

Inteligência Artificial

Tópico 04 - Parte 03

Aprendizado de Máquina - Experimentos e Weka

Profa. Dra. Priscila Tiemi Maeda Saito
✉ priscilasaito@ufscar.br

Roteiro

- 1 Aprendizado de Máquina
 - Experimentos

Aprendizado de Máquina

- É uma área de IA cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado, bem como a construção de sistemas capazes de adquirir conhecimento de forma automática

Aprendizado de Máquina

- Duas importantes áreas do AM são
 - ▶ classificação (aprendizado supervisionado)
 - ▶ agrupamento (aprendizado não-supervisionado)

Aprendizado de Máquina

- Um artigo típico de AM propõe
 - ▶ um novo algoritmo
 - ▶ uma modificação em um algoritmo existente
 - ▶ um novo pré-processamento
 - ▶ um novo pós-processamento

Hipótese implícita

a nova proposta produz uma melhora no desempenho em comparação aos algoritmos da literatura

Aprendizado de Máquina

- A avaliação em um artigo típico de AM
 - ▶ usa um conjunto de datasets de teste
 - ▶ usa medidas de avaliação apropriadas
 - ▶ estabelece as conclusões com base nos datasets e nas medidas de avaliação

Teorema “no free lunch”

Lembrando que: não é possível encontrar um algoritmo que seja o melhor para todos os problemas

Aprendizado de Máquina

- Como podemos verificar a hipótese de que o desempenho será melhorado com a nova proposta?

Aprendizado de Máquina

- Questões a serem tratadas?
 - ▶ dados dois algoritmos de aprendizado A e B e um dataset, qual é o melhor algoritmo?
 - ▶ dados vários algoritmos de aprendizado e vários datasets, qual deles é o melhor?
 - ★ há diferença entre os resultados dos algoritmos?

Aprendizado de Máquina

Questões a serem tratadas

	Alg. 1	Alg. 2	Alg. 3	Alg. 4	Alg. 5	Alg. 6	Alg. 7
aud	25.3	76.0	68.4	69.6	79.0	81.2	57.7
aus	55.5	81.9	85.4	77.5	85.2	83.3	85.7
bal	45.0	76.2	87.2	90.4	78.5	81.9	79.8
bpa	58.0	63.5	60.6	54.3	65.8	65.8	68.2
bps	51.6	83.2	82.8	78.6	80.1	79.0	83.3
bre	65.5	96.0	96.7	96.0	95.4	95.3	96.0
cmc	42.7	44.4	46.8	50.6	52.1	49.8	52.3
gls	34.6	66.3	66.4	47.6	65.8	69.0	72.6
h-c	54.5	77.4	83.2	83.6	73.6	77.9	79.9
hep	79.3	79.9	80.8	83.2	78.9	80.0	83.2
irs	33.3	95.3	95.3	94.7	95.3	95.3	94.7
krk	52.2	89.4	94.9	87.0	98.3	98.4	98.6
lab	65.4	81.1	92.1	95.2	73.3	73.9	75.4
led	10.5	62.4	75.0	74.9	74.9	75.1	74.8
lym	55.0	83.3	83.6	85.6	77.0	71.5	79.0
mmg	56.0	63.0	65.3	64.7	64.8	61.9	63.4
mus	51.8	100.0	100.0	96.4	100.0	100.0	99.8
mux	49.9	78.6	99.8	61.9	99.9	100.0	100.0
pmi	65.1	70.3	73.9	75.4	73.1	72.6	76.0
prt	24.9	34.5	42.5	50.8	41.6	39.8	43.7
seg	14.3	97.4	96.1	80.1	97.2	96.8	96.1
sick	93.8	96.1	96.3	93.3	98.4	97.0	96.7
soyb	13.5	89.5	90.3	92.8	91.4	90.3	76.2
tao	49.8	96.1	96.0	80.8	95.1	93.6	88.4
thy	19.5	68.1	65.1	80.6	92.1	92.1	86.3
veh	25.1	69.4	69.7	46.2	73.6	72.6	72.2
vote	61.4	92.4	92.6	90.1	96.3	96.5	95.4
vow	9.1	99.1	96.6	65.3	80.7	78.3	87.6
wne	39.8	95.6	96.8	97.8	94.6	92.9	96.3
zoo	41.7	94.6	92.5	95.4	91.6	92.5	92.6
Avg	44.8	80.0	82.4	78.0	82.1	81.8	81.7

- Grandes variações na acurácia de diferentes classificadores!
- O algoritmo 3 é o melhor porque ele obtém a melhor média?

Aprendizado de Máquina

Questões a serem tratadas

	Alg. 1	Alg. 2	Alg. 3	Alg. 4	Alg. 5	Alg. 6	Alg. 7
aud	25.3	76.0	68.4	69.6	79.0	81.2	57.7
aus	55.5	81.9	85.4	77.5	85.2	83.3	85.7
bal	45.0	76.2	87.2	90.4	78.5	81.9	79.8
bpa	58.0	63.5	60.6	54.3	65.8	65.8	68.2
bps	51.6	83.2	82.8	78.6	80.1	79.0	83.3
bre	65.5	96.0	96.7	96.0	95.4	95.3	96.0
cmc	42.7	44.4	46.8	50.6	52.1	49.8	52.3
gls	34.6	66.3	66.4	47.6	65.8	69.0	72.6
h-c	54.5	77.4	83.2	83.6	73.6	77.9	79.9
hep	79.3	79.9	80.8	83.2	78.9	80.0	83.2
irs	33.3	95.3	95.3	94.7	95.3	95.3	94.7
krk	52.2	89.4	94.9	87.0	98.3	98.4	98.6
lab	65.4	81.1	92.1	95.2	73.3	73.9	75.4
led	10.5	62.4	75.0	74.9	74.9	75.1	74.8
lym	55.0	83.3	83.6	85.6	77.0	71.5	79.0
mmg	56.0	63.0	65.3	64.7	64.8	61.9	63.4
mus	51.8	100.0	100.0	96.4	100.0	100.0	99.8
mux	49.9	78.6	99.8	61.9	99.9	100.0	100.0
pmi	65.1	70.3	73.9	75.4	73.1	72.6	76.0
prt	24.9	34.5	42.5	50.8	41.6	39.8	43.7
seg	14.3	97.4	96.1	80.1	97.2	96.8	96.1
sick	93.8	96.1	96.3	93.3	98.4	97.0	96.7
soyb	13.5	89.5	90.3	92.8	91.4	90.3	76.2
tao	49.8	96.1	96.0	80.8	95.1	93.6	88.4
thy	19.5	68.1	65.1	80.6	92.1	92.1	86.3
veh	25.1	69.4	69.7	46.2	73.6	72.6	72.2
vote	61.4	92.4	92.6	90.1	96.3	96.5	95.4
vow	9.1	99.1	96.6	65.3	80.7	78.3	87.6
wne	39.8	95.6	96.8	97.8	94.6	92.9	96.3
zoo	41.7	94.6	92.5	95.4	91.6	92.5	92.6
Avg	44.8	80.0	82.4	78.0	82.1	81.8	81.7

- O algoritmo 4 é o vencedor em 8 problemas com média de 78.0
- O algoritmo 2 é o vencedor para 4 problemas com média 80.0
- Qual o melhor entre eles?

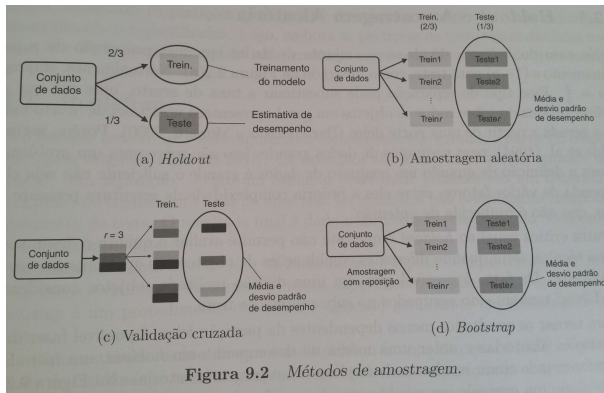
Aprendizado de Máquina

- Amostragem

- ▶ Resubstitution
- ▶ Holdout e amostragem aleatória
- ▶ Validação cruzada
 - ★ k-fold cross-validation
 - ★ k-fold stratified cross-validation
 - ★ Leave-one-out
- ▶ Bootstrap

Aprendizado de Máquina

• Amostragem



Amostragem - Substitution

- Utiliza o mesmo conjunto de amostras no treinamento e avaliação do modelo
- Produz estimativas otimistas, uma vez que todos os algoritmos de AM tentam melhorar de alguma forma o seu desempenho preditivo nessas amostras durante a fase indutiva
- Deve-se utilizar métodos de amostragem alternativos para obter estimativas de desempenho preditivo mais confiáveis, definindo os subconjuntos de treinamento e de teste

Amostragem - Holdout

- Divide-se o conjunto de dados em uma proporção de p para treinamento e $(1 - p)$ para teste
- Normalmente, emprega-se $p = \frac{2}{3}$
- Amostragem aleatória (random subsampling)
 - ▶ realiza diversas partições aleatórias e obtém média de desempenho em holdout
 - ▶ torna os resultados menos dependentes da partição realizada

Amostragem - Validação Cruzada

- k-fold cross-validation
 - ▶ conjunto de amostras é dividido em k subconjuntos de tamanho \approx igual
 - ▶ as amostras de $k-1$ partições são utilizadas no treinamento do modelo
 - ▶ o modelo é testado na partição restante
 - ▶ o processo é repetido k vezes, utilizando em cada iteração uma partição diferente para teste
 - ▶ desempenho final do modelo é dado pela média dos desempenhos observados sobre cada subconjunto de teste

Amostragem - Validação Cruzada

- k-fold cross-validation estratificado
 - ▶ mantém em cada partição a proporção de amostras de cada classe semelhante à proporção contida no conjunto de dados total
 - ▶ exemplo:
 - ★ se o conjunto de dados original tem 20% das amostras na classe c_1 e 80% na classe c_2
 - ★ cada partição também procura manter essa proporção, apresentando 20% de suas amostras na classe c_1 e 80% na classe c_2

Amostragem - Validação Cruzada

- Leave-one-out

- ▶ caso em que $k = n$, em que n representa o número de amostras disponíveis
- ▶ a cada iteração, uma amostra é separada para teste, enquanto as $k-1$ amostras restantes são utilizadas no treinamento do modelo
- ▶ o desempenho é dado pela soma dos desempenhos verificados para cada amostra de teste individual
- ▶ produz uma estimativa mais fiel do desempenho preditivo do modelo
- ▶ computacionalmente caro e geralmente aplicado somente em amostras de dados pequenas

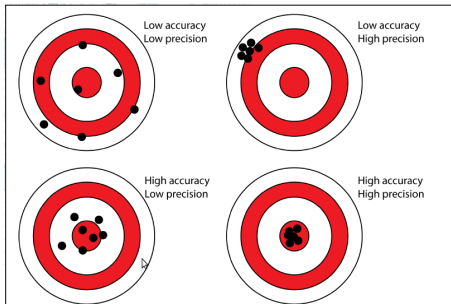
Amostragem - Validação Cruzada

• Bootstrap

- ▶ a partir do conjunto de dados original, k subconjuntos de treinamento são gerados
- ▶ as amostras são selecionadas aleatoriamente do conjunto de dados original, com reposição
 - ★ uma amostra pode estar presente em um determinado subconjunto de treinamento mais de uma vez
- ▶ as amostras não selecionadas compõem os subconjuntos de teste
- ▶ o resultado final é dado pela média do desempenho observado em cada subconjunto de teste
- ▶ normalmente, adota-se $k \geq 100$
 - ★ ideia → repetir o experimento um número alto de vezes e estimar o desempenho nesses experimentos replicados
- ▶ processo custoso e aplicado em amostras de dados pequenas

Aprendizado de Máquina

- Acurácia e precisão



Matriz de Confusão

- Matriz que permite visualizar as principais confusões do sistema
- Considere um sistema com 3 classes, 100 exemplos por classe

100% de classificação

	c_1	c_2	c_3
c_1	100		
c_2		100	
c_3			100

Erros de classificação

	c_1	c_2	c_3
c_1	90	10	
c_2		100	
c_3	5		95

- 10 → 10 exemplos de C_1 foram classificados como C_2

Aprendizado de Máquina

Matriz de Confusão

		Predicted	
		Negative	Positive
Actual	Negative	a	b
	Positive	c	d

- **a**: número de predições corretas que uma instância é negativa (**TN**)
- **b**: número de predições incorretas que uma instância é positiva (**FP**)
- **c**: número de predições incorretas que uma instância é negativa (**FN**)
- **d**: número de predições corretas que uma instância é positiva (**TP**)

Medidas

$$\text{Acurácia} \rightarrow \text{Acc} = \frac{(a+d)}{a+b+c+d}$$

$$\text{Recall (R) ou true positive (TP) rate} \rightarrow \text{TPR} = \frac{d}{(c+d)}$$

$$\text{false positive (FP) rate} \rightarrow \text{FPR} = \frac{b}{(a+b)}$$

$$\text{true negative (TN) rate} \rightarrow \text{TNR} = \frac{a}{(a+b)}$$

$$\text{false negative (FN) rate} \rightarrow \text{FNR} = \frac{c}{(c+d)}$$

$$\text{Precision} \rightarrow P = \frac{d}{(b+d)}$$

$$\text{F-score} = \frac{2(R \cdot P)}{(R+P)}$$

Aprendizado de Máquina

- Eficácia x Eficiência
- Outra métrica é também importante:
 - ▶ quão fácil é usar a técnica em questão, ou quantos graus de liberdade é preciso controlar para se chegar no desempenho desejado

Aprendizado de Máquina

Weka

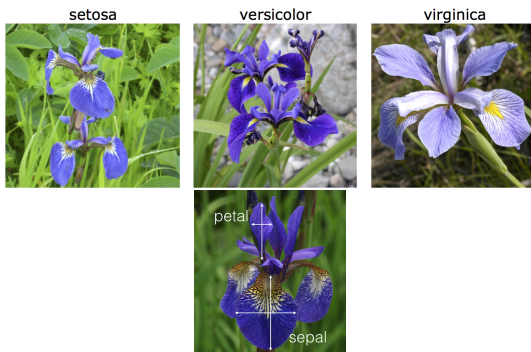
- Ferramenta e biblioteca para tarefas de aprendizado de máquina
- Possui grande quantidade de algoritmos de aprendizado de máquina
- Implementado na linguagem Java
- Versão weka-3-6-11
 - ▶ <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>
- Datasets
 - ▶ www.cs.waikato.ac.nz/ml/weka/index_datasets.html
- `java -jar weka.jar`



Aprendizado de Máquina

Weka

- Dataset Iris



Aprendizado de Máquina

Weka

- Formato do arquivo de treinamento/testes:
 - ▶ formato ARFF: Attribute-Relation File Format

Cabeçalho do arquivo

```
@relation (DatasetName)
@attribute (AtribName1) (AtribType1)
@attribute (AtribName2) (AtribType2)
.
.
.
@attribute (AtribNameN) (AtribTypeN)
@attribute (Class) {(Class1, Class2, ..., ClassN)}
@data (Atrib1), (Atrib2), ..., (AtribN), (Class)
.
.
.
```

Aprendizado de Máquina

Weka

- Formato do arquivo de treinamento/testes:

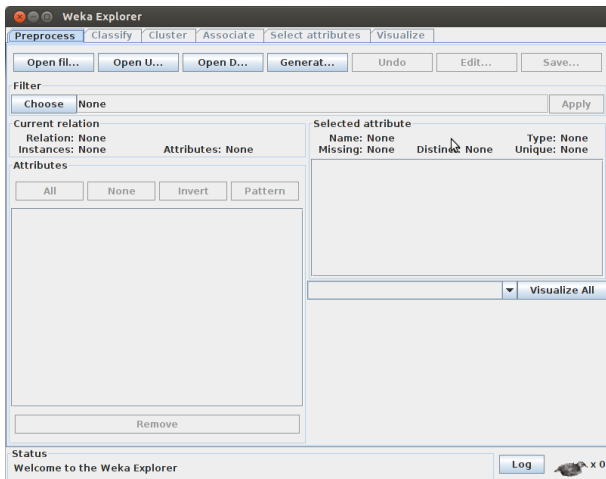
Exemplo

```
@relation Activity_Recognition
@attribute 'valor1' real
@attribute 'valor2' real
@attribute 'valor3' real
@attribute 'class' {0,1,2,3,4,5,6,7,8,9,10}
@data
0.24679, 0.210083, 0.0873606, 0
0.546452, 0.811992, 0.0163704, 1
0.745887, 0.114372, 0.0957822, 3
0.245887, 0.214372, 0.0857822, 0
```

Aprendizado de Máquina

Weka

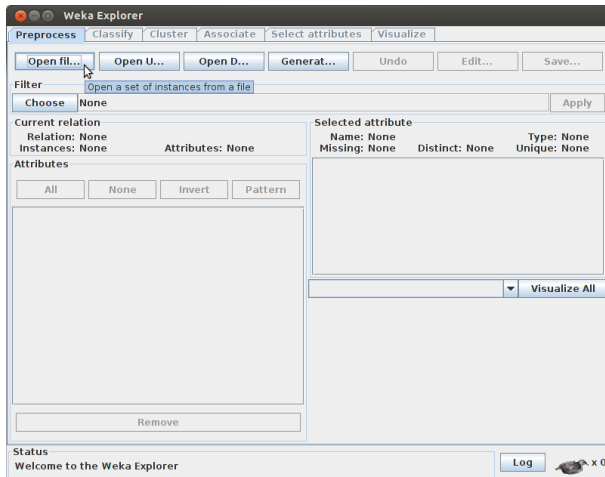
- Interface para testes



Aprendizado de Máquina

Weka

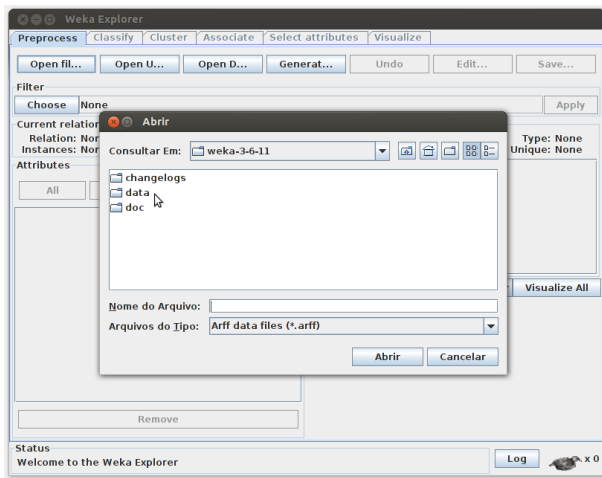
- Selecionando uma base de treinamento



Aprendizado de Máquina

Weka

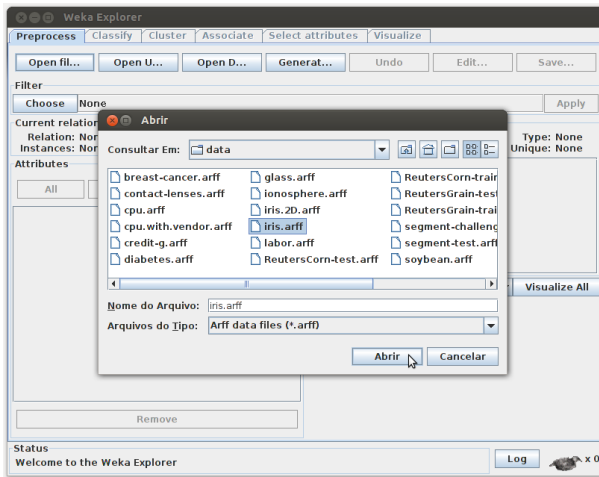
- Selecionando uma base de treinamento



Aprendizado de Máquina

Weka

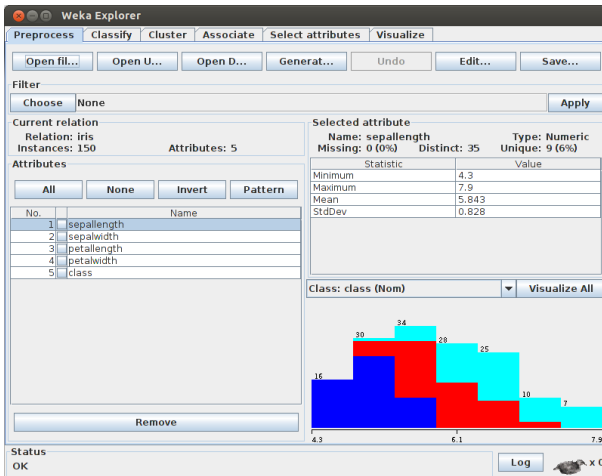
- Selecionando uma base de treinamento



Aprendizado de Máquina

Weka

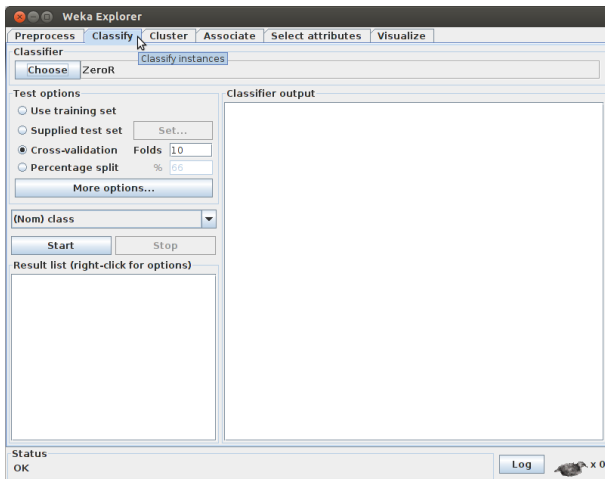
- Selecionando uma base de treinamento



Aprendizado de Máquina

Weka

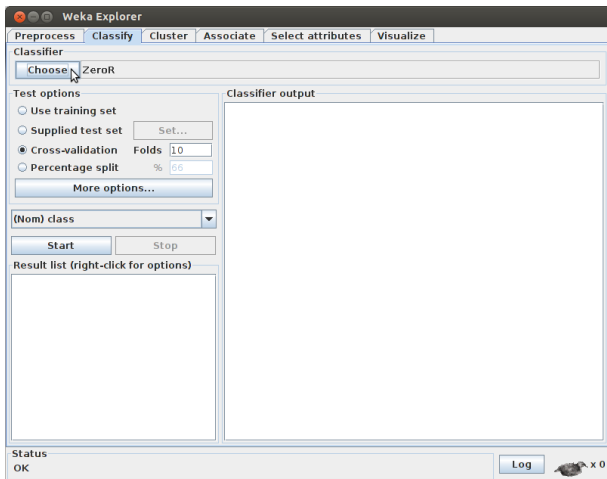
- Selecionando um algoritmo



Aprendizado de Máquina

Weka

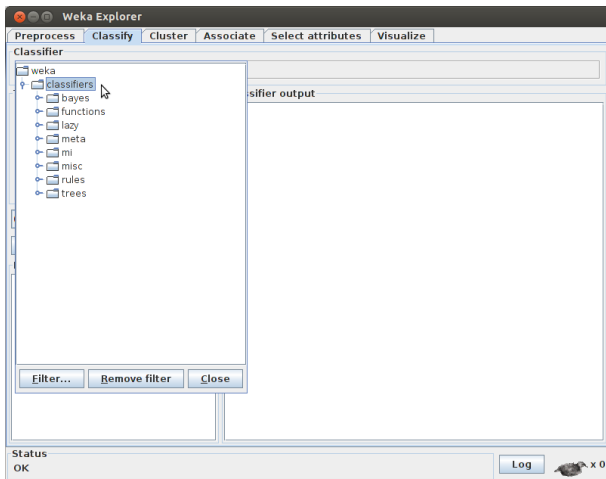
- Selecionando um algoritmo



Aprendizado de Máquina

Weka

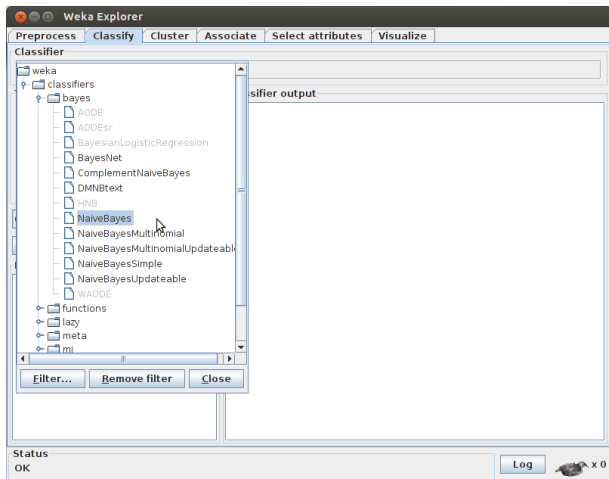
- Selecionando um algoritmo



Aprendizado de Máquina

Weka

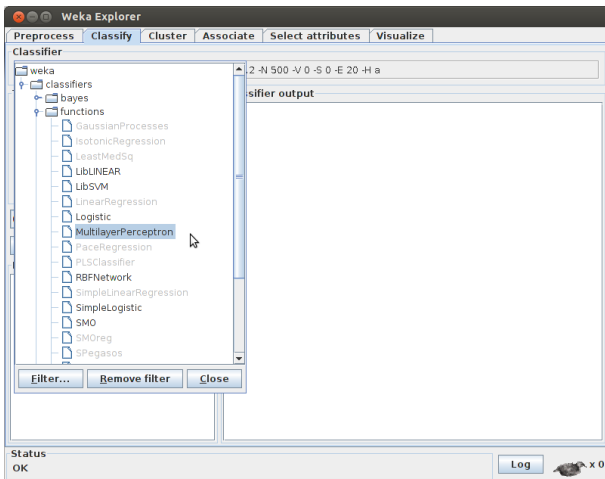
- Selecionando um algoritmo



Aprendizado de Máquina

Weka

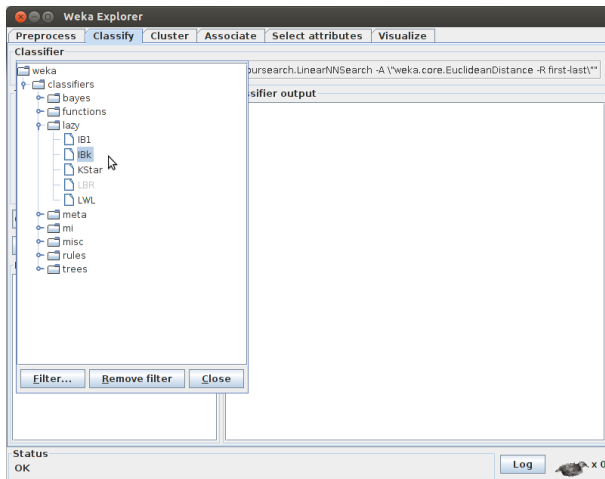
- Selecionando um algoritmo



Aprendizado de Máquina

Weka

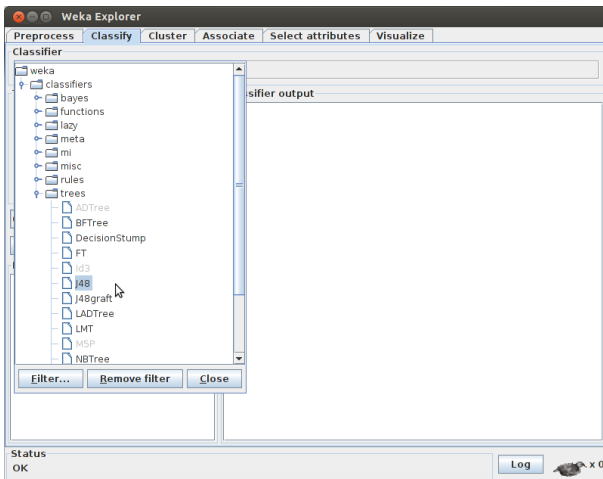
- Selecionando um algoritmo



Aprendizado de Máquina

Weka

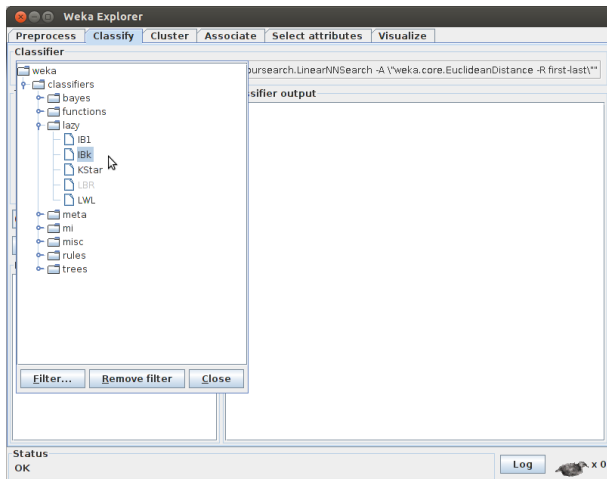
- Selecionando um algoritmo



Aprendizado de Máquina

Weka

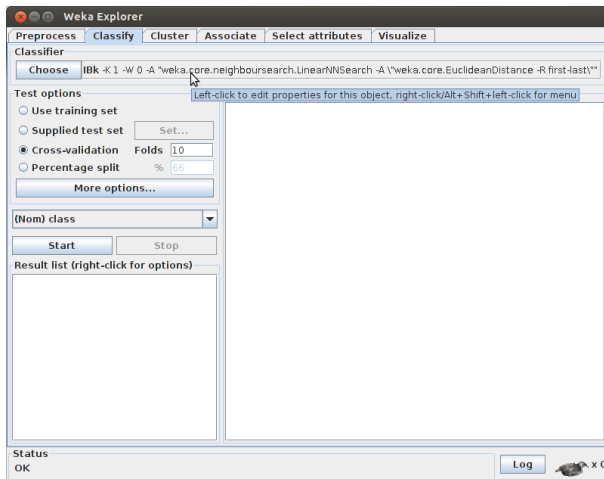
- Alterando parâmetros do algoritmo



Aprendizado de Máquina

Weka

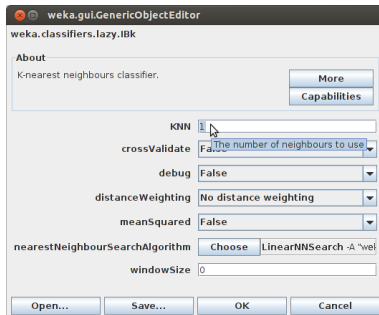
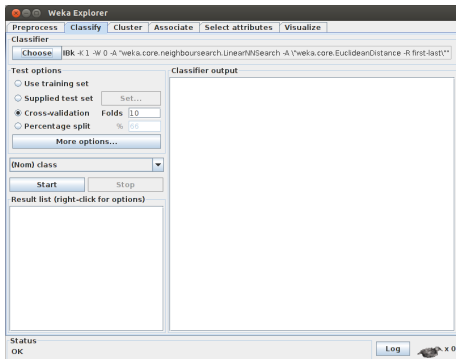
- Alterando parâmetros do algoritmo



Aprendizado de Máquina

Weka

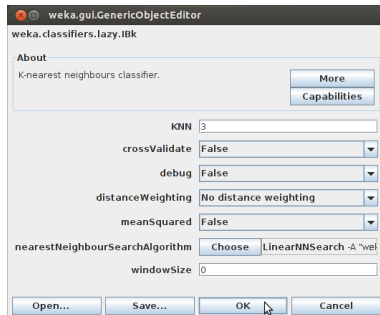
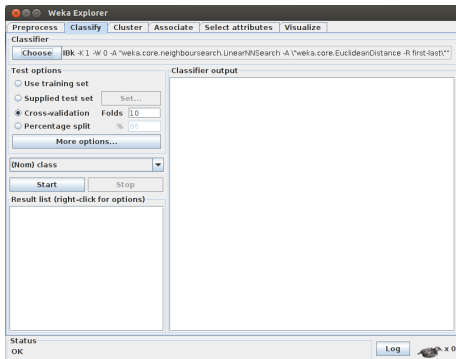
- Alterando parâmetros do algoritmo



Aprendizado de Máquina

Weka

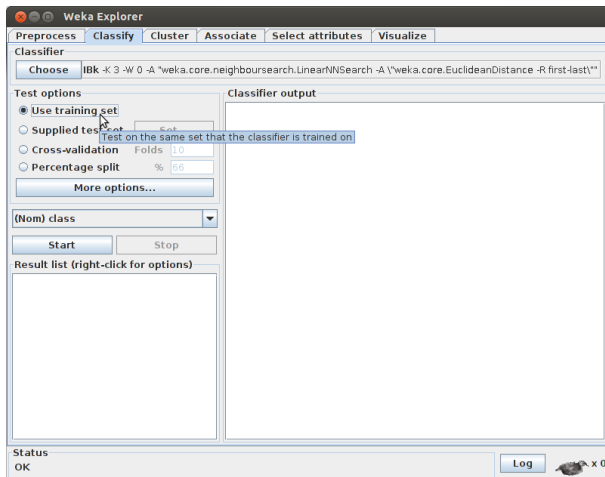
- Alterando parâmetros do algoritmo



Aprendizado de Máquina

Weka

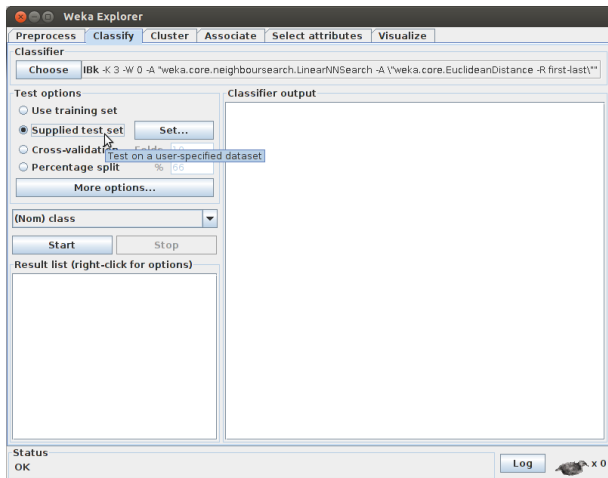
- Realizando testes



Aprendizado de Máquina

Weka

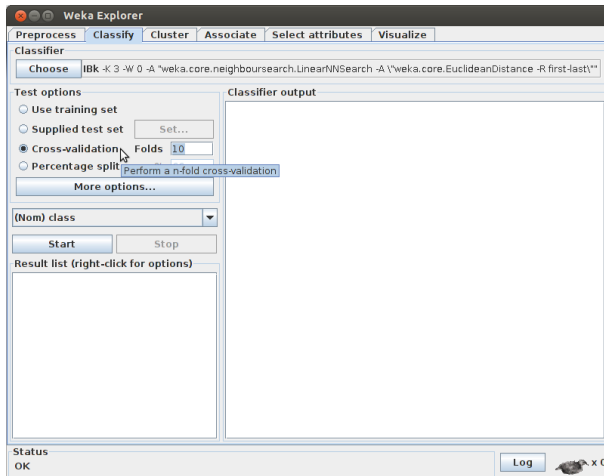
- Realizando testes



Aprendizado de Máquina

Weka

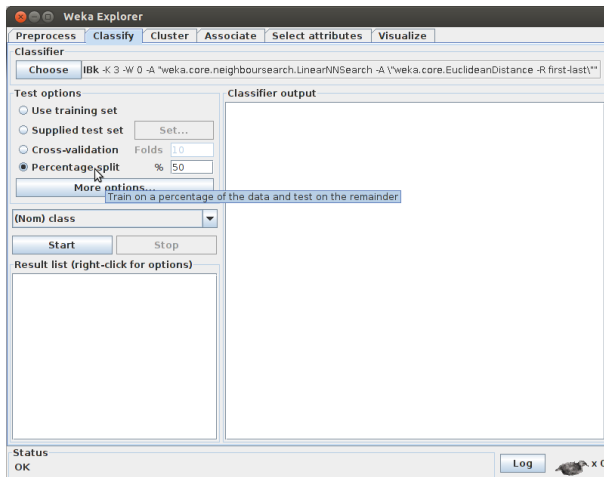
- Realizando testes



Aprendizado de Máquina

Weka

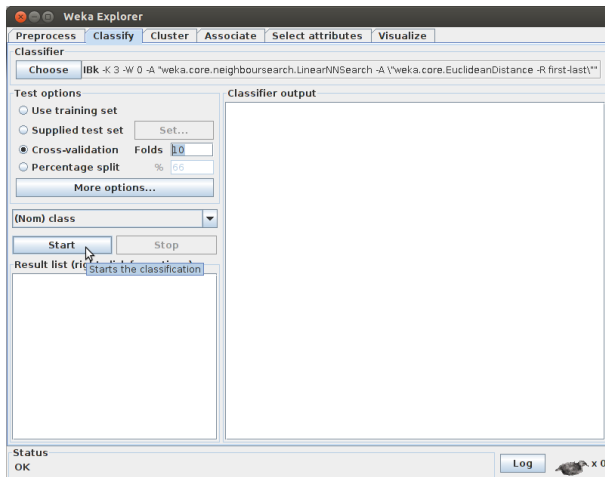
- Realizando testes



Aprendizado de Máquina

Weka

- Realizando testes



Aprendizado de Máquina

Weka

- Realizando testes

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"'. The test options are set to 'Cross-validation' with 'Folds' set to 10. The result list shows '12:11:31 - lazy.IBk'. The classifier output displays the following information:

=== Run information ===

Scheme: weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"

Relations: iris

Instances: 150

Attributes: 5

sepalwidth
sepalwidth
petalwidth
petalwidth
class

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	143	95.3333 %
Incorrectly Classified Instances	7	4.6667 %
Kappa statistic	0.93	
Mean absolute error	0.04	
Root mean squared error	0.1703	
Relative absolute error	9.0013 %	
Root relative squared error	36.1192 %	
Total Number of Instances	150	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	1	Iris-setosa
0.94	0.04	0.922	0.94	0.931	0.963		Iris-versicolor
0.92	0.03	0.939	0.92	0.929	0.958		Iris-virginica
Weighted Avg.	0.953	0.023	0.953	0.953	0.953	0.974	

=== Confusion Matrix ===

a	b	c	-- classified as
50	0	0	a = Iris-setosa
0	47	3	b = Iris-versicolor
0	4	46	c = Iris-virginica

Status: OK

Aprendizado de Máquina

Weka

● Realizando testes

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last""

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

12:11:31 - k

Classifier output

Scheme: weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last""

Relation: iris

Instances: 150

Attributes: 5

sepalwidth

sepalwidth

petalwidth

petalwidth

class

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

IBk instance-based classifier

Test neighbour(s) for classification

to build model: 0 seconds

ed cross-validation ===

===

classified Instances 143 95.3333 %

Classified Instances 7 4.6667 %

tic 0.93

e error 0.04

uared error 0.1703

bsolute error 9.0013 %

e squared error 36.1192 %

of Instances 150

=== detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Iris-setosa
0.94	0.04	0.922	0.94	0.931	0.963	Iris-versicolor
0.92	0.03	0.939	0.92	0.929	0.958	Iris-virginica
Weighted Avg.	0.953	0.023	0.953	0.953	0.974	

=== Confusion Matrix ===

a b c <-- classified as

50 0 0 | a = Iris-setosa

0 47 3 | b = Iris-versicolor

0 4 46 | c = Iris-virginica

Status OK

Log x0

Aprendizado de Máquina

Weka

● Realizando testes

Weka Explorer

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

Classifier: Choose IB1 -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Test options

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds: 10
- ☐ Percentage split %: 66

More options...

(Item) class

Start Stop

Result list (right-click for options)

12:11:31 - lazy-IB1

Classifier output

```
Scheme:weka.classifiers.lazy.IB1 -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""
```

Relation: iris
Instances: 150
Attributes: 5
sepalength
sepalwidth
petallength
petalwidth
class

Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification

Model time: 0 seconds

=== Cross-validation ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0.94	0.04	0.922	0.94	0.931	0.963	Iris-setosa
2	0.92	0.03	0.939	0.92	0.929	0.958	Iris-versicolor
3	0.953	0.023	0.953	0.953	0.953	0.974	Iris-virginica

Weighted Avg.

=== Confusion Matrix ===

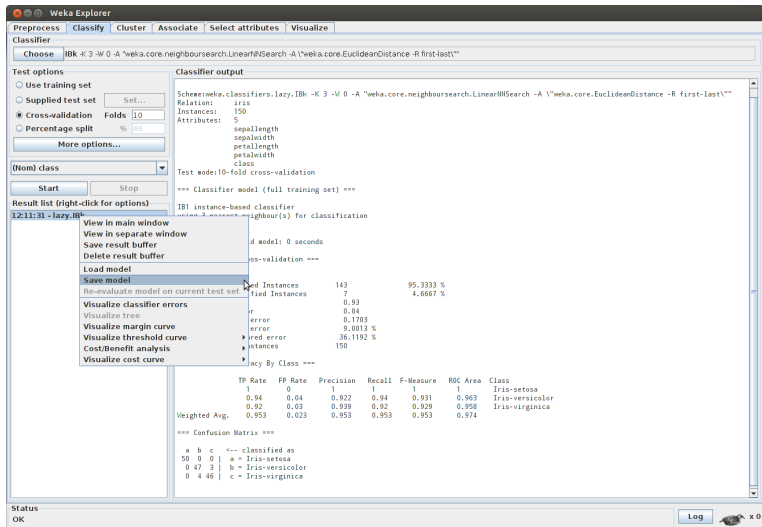
```
a b c <-- classified as
50 0 0 | a = Iris-setosa
0 47 3 | b = Iris-versicolor
0 4 46 | c = Iris-virginica
```

Status: OK

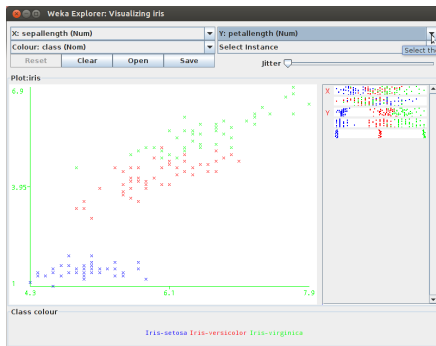
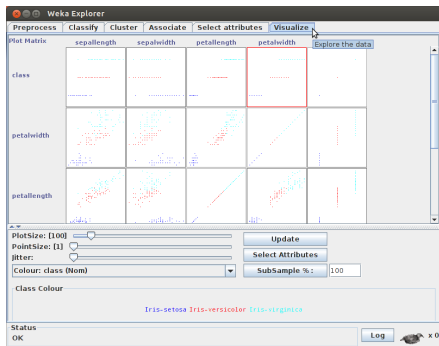
Log x0

Weka

- Realizando testes



Visualizando resultados



Aprendizado de Máquina

Weka

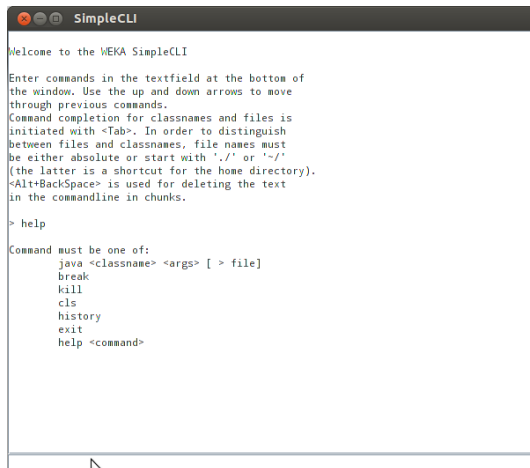
- Simple CLI



Aprendizado de Máquina

Weka

- Simple CLI



Aprendizado de Máquina

Weka

- Simple CLI



```
SimpleCLI

Welcome to the WEKA SimpleCLI

Enter commands in the textfield at the bottom of
the window. Use the up and down arrows to move
through previous commands.
Command completion for classnames and files is
initiated with <Tab>. In order to distinguish
between files and classnames, file names must
be either absolute or start with './' or '~/'
(the latter is a shortcut for the home directory).
<Alt+BackSpace> is used for deleting the text
in the commandline in chunks.

> help

Command must be one of:
  java <classname> <args> [ > file]
  break
  kill
  cls
  history
  exit
  help <command>

java weka.classifiers.trees.J48 -t data/iris.arff
```

Aprendizado de Máquina

Weka

- Simple CLI

```
SimpleCLI
> java weka.classifiers.trees.J48 -t data/iris.arff

J48 pruned tree
-----
petalwidth <= 0.6: Iris-setosa (50,0)
petalwidth > 0.6
| petalwidth <= 1.7
| | petalwidth <= 4.9: Iris-versicolor (48,0/1,0)
| | petalwidth > 4.9
| | | petalwidth <= 1.5: Iris-virginica (3,0)
| | | petalwidth > 1.5: Iris-versicolor (3,0/1,0)
| | petalwidth > 1.7: Iris-virginica (46,0/1,0)

Number of Leaves :    5
Size of the tree :    9

Time taken to build model: 0 seconds
Time taken to test model on training data: 0.01 seconds

=== Error on training data ===
Correctly Classified Instances      147          98 %
Incorrectly Classified Instances      3           2 %
Kappa statistic                    0.97
Mean absolute error                 0.0239
Root mean squared error             0.108
Relative absolute error              5.2482 %
Root relative squared error         22.9089 %
Total Number of Instances          150

=== Confusion Matrix ===
  a b c  <-- classified as
50 0 0 | a = Iris-setosa
 0 49 1 | b = Iris-versicolor
 0 2 46 | c = Iris-virginica

=== Stratified cross-validation ===
Correctly Classified Instances      144          96 %
Incorrectly Classified Instances      6           4 %
Kappa statistic                    0.94
Mean absolute error                 0.035
Root mean squared error             0.1586
Relative absolute error             7.8705 %
Root relative squared error        33.6353 %
Total Number of Instances          150

=== Confusion Matrix ===
  a b c  <-- classified as
49 1 0 | a = Iris-setosa
 0 47 3 | b = Iris-versicolor
 0 2 46 | c = Iris-virginica
```


Weka

- weka.classifiers.**Classifier**
 - ▶ trees.J48
 - ▶ bayes.NaiveBayes
 - ▶ functions.MultilayerPerceptron
 - ▶ lazy.IBk
 - ★ lazy.IBk -K 1

```
java weka.classifiers.trees.J48 -t data/iris.arff -i
```

Aprendizado de Máquina

Weka

```
java weka.classifiers.trees.J48 -t data/iris.arff -i
```

```
java weka.classifiers.bayes.NaiveBayes -K -t data/iris-train.arff -T  
iris-test.arff -p 0
```

```
java weka.classifiers.trees.J48 -t data/iris.arff -i -k -d J48-data.model >&!  
J48-data.out &
```

```
java weka.classifiers.trees.J48 -l J48-data.model -T new-data.arff
```

```
java weka.classifiers.trees.J48 -t test.arff > j48.txt
```

Weka

• WekaManual.pdf

- t especifica o arquivo de treinamento (.arff)
- T especifica o arquivo de teste (.arff). Se não definir tal parâmetro, faz 10-fold cross-validation
- x determina o número de folds para cross-validation
- d salva o modelo após o treinamento
- l carrega o modelo salvo
- p # exibe as predições e um atributo (0 para none) para todas as amostras de teste, se um arquivo de teste é especificado
- i exibe uma descrição de desempenho mais detalhada (precisão, revocação, taxas true e false positive)