGM) = \frac{TP + TN}{TP + FP + FN + TN}$$
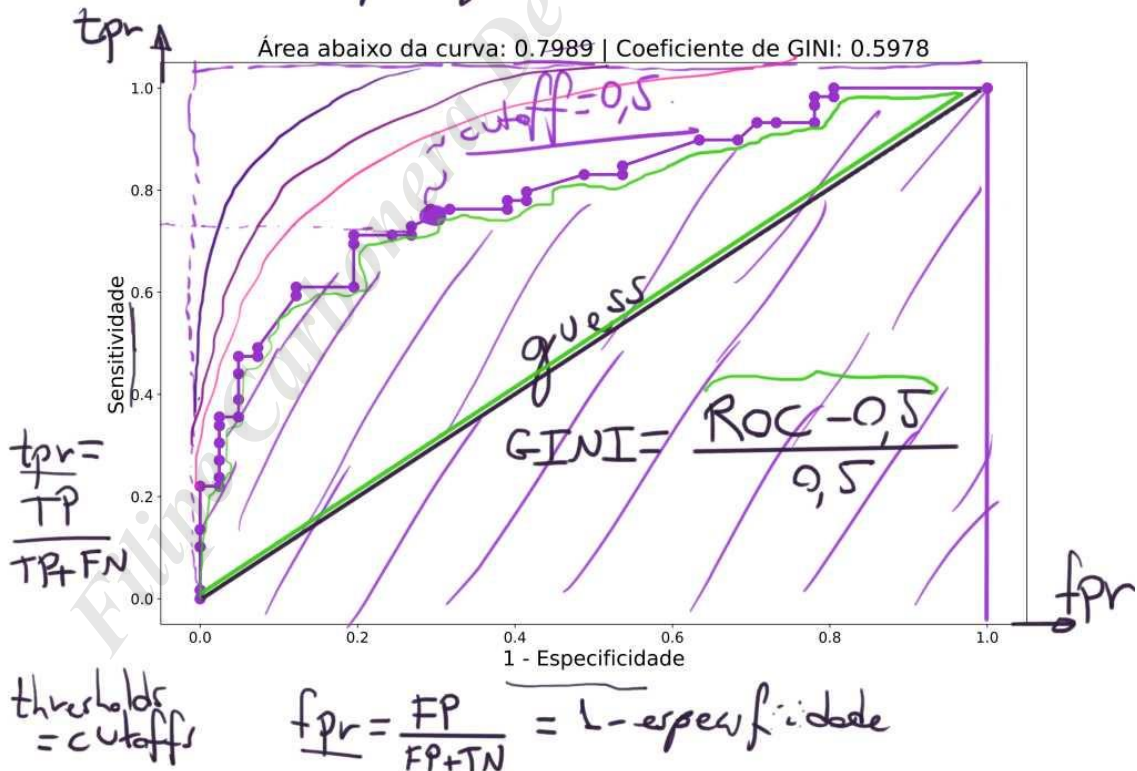
$$F1\_score = \frac{2 \times Recall \times Precision}{(Recall + Precision)}$$

Média Harmônica entre Recall e Precision.
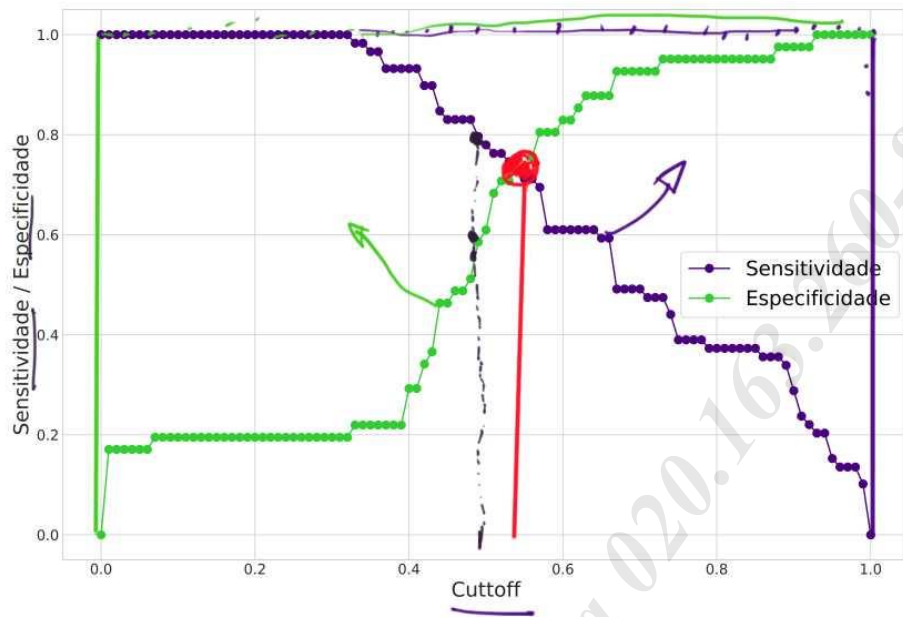
ROC: Receiver Operating Characteristic

Área abaixo da curva: 0.7989 | Coeficiente de GINI: 0.5978

cutoff = 0,5

guess

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

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

thresholds = cutoffs

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

Indicadores / Estatísticas : *Independem do cutoff*

- LogLik $(LL_m \to LLf)$ ↑

- χ^2 (p-value $\chi^2 \to$ análogo ao p-value F).

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

- $AIC = -2 \cdot LL_m + 2 \cdot \underset{\beta}{(\underline{k}+1)}$

- $BIC = -2 LL_m + (\underline{k}+1) \cdot l_n(n)$] ↓

sens.
AUROC
guess
1-esp
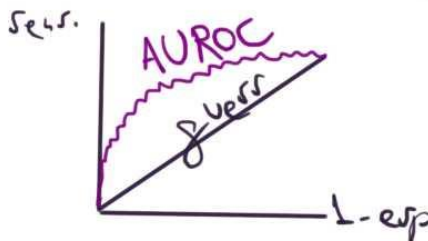
$$GINI = \frac{ROC - 0,5}{0,5}$$

Robert Clemen

Making Hard Decisions
(1997)

$$P_i = \frac{1}{1 + e^{-(\alpha + \beta_1 X_{1_i} + \cdots + \beta_k X_{k_i})}}$$

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

$$\ln(odds) = \alpha + \beta_1 X_{1_i} + \cdots + \beta_k X_{k_i}$$

$$odds = e^{\beta_i}$$

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
========================================================================
```

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

Logística Binária:

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

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

$$\text{Soma} = 1$$

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

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

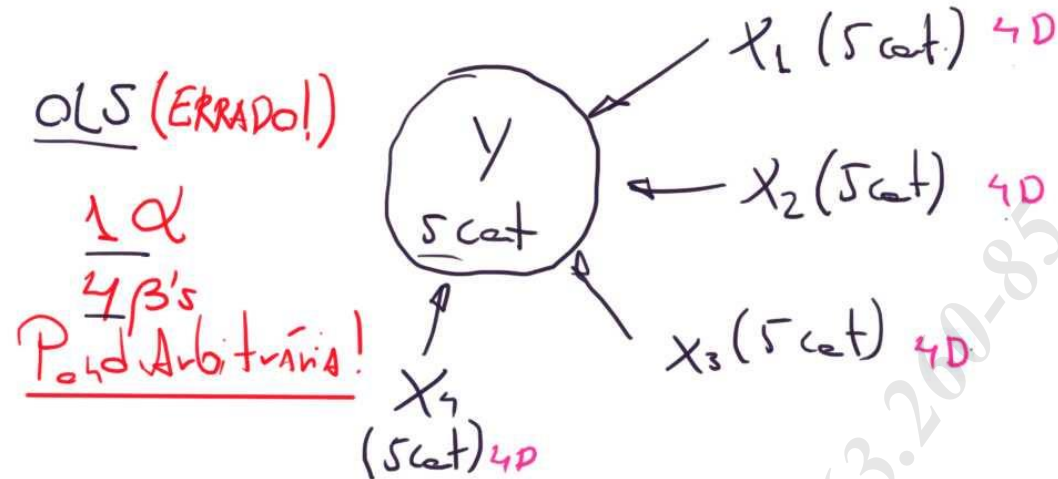$$\begin{cases} Z_1 = \alpha_1 + \beta_{11} X_{1i} + \beta_{21} X_{2i} + \cdots + \beta_{K1} X_{Ki} \\ Z_2 = \alpha_2 + \beta_{12} X_{1i} + \beta_{22} X_{2i} + \cdots + \beta_{K2} X_{Ki} \end{cases}$$

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

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

$$\text{Soma} = 1$$

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

OLS (ERRADO!)

$\dfrac{1\alpha}{4\beta's}$

P. 4D Arbitrária!

X_1 (5 cat) 4D

X_2 (5 cat) 4D

X_3 (5 cat) 4D

X_4 (5 cat) 4D

Y 5 cat

Logística Multinomial com Dummies (CORRETO!)

4 logitos com 1α e $16\beta's$ cada.

$$4\alpha's$$
$$64\beta's$$

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

z_1

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

z_2

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

z_1 ; z_2

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

$\ast$ MESMO DENOMINADOR

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

$\ast$ MESMO DENOMINADOR   Soma = 1

true

clsr

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)

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

ACURÁCIA.

Seh