



Aula Laboratório

Aprendizado Indutivo

Introdução

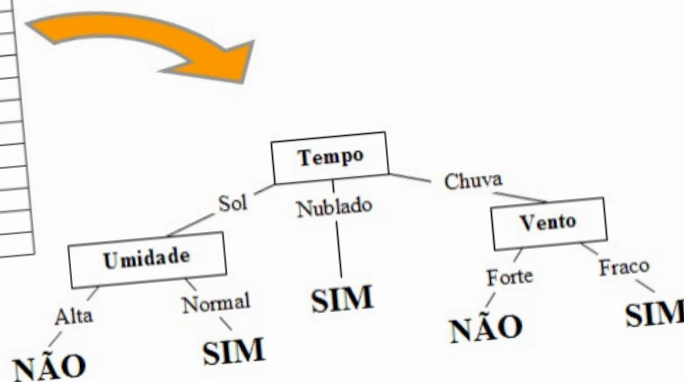
Inteligência Artificial e Robótica - CC 7711

Aprendizado

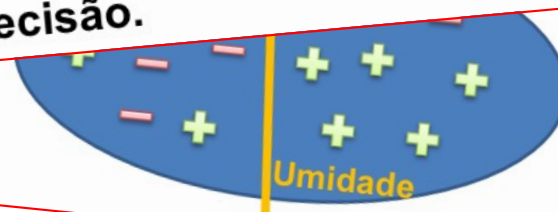
Aprendizado Indutivo

- Dado um conjunto de exemplos pré-classificados

Dia	Tempo	Temperatura	Umidade	Vento	Partida
1	Sol	Quente	Alta	Fraco	NÃO
2	Sol	Quente	Alta	Forte	NÃO
3	Nublado	Quente	Alta	Fraco	SIM
4	Chuva	Morno	Alta	Fraco	SIM
5	Chuva	Frio	Normal	Fraco	SIM
6	Chuva	Frio	Normal	Forte	NÃO
7	Nublado	Frio	Normal	Forte	SIM
8	Sol	Morno	Alta	Fraco	NÃO
9	Sol	Frio	Normal	Fraco	SIM
10	Chuva	Morno	Normal	Forte	SIM
11	Sol	Morno	Alta	Forte	SIM
12	Nublado	Quente	Normal	Fraco	SIM
13	Nublado	Quente	Normal	Forte	NÃO
14	Chuva	Morno	Alta	Forte	NÃO



- ... gerar uma árvore de decisão.

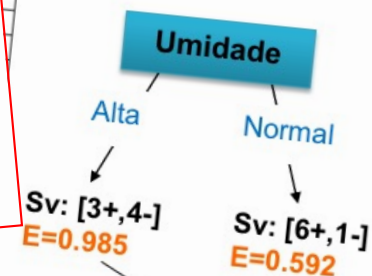


de decisão ?

$$S = [D1, D2, \dots, D14]$$

$$[9+, 5-]$$

$$\text{Entropia} = 0.940$$



$$\text{Ganho}(S, \text{Umidade})$$

$$= 0.940 - (7/14) * 0.985 - (7/14) * 0.592$$

$$= 0.151$$

Introdução

Inteligência Artificial e Robótica - CC 7711

Aprendizado

Aprendizado Indutivo

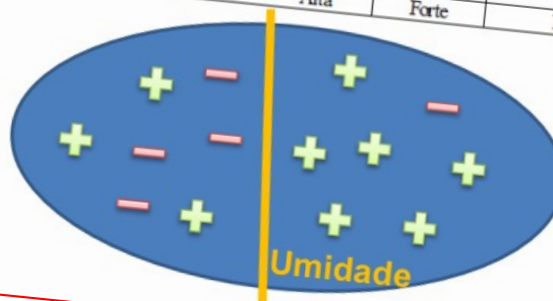
■ Dado um conjunto de exemplos

Dia	Tempo	Temperatura	Umidade	Vento	Partida
1	Sol	Quente	Alta	Fraco	NÃO
2	Sol	Quente	Alta	Forte	NÃO
3	Nublado	Quente	Alta	Fraco	SIM
4	Chuva	Morno	Alta	Fraco	SIM
5	Chuva	Frio	Normal	Fraco	NÃO
6	Chuva	Frio	Normal	Forte	SIM
7	Nublado	Frio	Normal	Forte	NÃO
8	Sol	Morno	Alta	Fraco	SIM
9	Sol	Frio	Normal	Fraco	SIM
10	Chuva	Morno	Normal	Forte	SIM
11	Sol	Morno	Alta	Forte	SIM
12	Nublado	Morno	Normal	Fraco	NÃO
13	Nublado	Quente	Normal	Forte	SIM
14	Chuva	Morno	Alta	Forte	NÃO

■ ... gerar uma árvore de

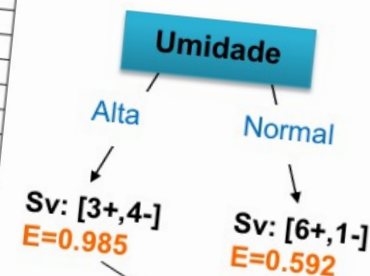
Como gerar uma árvore de decisão ?

Dia	Tempo	Temperatura	Umidade	Vento	Partida
1	Sol	Quente	Alta	Fraco	NÃO
2	Sol	Quente	Alta	Forte	NÃO
3	Nublado	Quente	Alta	Fraco	SIM
4	Chuva	Morno	Alta	Fraco	SIM
5	Chuva	Frio	Normal	Fraco	NÃO
6	Chuva	Frio	Normal	Forte	SIM
7	Nublado	Frio	Normal	Forte	NÃO
8	Sol	Morno	Alta	Fraco	SIM
9	Sol	Frio	Normal	Fraco	SIM
10	Chuva	Morno	Normal	Forte	SIM
11	Sol	Morno	Alta	Forte	SIM
12	Nublado	Morno	Normal	Fraco	NÃO
13	Nublado	Quente	Normal	Forte	SIM
14	Chuva	Morno	Alta	Forte	NÃO



$S = [D1, D2, \dots, D14]$
[9+, 5-]

Entropia = 0.940



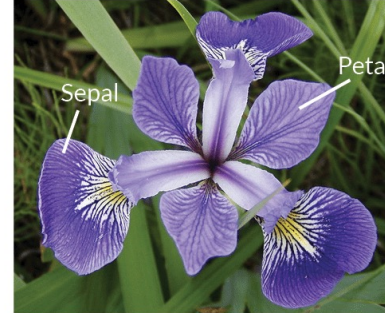
Ganho(S, Umidade)
= 0.940 - (7/14) * 0.985 - (7/14) * 0.592
= 0.151

Dataset IRIS

Conjunto de dados clássico para iniciantes

Objetivo:
classificar as flores considerando as
características de Sépala e Pétala

https://s3.amazonaws.com/assets.datacamp.com/blog_assets/Machine+Learning+R/iris-machinelearning.png



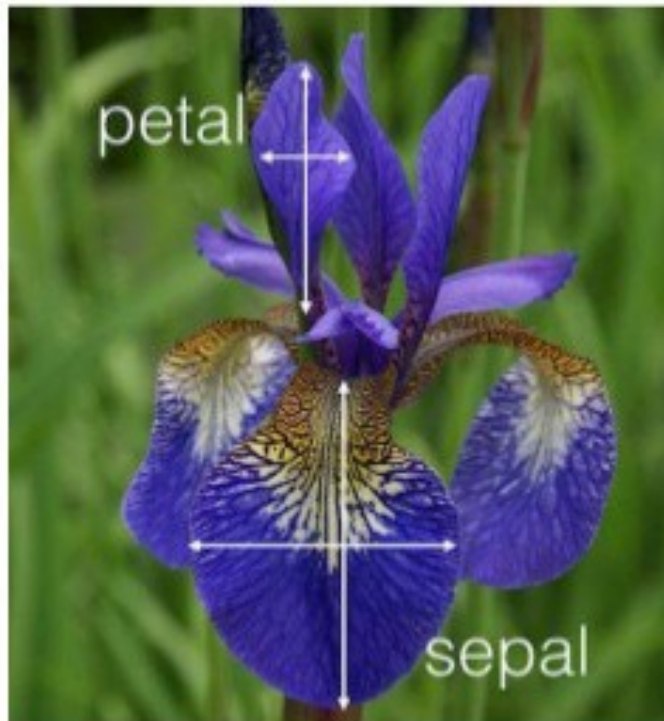
Iris Versicolor



Iris Setosa



Iris Virginica



Training / test data

Features

Labels

Sepal length	Sepal width	Petal length	Petal width	Species
5.1	3.5	1.4	0.2	Iris setosa
4.9	3.0	1.4	0.2	Iris setosa
7.0	3.2	4.7	1.4	Iris versicolor
6.4	3.2	4.5	1.5	Iris versicolor
6.3	3.3	6.0	2.5	Iris virginica
5.8	3.3	6.0	2.5	Iris virginica

Decision Tree Classifier

```
#Training the decision tree classifier. |
clf_entropy.fit(X_train, y_train)

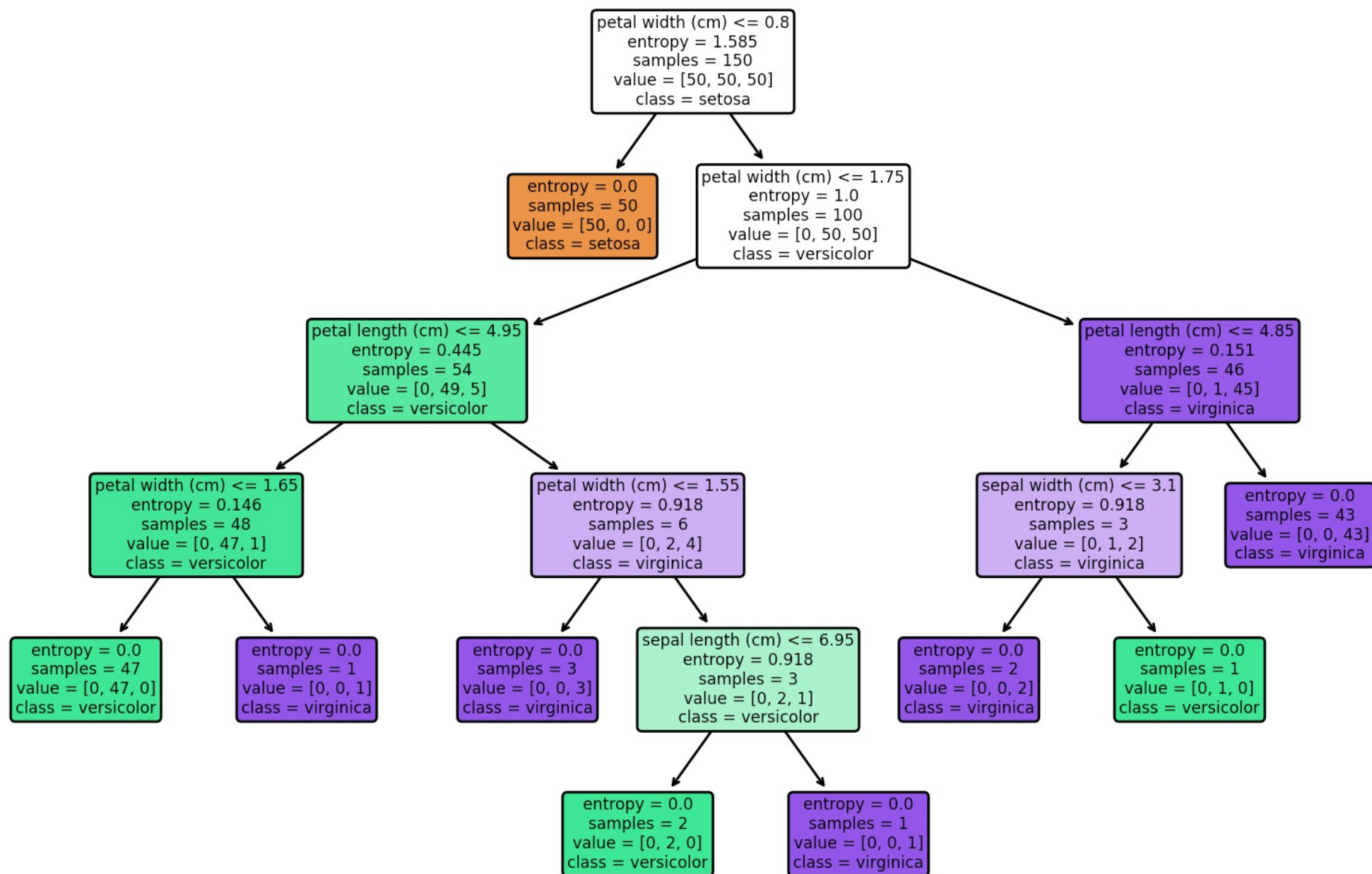
DecisionTreeClassifier(ccp_alpha=0.0, class_weight=None, criterion='entropy',
                      max_depth=None, max_features=None, max_leaf_nodes=None,
                      min_impurity_decrease=0.0, min_impurity_split=None,
                      min_samples_leaf=1, min_samples_split=2,
                      min_weight_fraction_leaf=0.0, presort='deprecated',
                      random_state=None, splitter='best')
```

```
data = load_iris()
features = data.data
target = data.target
```

```
Arvore = DecisionTreeClassifier(criterion = 'entropy').fit(features, target)
```

```
plt.figure(figsize=(10,6.5))
tree.plot_tree(Arvore, feature_names=data.feature_names, class_names=data.target_names,
               filled=True, rounded=True)
plt.show()
```

```
# só lembrando: ['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal width (cm)']
Amostra = [[1.9, 3, 3.4, 0.2]]
class_Amostra = Arvore.predict(Amostra)
```



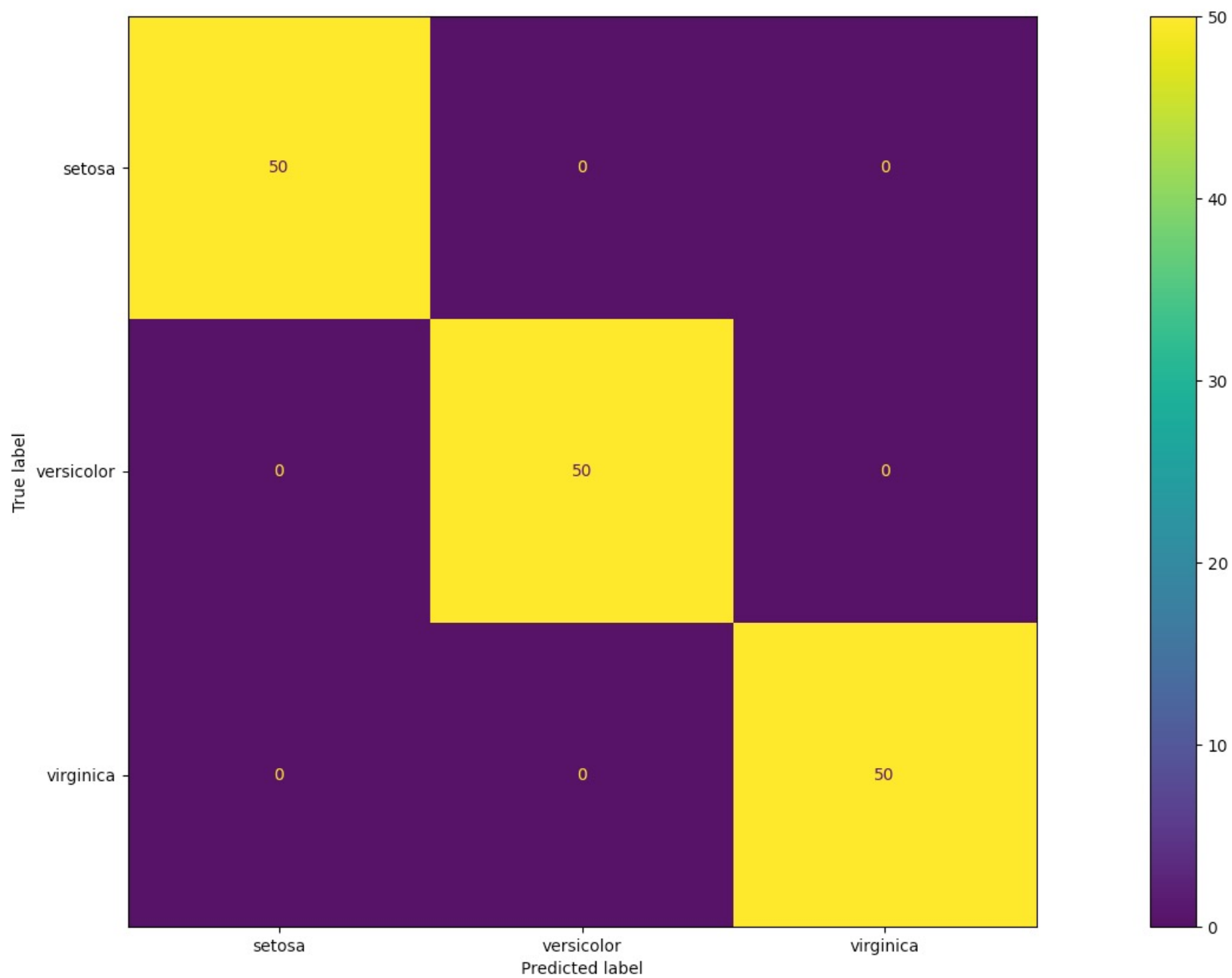
Detalhando a matriz de confusão

=== Confusion Matrix ===

a	b	<-- classified as
43	0	a = Aprovado
1	160	b = Reprovado

A **matriz de confusão** é uma tabela que permite a visualização do desempenho de um algoritmo de classificação

1. Neste exemplo, temos duas classificações possíveis (indicadas na terceira coluna):
(a)Aprovado e (b)Reprovado.
2. A primeira linha (a b ← classified as) identifica nas colunas a saída do algoritmo.
3. Lendo os valores da tabela:
 - 43 -> Classificado como a pelo algoritmo (primeira coluna) e no arquivo ARRF também está como a (primeira linha). Ou seja, o algoritmo classificou corretamente!
 - 1 -> Classificado como a pelo algoritmo (primeira coluna) porém, no arquivo ARRF está como b (segunda linha). Ou seja, classificou erroneamente um b como a!
 - 160 -> Classificado como b pelo algoritmo (segunda coluna) e no arquivo ARRF também está como b (segunda linha). Ou seja, o algoritmo classificou corretamente!
 - 0 -> Nenhum registro foi classificado como b pelo algoritmo (segunda coluna) quando no arquivo ARRF estava como a (primeira linha). Neste caso, o algoritmo teria classificado erroneamente um a como b!





Arquivos ARFF

- Attribute-Relation File Format
- @relation <relation-name>
 - relation-name é uma string que vai definir o título da análise
- @attribute <name> <type>
 - Tipos: numeric, integer, string, <lista>
 - Lista: conjunto de valores → Ex.: {True,False} ou {Vermelho,Amarelo,Verde}
- @data
 - Lista de dados, separados por virgula



Arquivos ARFF - Exemplo

@RELATION CriterioMedia

@ATTRIBUTE Nota integer

@ATTRIBUTE PercFalta integer

@ATTRIBUTE resultado {Aprovado,Reprovado}

@DATA

9,5,Aprovado

0,5,Reprovado

7,12,Aprovado

0,28,Reprovado



Atividade Laboratório

- Utilizar os arquivos ARFF fornecidos (na descrição da atividade no MOODLE!)
- Crie seu próprio arquivo ARFF...
 - ☐ Sem idéias? Utilize o exemplo da aula (Partida de Tennis!)
 - ☐ Ou utilize uma base pronta: <https://waikato.github.io/weka-wiki/datasets/>
 - ☐ Cuidado... é necessário explicar a base de dados
- No relatório, além de responder as questões acima, inclua todas as árvores encontradas e a matriz de confusão e discuta os resultados, sempre comparando os problemas e os algoritmos utilizados!
 - ☐ Não é necessário incluir os arquivos ARFF