



# Prova IA

## antiga





PUC Minas

Nome: Ysmela Helen Carlele Borges

Data: 26/09/2023

1. Esta avaliação vale 26 pontos
2. É individual e SEM consulta
3. A interpretação faz parte da avaliação

23,0

**Questão 01 (1,5 pontos)**

Considere a seguinte matriz de confusão obtida por meio do classificador, Árvore de decisão, para um problema de quatro classes:

TOTAL = 264

Foi classificado como

	A	B	C	D	E
A	20	2	1	0	2
B	2	30	5	0	1
C	15	10	40	3	5
D	4	6	10	20	10
E	2	4	0	5	67

Handwritten notes: 25 = 20, 28 = 2, 73 = 15, 50 = 10, 78 = 40. Red boxes around cells (A,C), (B,C), (C,C), (D,C), (E,C). Red curly braces on the right: {63} for row C, {128} for column C, {101} for column E.

O valor da Taxa de Falso Positivo (TFP) para a classe C:

<input checked="" type="checkbox"/> a) 16/191	SOMA LINHAS	COLUMNA
<input type="checkbox"/> b) 40/191	A = 25	C = 16
<input type="checkbox"/> c) 40/73	B = 38	
<input type="checkbox"/> d) 16/56	D = 50	
<input type="checkbox"/> e) 40/264	E = 78	
	TOTAL = 191	191

Handwritten: TFP = 16/191

NUMERADOR

	A	B	C	D	E
VP	20/25	30/38	40/73	20/50	67/78
FN	5/25	8/38	33/73	30/50	11/78
FP	23/229	22/226	16/191	8/214	18/186
VN					

**Questão 02 (1,5 pontos)**

Considerando-se a matriz de confusão da questão anterior, e as seguintes afirmações a respeito a respeito dela:

- I. A maior precisão é para a Classe E com valor de 78,82%
- II. O recall da classe E é mais que o dobro da Classe D
- III. A maior taxa de Falsos negativos é para a classe D com mais de 70%

TFN(D) = 60%

É correto o que se afirma em: recall<sub>(E)</sub> = 85,897%.

recall<sub>(D)</sub> = 40%

- a) I, apenas.
- ☒ b) II, apenas.
- c) III, apenas.
- ☒ d) I e II, apenas.
- e) I, II e III.

	PRECISÃO
A	89,20%
B	
C	
D	
E	

Handwritten: 40,0

precisão = VP

VP + FP

recall = VP

VP + FN

# TAXAS

Q1-

	A	B	C	D	E
TVP	20/25	30/38	33/73	20/50	67/78
TEN	5/25	6/38	40/73	30/50	11/78
TVN	216/129	201/226	175/191	200/219	18/136
TFP	23/25	22/26	16/191	8/219	118/136
Precisão	20/43				67/89
Recall	20/25			20/50	67/78

SOMA=101  
NÃO FORAM  
class. como C = 10

$$P = \frac{67/78}{67/78} \cdot 118/136$$

$$= \frac{67/78}{0,745}$$

$$= 1,15$$

$$F_1 \text{ score} = \frac{2 \cdot \text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}}$$

TFP = não são da classe que quero, mas foram classificados como.

TVP = número de acerto do classificador p/ cada classe

Q2- Precisão =  $\frac{VP}{VP + FP}$

Recall =  $\frac{VP}{VP + FN}$

$$A = \frac{20/25}{20/25 + 23/25} = \frac{0,8}{0,8 + 0,096} = 0,892$$

Positivos verdadeiros  
Positivos verdadeiros  
+  
Falsos positivos

Positivos verdadeiros  
Positivos verdadeiros  
+  
Falsos Negativos

(coluna)

(linha)

bits)

ite matriz de confusão obtida por meio do classificador, Árvore de decisão, para um o classes:

Foi classificado como

	A	B	C	D	E
A	20	2	1	0	2
B	2	30	5	0	1
C	15	10	40	3	5
D	4	6	10	20	10
E	2	4	0	5	67

25= 20  
38= 30  
73= 40  
50= 20  
78= 67

63  
128  
101

also Positivo (TFP) para a classe C

↑  
NUMERADOR

$$B = \frac{30/38}{30/38 + 22/26} = \frac{0,78}{0,78 + 0,857} = 0,889$$

$$C = \frac{33/73}{\frac{33}{73} + \frac{16}{191}} = \frac{0,45}{0,45 + 0,083} = 0,843$$

$$D = \frac{20/50}{20/50 + 8/214} = \frac{0,4}{0,4 + 0,037} = 0,914$$

menor  
preço

$$E = \frac{67/78}{67/78 + 118/136} = \frac{0,858}{0,856 + 0,86} = 0,496$$

$$\text{Recall } E = \frac{67/78}{67/78 + 118/136} = \frac{0,856}{1} = 0,856$$

$$D = \frac{20/50}{20/50 + 20/50} = 0,5$$

NÃO FAÇA

ASSIM!!!

(acabei calculando  
pelos taxas e não  
pelos valores em si)

(me pito de  
br não tá  
dando certo)

Questão 03 (1,5 ponto)

A respeito do algoritmo ID3, avalie as afirmações abaixo:

- I. Ele resolve apenas problemas de classificação nominal ✓
- II. Ele trabalha com atributos nominais e numéricos X
- III. Ele utiliza o ganho de informação para avaliar a importância do atributo ✓

É correto o que se afirma em:

- f) I, apenas.
- g) II, apenas.
- h) III, apenas.
- ☒ i) I e III, apenas.
- j) I, II e III.

discretos e categóricos (sem ordem ou hierarquia)

valores contínuos ou ordenados

Questão 04 (1,5 ponto)

A respeito do algoritmo C45, avalie as afirmações abaixo:

- I. Ele resolve problemas de classificação nominal e numérica regressão X
- II. Ele trabalha com base de dados com dados ausentes ✓
- III. Ele avalia a importância do atributo utilizando a mesma equação do ID3; ou seja, o ganho de informação. X

É correto o que se afirma em:

- a) I, apenas.
- ☒ b) II, apenas.
- c) III, apenas.
- ☒ d) II e III, apenas.
- e) I, II e III.

não é de regressão

Unknown (?)

ele usa o gain ratio

Questão 05 (1,5 ponto)

Considere os algoritmos GridSearch e RandomSearch, e as seguintes afirmações:

- I. Eles são importantes para encontrar o melhor algoritmo de aprendizado a ser aplicado em uma determinada base de dados
- II. Eles são importantes para encontrar os melhores hiperparâmetros de um algoritmo de aprendizado, desde que o usuário forneça a faixa de variação a ser investigada ✓
- III. Eles são importantes para encontrar os melhores hiperparâmetros de um algoritmo de aprendizado, sem fornecimento de faixa de variação fornecida pelo usuário.

vai cair?

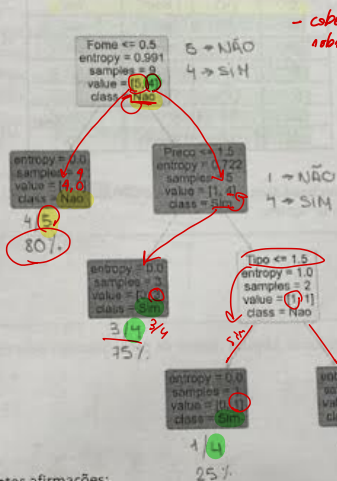
É correto o que se afirma em:

- a) I, apenas.  
 X II, apenas.  
 c) III, apenas.  
 d) I e II, apenas.  
 e) I, II e III.

C 1,5

**Questão 06 (1,5 ponto)**

Considere a árvore de decisão obtida a partir da base de dados 'Esperar ou não pelo Restaurante'.



Considere ainda as seguintes afirmações:

- I. A ~~menor~~ <sup>MAIOR</sup> cobertura por classe a partir desta árvore é de 80% e corresponde à classe 'Não' ✓  
 II. A menor cobertura por classe a partir desta árvore é de 20% e corresponde à classe 'Sim' X  
 III. Há duas regras nesta árvore com a mesma cobertura X

É correto o que se afirma em:

- X I, apenas.  
 b) III, apenas.  
 c) I e II, apenas.  
 d) I e III, apenas.  
 e) I, II e III.

C 1,5

Questão 07 (8,5 pontos)  
 Considere a seguinte base de dados, em que se deseja avaliar se o cogumelo é comestível ou não.

	Forma	Cor	Odor	Comestível
1	C	B	Ruim	Yes
2	D	B	Ruim	Yes
3	D	W	Ruim	Yes
4	D	W	Bom	Yes
5	C	B	Bom	Yes
6	D	B	Bom	No
7	D	G	Bom	No
8	C	U	Bom	No
9	C	B	Ótimo	No
10	C	W	Ótimo	No
11	D	W	Ótimo	No

Usando Naive Bayes, qual a probabilidade de o cogumelo ser comestível (YES) ou não comestível (NO), respectivamente, para o seguinte registro:

Forma = D

Cor = B

Odor = Bom

Obs: É necessário apresentar todos os cálculos (matriz de probabilidades para todos os atributos, cálculos das porcentagens e etc).

- a) 100% e 0%  
 b) 0% e 100%  
 c) 59,0164% e 40,9836%  
 d) 48,9796% e 51,0204%  
 e) 40,9836% e 59,0164%

FORMA		COR							
C	D	B	W	G	U		RUIM	BOM	ÓTIMO
2/5	3/5	3/5	2/5	0/5	0/5		3/5	2/5	0/5
3/6	3/6	2/6	2/6	1/6	1/6		0/6	3/6	3/6

$$\text{YES} \rightarrow 0,06545 \div 0,1109 = 0,590171 \approx 59,01\%$$

$$\text{NO} \rightarrow 0,04545 \div 0,1109 = 0,40982 \approx 40,98\%$$

Q7-

Format D  
Cor B  
Order from  
Yes e No

SIM:

$$\frac{\cancel{5}}{11} \cdot \frac{3}{\cancel{5}} \cdot \frac{3}{5} \cdot \frac{2}{5} = \frac{18}{275} = \frac{0,0654}{0,1108} = 0,5901$$

$$0,0654 + 0,0454 = 0,1108$$

Yes = 59,01%  
No = 40,97%

NAB:

$$\frac{\cancel{6}}{11} \cdot \frac{3}{\cancel{6}} \cdot \frac{2}{6} \cdot \frac{3}{6} = \frac{18}{396} = \frac{0,0454}{0,1108} = 0,4097$$

Retna c)

40,97%



**Questão 08 (8,5 pontos)**

Considerando a mesma base de dados do exercício anterior e o algoritmo de **árvore de decisão**, qual a raiz da árvore e qual o ganho do atributo, respectivamente?

Obs: 1. É necessário apresentar todos os cálculos. Ou seja, não será considerada questão sem apresentação dos cálculos.

- a) A raiz da árvore é o atributo **Ganho da Forma** com ganho de 0,007234 ✗
- b) A raiz da árvore é o atributo **Cor** com ganho de 0,189053 ✗
- ✗ c) A raiz da árvore é o atributo **Odor** com ganho de 0,552689 ○
- d) A raiz da árvore é o atributo **Cor** com ganho de 0,552689 ✗
- e) A raiz da árvore é o atributo **Odor** com ganho de 0,798967 ✗

Coloque na tabela abaixo os valores encontrados para:

Entropia da classe	0,99403
Entropia do atributo Forma	0,98679
Entropia do atributo Cor	0,80498
Entropia do atributo Odor	0,44134
Ganho do atributo Forma	0,00724
Ganho do atributo <del>Forma</del> COR	0,18905
Ganho do atributo <del>Forma</del> ODOR	0,55269

C 8,5

$$\text{Entropia (classe)} = \left( \frac{5}{11}, \frac{6}{11} \right) = \underbrace{-\frac{5}{11} \cdot \log_2 \left( \frac{5}{11} \right)}_{0,51705} - \underbrace{\frac{6}{11} \cdot \log_2 \left( \frac{6}{11} \right)}_{0,47698} = 0,99403$$

$$\text{Entropia (forma)} = \left( \frac{5}{11}, \frac{6}{11} \right) = \frac{5}{11} \cdot E\left(\frac{2}{5}, \frac{3}{5}\right) + \frac{6}{11} \cdot E\left(\frac{1}{2}, \frac{1}{2}\right) =$$

$$E\left(\frac{2}{5}, \frac{3}{5}\right) = \underbrace{-\frac{2}{5} \cdot \log_2 \left( \frac{2}{5} \right)}_{0,52877} - \underbrace{\frac{3}{5} \cdot \log_2 \left( \frac{3}{5} \right)}_{0,44218} = 0,97095$$

$$E\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{2} \cdot \log_2 \left( \frac{1}{2} \right) - \frac{1}{2} \cdot \log_2 \left( \frac{1}{2} \right) = 1$$

**Equações**

$$Pr = \frac{VP}{VP + FP}$$

$$Sens = \frac{VP}{VP + FN}$$

$$I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$$

CÁLCULOS NO  
VERSO →

08-

base de dados, em que se deseja avaliar a qualidade da classificação.

C-D	B-V-G-U	Ruim-Bom	Forma	Cor	Odor	Comestível
1	C	B	Ruim	Yes		
2	D	B	Ruim	Yes		
3	D	W	Ruim	Yes		
4	D	W	(Bom)	Yes		
5	C	B	(Bom)	Yes		
6	D	B	(Bom)	No		
7	D	G	(Bom)	No		
8	C	U	(Bom)	No		
9	C	B	Ótimo	No		
10	C	W	Ótimo	No		
11	D	W	Ótimo	No		

al a probabilidade de o cogumelo ser comestível (YES) ou não  
seguinte registro:

$$\text{Entropia (classe)} = Y_{us}: \frac{5}{11} \cdot \log_2 \frac{5}{11} \quad N_{o}: \frac{6}{11} \cdot \log_2 \frac{6}{11}$$

$$- (-0,517 + -0,476) = \underline{0,993}$$

$$\text{Entropia (forma)} = C: Y_{us}: \frac{2}{5} \cdot \log_2 \frac{2}{5} \quad N_{o}: \frac{3}{5} \cdot \log_2 \frac{3}{5}$$

$$- (-0,528 + -0,472) = 0,970 //$$

$$D: Y_{us}: \frac{3}{6} \cdot \log_2 \frac{3}{6} \quad N_{o}: \frac{3}{6} \cdot \log_2 \frac{3}{6}$$

$$-0,5 \quad -0,5$$

$$- (-0,5 + -0,5) = 1 //$$

$$\begin{array}{r} 0,993 \\ - 0,986 \\ \hline 0,007 \end{array}$$

$$\rightarrow \frac{5}{11} \cdot 0,970 + \frac{6}{11} \cdot 1 = 0,986$$

$$\text{ganho} = \underline{0,993} - 0,986 = \underline{0,007} //$$

$$\text{Entropia (cor)} = B: \text{sim: } \frac{3}{5} \cdot \log_2 \frac{3}{5} \quad \text{N\tilde{A}O: } \frac{2}{5} \cdot \log_2 \frac{2}{5} \\ = -0,492 \quad = 0,970 \quad = -0,528$$

$$W: \text{sim: } \frac{2}{4} \quad \text{N\tilde{A}O: } \frac{2}{4} \\ = -0,5 \quad = 1 \quad = -0,5$$

$$G: \text{sim: } \frac{0}{1} \cdot \log_2 0 = 0 \quad \text{N\tilde{A}O: } \frac{1}{1} \cdot \log_2 1 \\ = 0 \quad = 0$$

$$M: \text{sim: } \frac{0}{1} \quad \text{N\tilde{A}O: } \frac{1}{1} \\ = 0 \quad = 1$$

$$\rightarrow \frac{5}{11} \cdot 0,970 + \frac{4}{11} \cdot 1 + \frac{1}{11} \cdot 0 + \frac{1}{11} \cdot 0 = \frac{117}{220} = 0,8045$$

$$\text{ganho} = 0,993 - 0,8045 = 0,1885 //$$

$$\text{Entropia (odor)} = \text{ruim: sim: } \frac{3}{3} \quad \text{N\tilde{A}O: } \frac{0}{3} \\ = 0$$

$$\text{bom: sim: } \frac{2}{5} \quad \text{N\tilde{A}O: } \frac{3}{5} \\ = 0,970$$

$$\text{ótimo: sim: } \frac{0}{3} \quad \text{N\tilde{A}O: } \frac{3}{3} \\ = 0$$

$$\rightarrow \frac{3}{11} \cdot 0 + \frac{5}{11} \cdot 0,970 + \frac{3}{11} \cdot 0 = \frac{97}{220} = \underline{0.4409}$$

$$ganho = 0.593 - 0.449 = \underline{0,544} //$$

Questão:

Considere a seguinte base de dados, em que se deseja avaliar se uma fruta é doce ou não (sim ou não):

Tipo	Cor	Cheiro	Doce
Maçã	Vermelha	Doce	Sim
Banana	Amarela	Doce	Sim
Limão	Verde	Ácido	Não
Laranja	Laranja	Doce	Sim
Abacaxi	Amarela	Ácido	Não
Maçã	Verde	Doce	Sim
Banana	Verde	Ácido	Não
Laranja	Verde	Ácido	Não

Usando **Naive Bayes**, qual a probabilidade de a fruta ser **doce** (Sim) ou **não doce** (Não), respectivamente, para o seguinte registro:

**Tipo = Banana, Cor = Amarela, Cheiro = Doce**

Obs: Apresente todos os cálculos (matriz de probabilidades para todos os atributos, cálculos das porcentagens, etc).

- a) 85,0% e 15,0%
- b) 60,5% e 39,5%
- c) 78,9% e 21,1%
- d) 90,0% e 10,0%

Tente resolver essa questão, e caso precise de ajuda com os cálculos do Naive Bayes, estarei por aqui!

Naive:

Tipo = banana

Cor = amarela

Chute = doce

$$\text{Sim: } \frac{9}{8} \cdot \frac{1}{9} \cdot \frac{1}{4} \cdot \frac{4}{4} = \frac{4}{128} = 0,03125$$

$$\text{NÃO: } \frac{5}{8} \cdot \frac{1}{9} \cdot \frac{1}{4} \cdot \frac{0}{4} = ?$$

### Questão:

Considere a seguinte base de dados, em que se deseja avaliar se uma fruta é doce ou não (Sim ou Não):

Tipo	Cor	Cheiro	Doce	
Maçã	Vermelha	Doce	Sim	1
Banana	Amarela	Doce	Sim	2
Limão	Verde	Ácido	Não	3
Laranja	Laranja	Doce	Sim	4
Abacaxi	Amarela	Ácido	Não	5
Maçã	Verde	Doce	Sim	6
Banana	Verde	Ácido	Não	7
Laranja	Amarela	Ácido	Não	8
Banana	Vermelha	Doce	Sim	9
Abacaxi	Vermelha	<del>Ácido</del> Doce	Não	10

Usando **Naive Bayes**, qual a probabilidade de a fruta ser **doce** (Sim) ou **não doce** (Não), respectivamente, para o seguinte registro:

**Tipo = Laranja, Cor = Amarela, Cheiro = Doce**

Obs: Apresente todos os cálculos (matriz de probabilidades para todos os atributos, cálculos das porcentagens, etc).

- a) 90,5% e 9,5%
- b) 85,0% e 15,0%
- c) 78,9% e 21,1%
- d) 60,5% e 39,5%

Naive:

Time = Longa

Cor = Amarela

Cheiro = Doce

SIM = 71,4%.

NÃO = 28,6%.

$$0,714 = \frac{0,02}{0,028}$$

$$\text{SIM: } \frac{5}{10} \cdot \frac{1}{5} \cdot \frac{1}{5} \cdot \frac{5}{5} = \frac{5}{250}$$

$$0,02 + 0,008 = 0,028$$

$$0,28\% \frac{0,008}{0,028}$$

$$\text{NÃO: } \frac{5}{10} \cdot \frac{1}{5} \cdot \frac{2}{5} \cdot \frac{1}{5} = \frac{2}{250}$$