

Weka e questão 4 de P2

- Baixar Weka
 - <http://www.cs.waikato.ac.nz/~ml/weka/downloading.html>
- Executar
 - Selecionar opção “Explorer”



Passo 1 – Criar uma base

- Exemplo: **weather.arff**

@relation weather

@attribute outlook {sunny, overcast, rainy}

@attribute temperature numeric

@attribute humidity numeric

@attribute windy {TRUE, FALSE}

@attribute play {yes, no}

@data

sunny,85,85,FALSE,no

sunny,80,90,TRUE,no

overcast,83,86,FALSE,yes

rainy,70,96,FALSE,yes

rainy,68,80,FALSE,yes

rainy,65,70,TRUE,no

overcast,64,65,TRUE,yes

sunny,72,95,FALSE,no

sunny,69,70,FALSE,yes

rainy,75,80,FALSE,yes

sunny,75,70,TRUE,yes

overcast,72,90,TRUE,yes

overcast,81,75,FALSE,yes

rainy,71,91,TRUE,no

Nome (@relation)

Atributos (@attribute) e tipos:

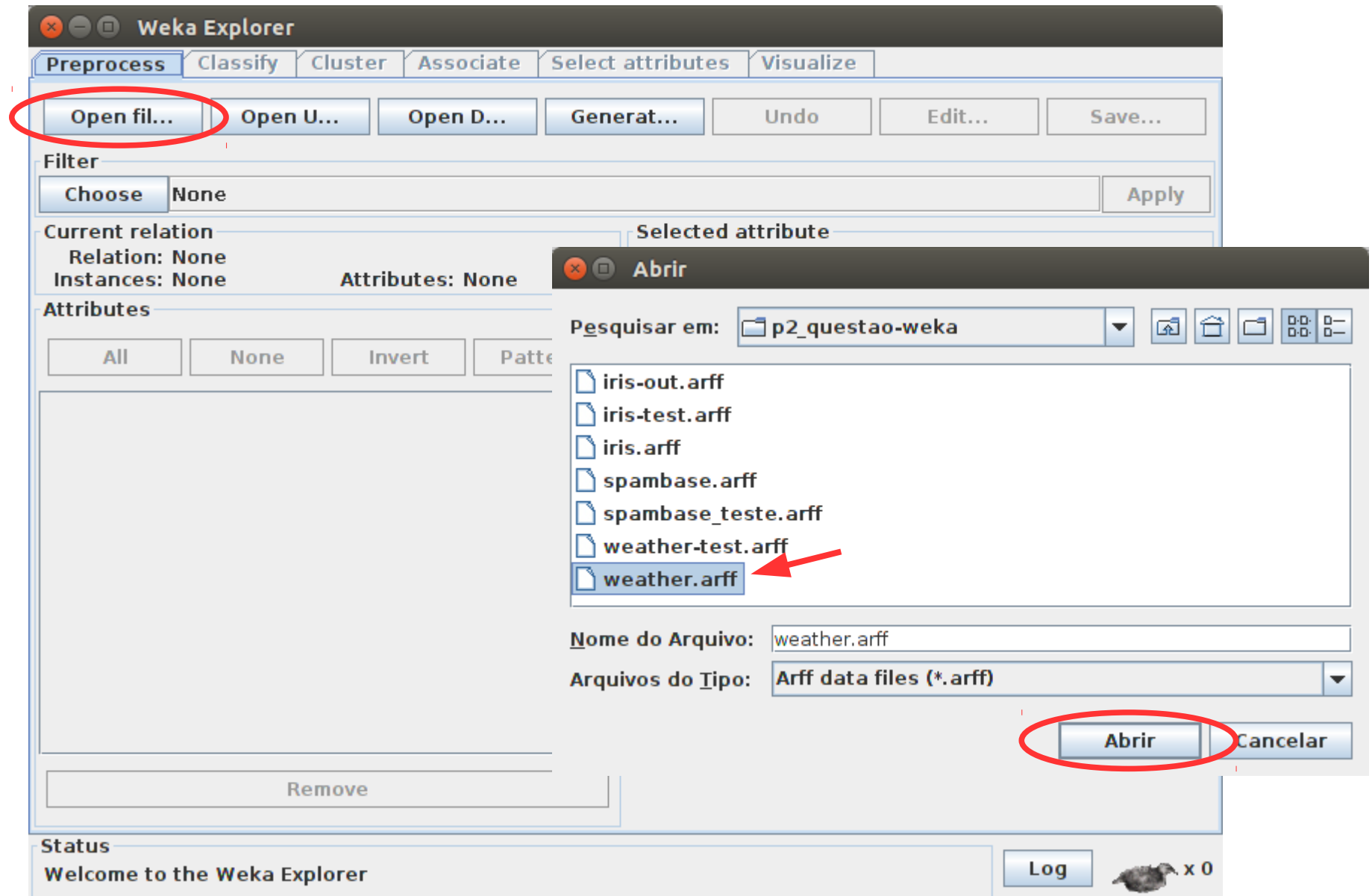
- numérico

- discreto {...}

Dados (@data)

- em cada linha, o valor de cada atributo separado por vírgula

Passo 2 – Abrir uma base (1)



Passo 2 – Abrir uma base (2)

The screenshot shows the Weka Explorer interface with the 'Visualize' tab selected. The 'Selected attribute' is 'outlook', which is a nominal attribute with 3 distinct values and 0 missing values. The 'Class' is 'play (Nom)'. A bar chart visualizes the distribution of 'play' values for each 'outlook' category. The chart shows that for 'sunny' and 'rainy' outlooks, there are 5 instances of 'play' (represented by red bars), and for 'overcast', there are 4 instances (represented by blue bars).

No.	Label	Count
1	sunny	5
2	overcast	4
3	rainy	5

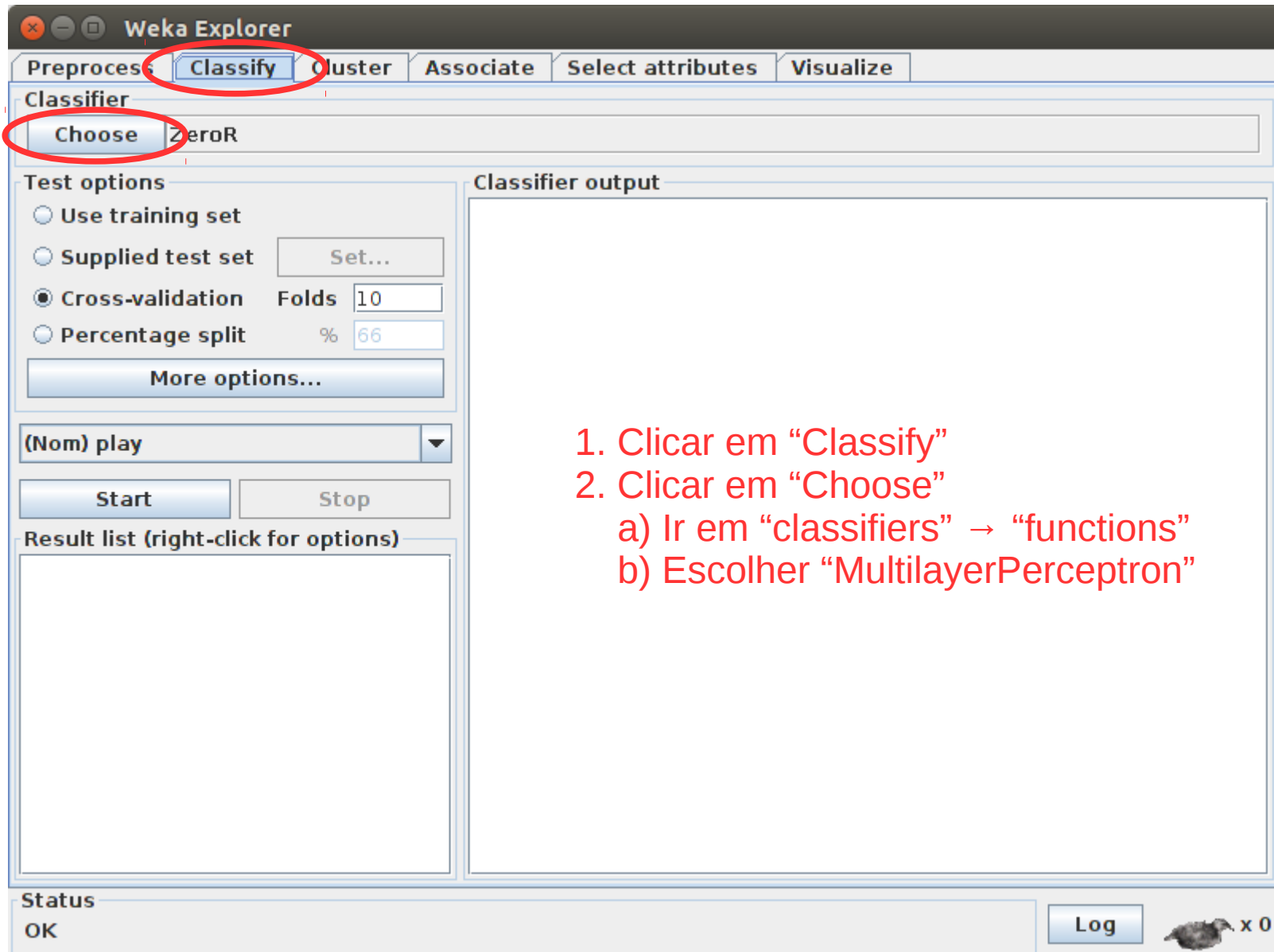
Current relation: Relation: weather, Instances: 14, Attributes: 5

Attributes:

No.	Name
1	<input type="checkbox"/> outlook
2	<input type="checkbox"/> temperature
3	<input type="checkbox"/> humidity
4	<input type="checkbox"/> windy
5	<input type="checkbox"/> play

Status: OK

Passo 3 – Treinamento (1)



The screenshot shows the Weka Explorer application window. The 'Classify' tab is selected in the top menu bar. Below it, the 'Classifier' section has 'Choose' selected. The 'Test options' section is visible, with 'Cross-validation' selected and 'Folds' set to 10. The 'Classifier output' area is empty. The status bar at the bottom shows 'Status OK' and a 'Log' button.

1. Clicar em “Classify”
2. Clicar em “Choose”
a) Ir em “classifiers” → “functions”
b) Escolher “MultilayerPerceptron”

Passo 3 – Treinamento (2)

The image shows the Weka Explorer interface with the 'Classify' tab selected. The 'Classifier' section shows 'MultilayerPerceptron' selected. The 'Test options' section has 'Percentage split' selected with a percentage of 66. A 'weka.gui.GenericObjectEditor' window is open, showing the configuration for 'weka.classifiers.functions.MultilayerPerceptron'. The 'GUI' property is set to 'True'. Red arrows and text annotations indicate the steps: '1. Clicar aqui' points to the 'Choose' button, '2. Colocar "True"' points to the 'GUI' dropdown, and '3. Selecionar' points to the 'Percentage split' radio button.

1. Clicar aqui

2. Colocar "True"

3. Selecionar

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose **MultilayerPerceptron** -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

Use training set

Supplied test set

Cross-validation Folds: 10

Percentage split % 66

(Nom) play

Result list (right-click for options)

Status
OK

weka.gui.GenericObjectEditor

weka.classifiers.functions.MultilayerPerceptron

About

A Classifier that uses backpropagation to classify instances.

GUI **True**

autoBuild True

debug False

decay False

hiddenLayers a

learningRate 0.3

momentum 0.2

nominalToBinaryFilter True

normalizeAttributes True

normalizeNumericClass True

reset False 0

seed 0

trainingTime 500

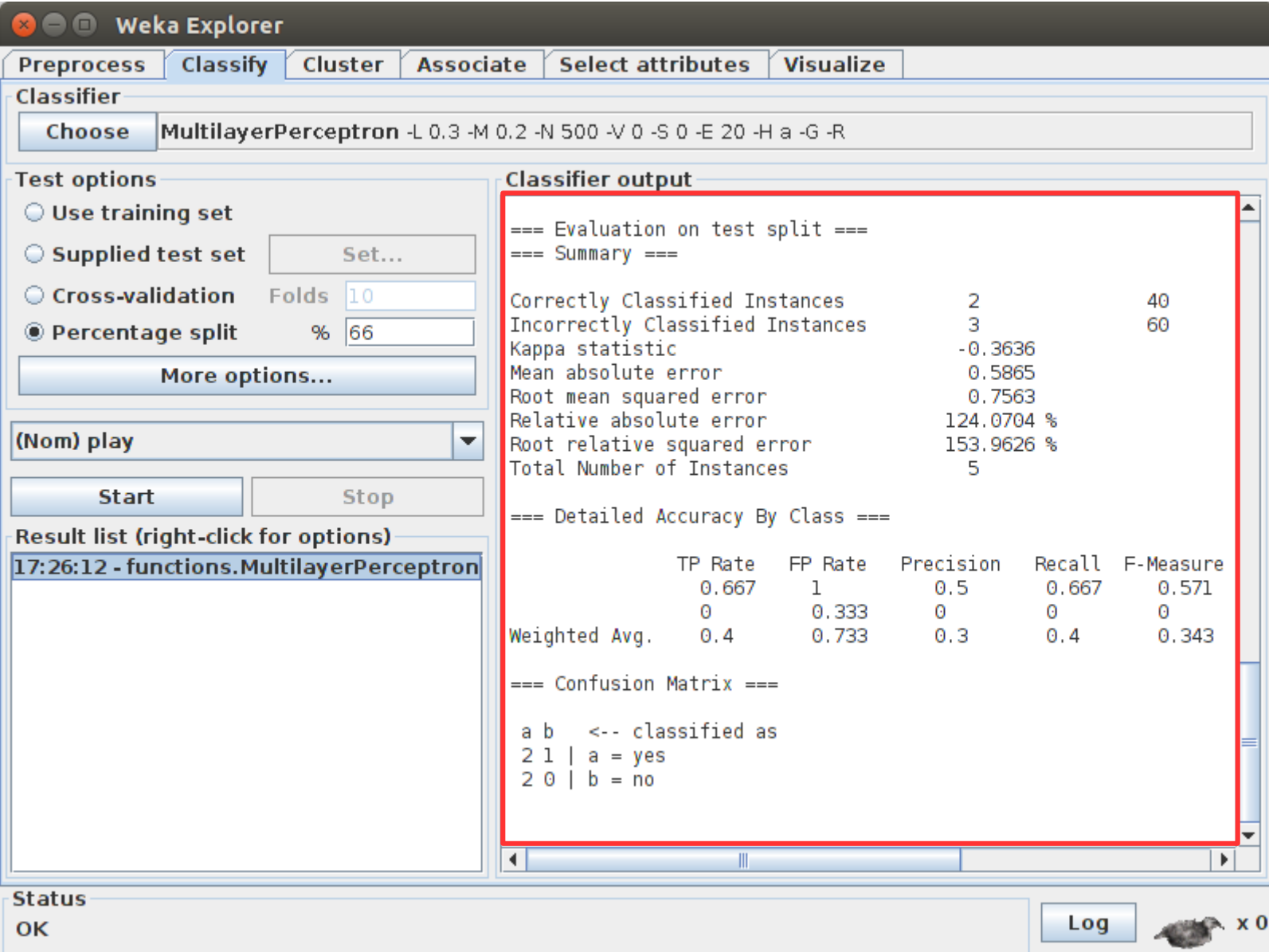
Passo 3 – Treinamento (3)

The image shows the Weka Explorer interface during the training of a Multilayer Perceptron classifier. The main window is titled "Weka Explorer" and has tabs for "Preprocess", "Classify", "Cluster", "Associate", "Select attributes", and "Visualize". The "Classify" tab is active, showing the "Classifier" panel with "MultilayerPerceptron -L 0.3 -M 0.2" selected. The "Test options" section includes radio buttons for "Use training set", "Supplied test set", "Cross-validation", and "Percentage split" (selected), with a "Start" button highlighted by a red arrow and labeled "1. Iniciar".

The "Neural Network" window is overlaid, showing a diagram of a neural network with 6 input nodes (outlook=sunny, outlook=overcast, outlook=rainy, temperature, humidity, windy), 3 hidden nodes, and 2 output nodes (yes, no). A red arrow points to the "Start" button in this window, labeled "2. Iniciar e esperar".

The "Controls" panel at the bottom of the Neural Network window shows "Epoch 0", "Num Of Epochs" set to 500, and "Error per Epoch = 0". A red arrow points to the "Accept" button, labeled "3. Aceitar". The "Learning Rate" is set to 0.3 and "Momentum" is set to 0.2. The "Status" bar at the bottom left shows "OK".

Passo 3 – Treinamento (4)



Weka Explorer

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

Classifier
Choose MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -R

Test options

- Use training set
- Supplied test set
- Cross-validation Folds
- Percentage split %

(Nom) play

Result list (right-click for options)

17:26:12 - functions.MultilayerPerceptron

Classifier output

```
=== Evaluation on test split ===
=== Summary ===

Correctly Classified Instances      2           40
Incorrectly Classified Instances    3           60
Kappa statistic                    -0.3636
Mean absolute error                 0.5865
Root mean squared error             0.7563
Relative absolute error             124.0704 %
Root relative squared error         153.9626 %
Total Number of Instances          5


=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure
                0.667    1        0.5        0.667   0.571
                0        0.333    0          0        0
Weighted Avg.   0.4      0.733    0.3        0.4     0.343

=== Confusion Matrix ===

 a b  <-- classified as
 2 1 | a = yes
 2 0 | b = no
```

Status
OK

 x 0

Passo 4 – Criar uma base de teste

- Exemplo: **weather-test.arff**

```
@relation weather-test
```

```
@attribute outlook {sunny, overcast, rainy}
```

```
@attribute temperature numeric
```

```
@attribute humidity numeric
```

```
@attribute windy {TRUE, FALSE}
```

```
@attribute play {yes, no}
```

```
@data
```

```
sunny,61,89,TRUE,yes
```

```
overcast,80,50,FALSE,yes
```

Aqui, à classe de saída, é atribuído um valor qualquer (pode ser *yes* ou *no*, neste caso). Esta classificação será revisada, após a aplicação dos atributos na rede neural treinada

Nome (@relation)

Atributos (@attribute) e tipos:

- numérico
- discreto {...}

Dados (@data)

- em cada linha, o valor de cada atributo separado por vírgula
- os valores podem ser totalmente diferentes da base de treinamento

Passo 5 – Teste (1)

1. Selecionar
2. Clicar em "Set"

3. Clicar em "Open file"

4. Escolher arquivo de teste

5. Clicar em "Abrir"

Weka Explorer

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

Classifier

Choose **MultilayerPerceptron** -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -R

Test options

Use training set

Supplied test set **Set...**

Cross-validation Folds

Percentage split %

More options...

(Nom) play

Start Stop

Result list (right-click for options)

17:26:12 - functions.MultilayerPerceptron

Test Instances

Relation: None
Instances: None Attributes: None

Open file... Open URL... Close

Abrir

Pesquisar em: p2_questao-weka

- iris-out.arff
- iris-test.arff
- iris.arff
- spambase.arff
- spambase_teste.arff
- weather-test.arff**
- weather.arff

Nome do Arquivo: weather-test.arff

Arquivos do Tipo: Arff data files (*.arff)

Abrir Cancelar

Classifier output

=== Evaluation on test split ===
=== Summary ===

Correctly Classified Instances	2	40
Incorrectly Classified Instances	3	60
Kappa statistic	-0.3636	
Mean absolute error	0.5865	
Root mean squared		
Relative absolute		
Root relative squa		
Total Number of In		

=== Detailed Accur

TP

C

C

Avg. C

ision Matr

- classif

= yes

= no

Status

OK

Passo 5 – Teste (3)

The image shows the Weka Explorer interface with the 'Classify' tab selected. The classifier is set to 'MultilayerPerceptron' with a learning rate of 0.3 and momentum of 0.2. The test options are set to 'Supplied test set'. The 'Start' button is circled in red. The 'Result list' shows the classifier name and time. The 'Neural Network' window displays a diagram of the neural network with input nodes (outlook, temperature, humidity, windy) and output nodes (yes, no). The 'Controls' section at the bottom shows 'Start' and 'Accept' buttons circled in red, along with 'Epoch 500', 'Num Of Epochs 500', 'Error per Epoch = 0.0020689', 'Learning Rate = 0.3', and 'Momentum = 0.2'.

Weka Explorer

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

Classifier

Choose MultilayerPerceptron -L 0.3 -M

Test options

Use training set

Supplied test set **Set...**

Cross-validation Folds 10

Percentage split % 66

More options...

(Nom) play

Start Stop

Result list (right-click for options)

17:26:12 - functions.MultilayerPerceptron

Status
OK

Neural Network

Diagram showing a neural network with input nodes (outlook, temperature, humidity, windy) and output nodes (yes, no). The input nodes are labeled with their values: outlook=sunny, outlook=overcast, outlook=rainy, temperature, humidity, windy.

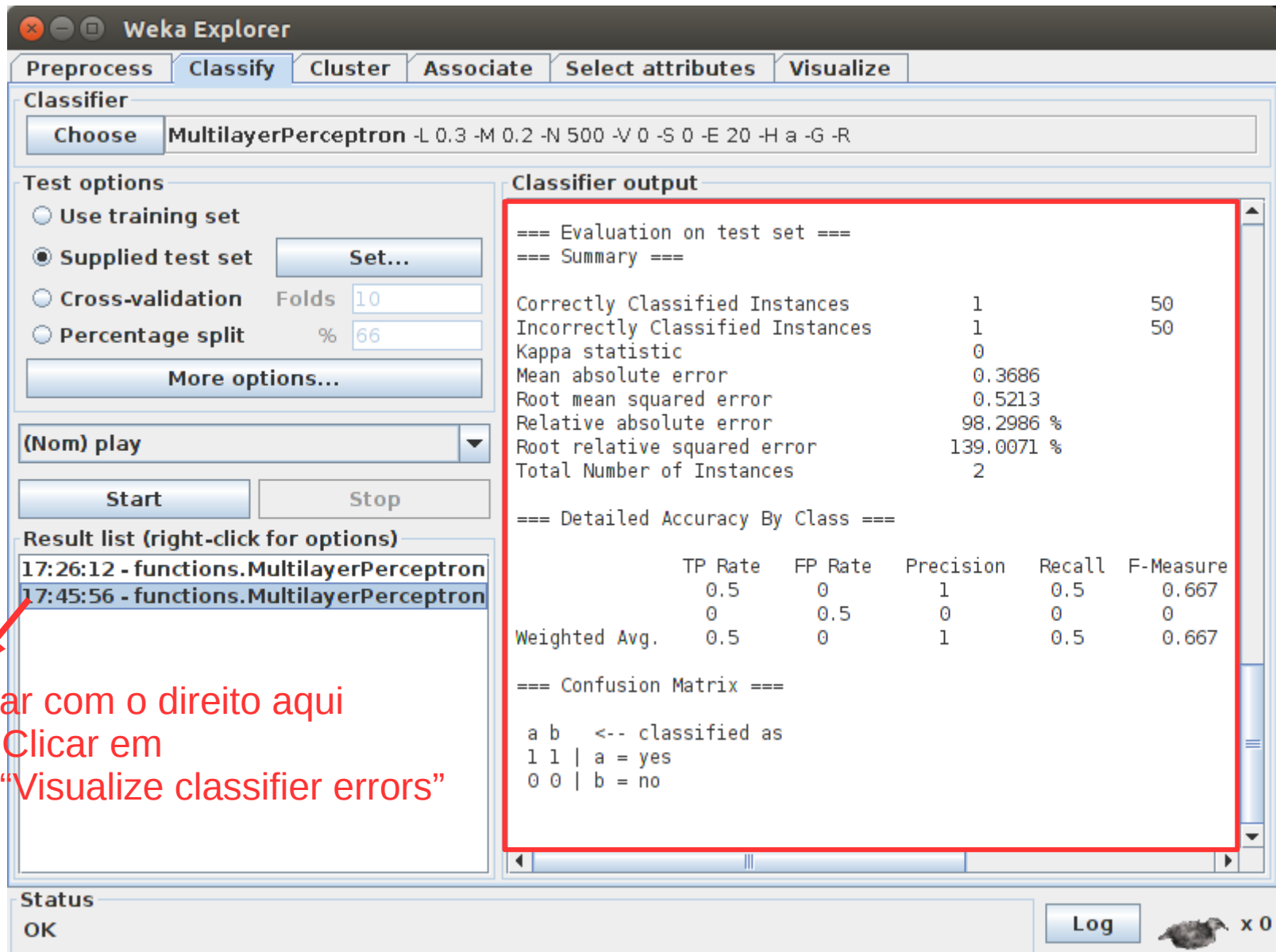
Controls

Start **Accept**

Epoch 500
Num Of Epochs 500
Error per Epoch = 0.0020689

Learning Rate = 0.3
Momentum = 0.2

Passo 5 – Teste (4)



The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier is 'MultilayerPerceptron' with parameters '-L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -R'. The 'Test options' section shows 'Supplied test set' selected with 'Folds' set to 10 and 'Percentage split' at 66%. The 'Classifier output' window displays the following evaluation results:

```
=== Evaluation on test set ===
=== Summary ===
Correctly Classified Instances      1      50
Incorrectly Classified Instances    1      50
Kappa statistic                     0
Mean absolute error                 0.3686
Root mean squared error             0.5213
Relative absolute error             98.2986 %
Root relative squared error        139.0071 %
Total Number of Instances          2

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure
0.5              0.5      0         1          0.5     0.667
0                0       0.5       0          0         0
Weighted Avg.   0.5      0         1          0.5     0.667

=== Confusion Matrix ===
 a b  <-- classified as
 1 1 | a = yes
 0 0 | b = no
```

The 'Result list' shows two entries for 'functions.MultilayerPerceptron' with timestamps 17:26:12 and 17:45:56. A red arrow points to the second entry with the text: 'Clicar com o direito aqui - Clicar em "Visualize classifier errors"'. The status bar at the bottom shows 'OK' and a 'Log' button.

Passo 5 – Teste (5)

The image shows two overlapping windows from the Weka software. The top window is titled "Weka Classifier Visualize: 17:45:56 - functions.MultilayerPerceptron (w)". It features a control panel with dropdown menus for "X: play (Nom)" and "Y: predictedplay (Nom)", a "Colour: play (Nom)" dropdown, and a "Select Instance" dropdown. Below these are buttons for "Reset", "Clear", "Open", and "Save", along with a "Jitter" slider. A plot titled "Plot: weather_predicted" is visible, showing a vertical axis with labels "n" and "0", and a horizontal axis with labels "y", "e", "s", and "yes". A "Class colour" section is at the bottom. A red arrow points to the "Save" button in this window.

The bottom window is titled "Salvar" and is a file save dialog. The "Salvar Em:" field shows the directory "p2_questao-weka". A list of files is displayed, including "iris-out.arff", "iris-test.arff", "iris.arff", "spambase.arff", "spambase_teste.arff", "weather-pred.arff", and "weather-test.arff". The "Nome do Arquivo:" field contains "weather-pred.arff" and the "Arquivos do Tipo:" dropdown is set to "Arff data files". At the bottom right are "Salvar" and "Cancelar" buttons. A red arrow points to the "Nome do Arquivo:" field, and another red arrow points to the "Salvar" button.

Passo 5 – Teste (6)

- Abrir o arquivo gerado: **weather-pred.arff**

```
@relation weather_predicted
```

```
@attribute outlook {sunny, overcast, rainy}
```

```
@attribute temperature numeric
```

```
@attribute humidity numeric
```

```
@attribute windy {TRUE, FALSE}
```

```
@attribute predictedplay {yes,no}
```

```
@attribute play {yes, no}
```

```
@data
```

```
sunny,61,89,TRUE,no,yes
```

```
overcast,80,50,FALSE,yes,yes
```

Resultados da
classificação
automática (após
o treinamento da
rede) para novas
instâncias

Nome (@relation)

Atributos (@attribute) e tipos:

- numérico
- discreto {...}

Dados (@data)

- em cada linha, o valor de cada atributo separado por vírgula
- o penúltimo valor é o predito, ou seja, consiste em uma variável variável nova (predicted) automaticamente criada para classificar novas instâncias

P2 – Questão 4 (25%) - Parte opcional

- Aquecimento
 - Repetir os passos anteriores para a classificação, utilizando redes neurais, do tipo de flor “Iris” (lírios), com base no comprimento e largura de suas pétalas:
 - Base de treinamento:
http://www.inf.ufsc.br/~alexandre.silva/ine5633/provas/p2_questao4/iris.arff
 - Base de teste:
http://www.inf.ufsc.br/~alexandre.silva/ine5633/provas/p2_questao4/iris-test.arff
 - Descrever a acurácia do treinamento e comentar a classificação efetuada pelos casos de teste

P2 – Questão 4 (25%) - Parte obrigatória

- Escolha uma das bases em “UCI Machine Learning Repository: Data Sets (Classification)”:
 - Bases:
<https://archive.ics.uci.edu/ml/datasets.html?format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table>
 - Transforme o arquivo para o formato “arff” do Weka, de modo que se torne o arquivo de treinamento
 - Alternativamente, alguns arquivos em “arff” podem ser obtidos diretamente em:
<http://repository.seasr.org/Datasets/UCI/arff/>
- Crie um arquivo de teste com algumas instâncias novas
- Repita os passos para a execução da rede neural multicamada (*Multilayer Perceptron*) com apoio do Weka
- Entrega:
 - PDF com um breve explicação sobre o problema tratado, além da reprodução e discussão dos resultados obtidos
- Prazo:
 - 03/12