

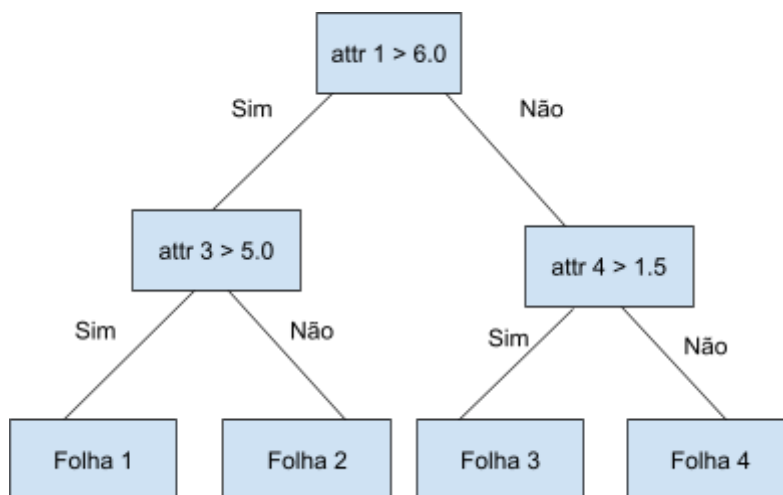
Aprendizagem Automática

Teste 1 (exemplos de questões)

1. Considere o seguinte conjunto de dados, onde cada exemplo é caracterizado por 4 atributos e pertence a uma de 3 classes. Se o conjunto for fornecido ao algoritmo dos K-vizinhos-mais-próximos, com $K=3$ e a distância euclidiana como métrica de distância, a que classe pertence o exemplo $\{5.9, 3.0, 5.1, 1.3\}$? Justifique a resposta apresentando os cálculos realizados.

ID	Atr1	Atr2	Atr3	Atr4	Classe
1	4.6	3.2	1.4	0.2	Iris-setosa
2	5.3	3.7	1.5	0.2	Iris-setosa
3	5.0	3.3	1.4	0.2	Iris-setosa
4	6.2	2.9	4.3	1.3	Iris-versicolor
5	5.1	2.5	3.0	1.1	Iris-versicolor
6	5.7	2.8	4.1	1.3	Iris-versicolor
7	6.3	2.5	5.0	1.9	Iris-virginica
8	6.5	3.0	5.2	2.0	Iris-virginica
9	6.2	3.4	5.4	2.3	Iris-virginica

2. Considere a seguinte árvore de decisão e o conjunto de dados da questão anterior.



- Indique o grau de pureza na Folha 4 considerando a entropia como função de impureza.
- Indique a distribuição das classes das instâncias classificadas na Folha 4.
- Indique qual deverá ser o resultado da árvore de decisão na Folha 4.

3. Considere o seguinte conjunto de dados, onde cada exemplo é caracterizado por 3 atributos, tendo associada uma classe. Se o conjunto for fornecido ao algoritmo Naive Bayes com estimador suavizado de Laplace qual a classificação para o exemplo {low, T, teen}? Justifique a resposta, apresentando os cálculos realizados

ID	income	student	age	buy-phone
1	low	T	teen	no
2	low	T	teen	yes
3	low	T	child	no
4	high	F	teen	yes
5	low	F	adult	no
6	low	F	teen	no
7	high	F	child	yes
8	high	F	adult	no
9	high	T	adult	yes

4. Classifique como Verdadeiro ou Falso. Justifique a resposta.
- Os dados de treino influenciam o modelo do classificador
 - Os dados de teste não influenciam o desempenho do classificador
 - Um modelo de classificação em funcionamento gera um conjunto de atributos
 - O algoritmo de aprendizagem gera um conjunto de resultados
 - O classificador é um modelo para previsão de números reais e depende dos dados de teste
 - Um classificador Naive Bayes exige que os atributos sejam correlacionados entre si
 - Uma tarefa de classificação é uma tarefa de aprendizagem não supervisionada

5. Considere os seguintes comandos:

```
a: print(" A exactidão é ", svc.score(x_train,y_train))
b: x_train, x_test, y_train, y_test = train_test_split(x_data, y_data)
c: svc = SVC()
d: print(" A exactidão é ", svc.score(x_test,y_test))
e: svc.fit(x_test,y_test)
f: x_data, y_data = load_data_XYZ()
g: svc.fit(x_train,y_train)
```

Selecione e ordene as instruções de modo apresentar o desempenho do modelo construído com o algoritmo de máquina de vetores de suporte para classificação aplicado ao conjunto de dados "XYZ".

6. Considere que os seguintes valores de exatidão foram obtidos utilizando o algoritmo de **Regressão Logística** com norma L1 e diferentes valores para o parâmetro C (parâmetro de regularização).

	C1	C2	C3
treino	0.84	0.93	1
teste	0.85	0.96	0.9

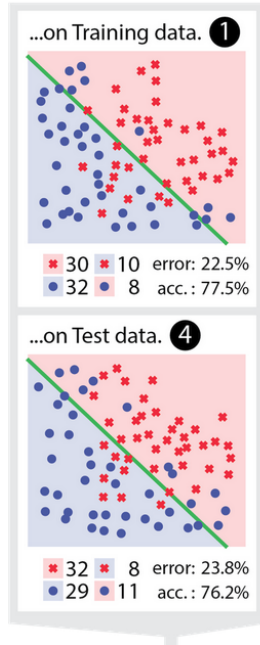
- Faça corresponder {C1, C2, C3} aos elementos do conjunto {0.1, 1, 100}. Justifique a resposta.
- Existem indícios de sobre-ajustamento em alguma configuração? Qual? Justifique a resposta.

7. Considere a seguinte matriz de confusão, e a classe "maligno" como classe positiva.

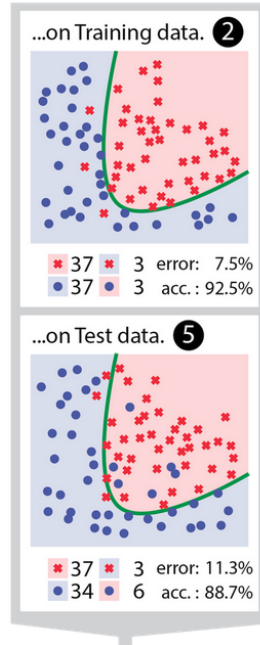
real\classificador	maligno	benigno
maligno	43	10
benigno	0	90

- Qual o nº total de exemplos positivos?
 - Qual o número de Falsos Negativos?
 - Indique o valor da exatidão.
8. Seja VP e FP o nº de verdadeiros positivos e falsos positivos, respectivamente. Sabendo que a precisão é calculada pela fórmula $VP/(VP+FP)$, qual das seguintes definições corresponde à precisão?
- Proporção de verdadeiros positivos entre os classificados como positivos.
 - Proporção de instâncias positivas classificadas corretamente.
 - Proporção de instâncias negativas classificadas corretamente
 - Proporção de instâncias corretamente classificadas
9. Considere um conjunto de dados com duas classes {vermelho, azul}. A imagem seguinte apresenta a fronteira de decisão dos modelos gerados por um algoritmo de aprendizagem sobre um conjunto de dados e apresenta o erro associado aos conjuntos de treino e teste.
- Que modelo poderá estar em situação de sobre-ajustamento?
 - Que modelo poderá estar em situação de sub-ajustamento?
 - Que modelo escolheria para o classificador?

Model 1...



Model 2...



Model 3...

