

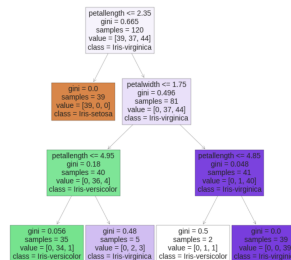
Lista 3

matéria	IA
feito	<input checked="" type="checkbox"/>

▼ Questão 1 - c

A figura abaixo mostra uma árvore de decisão construída por um algoritmo de aprendizado indutivo a partir de um conjunto de dados em que as instâncias são descritas por quatro atributos: Tamanho da Pétala, largura da Pétala, Tamanho da Sépala e Largura da Sépala.

Dado um objeto de classe desconhecida, essa árvore classifica o objeto nas classes: Iris_Setosa, Iris_Virginica e Iris_Versicolor. Esta árvore foi gerada com os hiperparâmetros (DecisionTreeClassifier(criterion='gini', max_depth=3)), usando a linguagem Python.



Com base nestas informações, qual as saídas da árvore para os seguintes registros de teste, respectivamente?

Registros de teste	Tamanho da Pétala	Largura da Pétala	Tamanho da Sépala	Largura da Sépala
Instância 1	3.46	0.87	2.45	1.78
Instância 2	1.67	1.89	0.78	1.32
Instância 3	2.56	2.34	2.45	1.78
Instância 4	6.67	2.34	2.45	1.78

- Iris_Virginica, Iris_Setosa, Iris_Versicolor, Iris_Virginica
- Iris_Virginica, Iris_Setosa, Iris_Virginica, Iris_Versicolor
- Iris_Versicolor, Iris_Setosa, Iris_Versicolor, Iris_Virginica**
- Iris_Setosa, Iris_Virginica, Iris_Virginica, Iris_Versicolor
- Iris_Versicolor, Iris_Setosa, Iris_Versicolor, Iris_Virginica

- Iris-versicular
- Iris-setosa
- Iris-versicular
- Iris virginica

▼ Questão 2 - c

Considerando a árvore da questão anterior, e as seguintes afirmações:

- Esta árvore possui 5 regras de classificação
- Das regras geradas, há apenas uma com cobertura por classe de 100%
- A menor cobertura por classe é de 6.8% e corresponde à classe Iris_Virginica

É correto o que se afirma em:

- apenas.

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

regras de classificação:

1. $\text{pentallength} \leq 2.35 \rightarrow \text{cobertura: } \frac{39}{39}$
2. $\text{pentallength} > 2.35$ e $\text{pentalwidth} \leq 1.75$ e $\text{pentallength} \leq 4.95 \rightarrow \text{cobertura: } \frac{34}{37}$
3. $\text{pentallength} > 2.35$ e $\text{pentalwidth} \leq 1.75$ e $\text{pentallength} > 4.95 \rightarrow \text{cobertura: } \frac{3}{44}$
4. $\text{pentallength} > 2.35$ e $\text{pentalwidth} > 1.75$ e $\text{pentallength} \leq 4.85 \rightarrow \text{cobertura: } \frac{1}{37}$
5. $\text{pentallength} > 2.35$ e $\text{pentalwidth} > 1.75$ e $\text{pentallength} > 4.85 \rightarrow \text{cobertura: } \frac{39}{44}$

▼ Questão 03

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

		Foi classificado como			
		A	B	C	D
Era da classe	A	10	4	2	1
	B	1	15	2	0
	C	2	3	20	5
	D	4	1	2	50

Quais os valores para as métricas abaixo para cada uma das classes A, B, C e D?

	precisão	recall	F1Score	TVP	TFN	TFP	TVN
A	89.8%	58.8%	71.0%	10/17	7/17	7/105	98/105
B	91.5%	83.3%	87.2%	15/18	3/18	8/104	96/104
C	91.1%	60.0%	72.3%	20/30	10/30	6/92	86/92
D	90.4%	87.7%	89.0%	50/57	7/57	6/65	59/65

▼ Questão 04

Considerando a base de dados abaixo. Mostre o primeiro e o segundo nível da árvore gerada

História do crédito	Dívida	Garantias	Renda Anual	Risco
Ruim	Alta	Nenhuma	<15000	Alto
Desconhecida	Alta	Nenhuma	>=15000 a <=35000	Alto
Desconhecida	Baixa	Nenhuma	>=15000 a <=35000	Moderado
Desconhecida	Baixa	Nenhuma	>35000	Alto
Desconhecida	Baixa	Nenhuma	>35000	Baixo
Desconhecida	Baixa	Adequada	>35000	Baixo
Ruim	Baixa	Nenhuma	<15000	Alto
Ruim	Baixa	Adequada	>35000	Moderado
Boa	Baixa	Nenhuma	>35000	Baixo
Boa	Alta	Adequada	>35000	Baixo
Boa	Alta	Nenhuma	<15000	Alto
Boa	Alta	Nenhuma	>=15000 a <=35000	Moderado
Boa	Alta	Nenhuma	>35000	Baixo
Ruim	Alta	Nenhuma	>=15000 a <=35000	Alto

▼ Primeiro nível → renda anual

$$\text{Entropia de classe} \rightarrow \left(\frac{6}{14}, \frac{3}{14}, \frac{5}{14} \right) = 0.52 + 0.47 + 0.54 = 1.53$$

ganho História do crédito

$$\text{entropia} \rightarrow 0.23 + 0.489 + 0.543 = 1.262$$

$$\text{ganho} = 1.53 - 1.262 = 0.268$$

ganho Dívida

$$\text{entropia} \rightarrow 0.689 + 0.778 = 1.467$$

$$\text{ganho} = 1.53 - 1.467 = 0.063$$

ganho Garantias

$$\text{entropia} \rightarrow 1.127 + 0.196 = 1.323$$

$$\text{ganho} = 1.53 - 1.323 = 0.207$$

ganho Renda Anual

$$\text{entropia} \rightarrow 0.285 + 0.574 = 0.859$$

$$\text{ganho} = 1.53 - 0.859 = 0.671$$

▼ Segundo nível → história de crédito

$$\text{Entropia de classe} \rightarrow \left(\frac{3}{11}, \frac{3}{11}, \frac{4}{11}\right) = 1.54$$

ganho História do crédito

$$\text{entropia} \rightarrow 0.691 + 0.188 + 0.295 = 1.174$$

$$\text{ganho} = 1.54 - 1.177 = 0.366$$

ganho Dívida

$$\text{entropia} \rightarrow 0.691 + 0.795 = 1.486$$

$$\text{ganho} = 1.54 - 1.486 = 0.054$$

ganho Garantias

$$\text{entropia} \rightarrow 1.135 + 0.25 = 1.385$$

$$\text{ganho} = 1.54 - 1.385 = 0.155$$

▼ Questão 05

Investigue a biblioteca <https://pypi.org/project/chefboost/> E implemente os algoritmos ID3 e C45 para gerar a árvore da base de dados da questão anterior.

https://colab.research.google.com/drive/1UygSsTmFP_S0epqGGjQvabuAynzrjlo3?usp=sharing

▼ Questão 06

Investigue como é o funcionamento do algoritmo CART. Mostre todos os cálculos necessários para a geração da árvore.

O algoritmo CART começa na raiz da árvore e efetua uma divisão, criando dois nós no próximo nível da árvore.

$$R_1 = \{X \in \mathbb{R}^p : X_j \leq t\} \text{ e } R_2 = \{X \in \mathbb{R}^p : X_j > t\}$$

Depois disso, descemos para o primeiro nível da árvore e repetimos o procedimento para os dois nós que foram criados.

Continuamos da mesma maneira nos níveis seguintes.

Em cada etapa, escolhemos a divisão que produz a maior queda no erro de classificação.

$$(j, t) = \min_{(j, t)} ((1 - p_{c_1}(R_1)) + (1 - p_{c_2}(R_2)))$$

O algoritmo CART cresce uma árvore alta e poda alguns dos seus ramos no final do processo.

▼ Questão 07

Utilizando a biblioteca acima, rode o CART e veja a árvore gerada. Compare as 3 árvores geradas a partir da mesma base de dados.

https://colab.research.google.com/drive/1UygSsTmFP_S0epqGGjQvabuAynzrjlo3?usp=sharing