

# Redes Neurais

**WEKA**

<http://www.cs.waikato.ac.nz/ml/weka/>

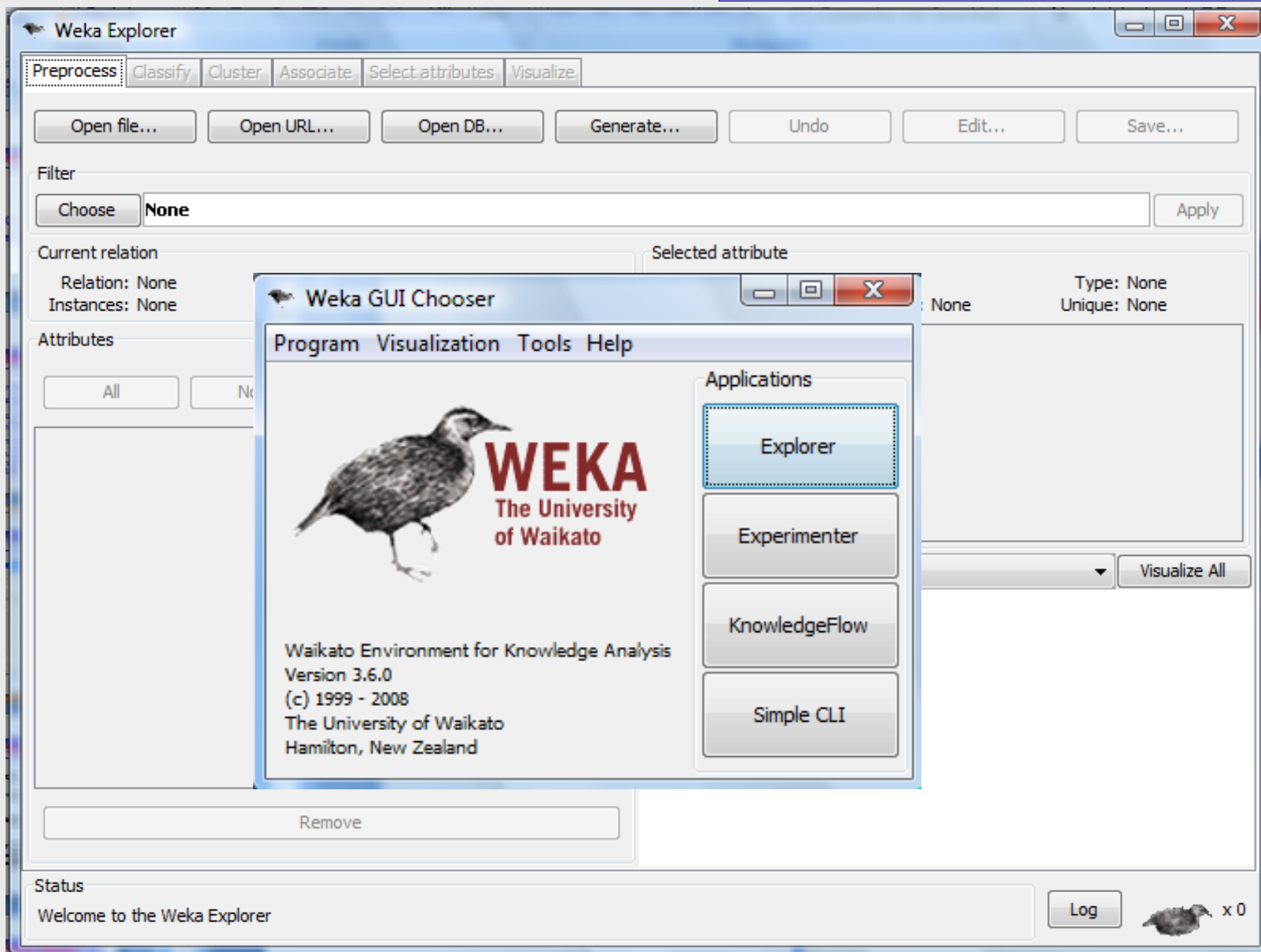
# Redes Neurais no WEKA

- **Introdução ao WEKA**
  - **Base *Benchmark***
- **Estudo de Casos**
  - **Análise de Crédito Bancário**

# Redes Neurais no Weka

- **WEKA (Waikaito Environment for Knowledge Analysis).**
- **Weka é uma ave que existe somente na Nova Zelândia.**
- **Weka é uma sistema escrito em Java.**  
Atenção: se o seu computador não tiver java  
baixar a versão weka with Java (caso contrário  
não roda!).

# WEKA



Weka GUI Choo...   

Waikato Environment for  
Knowledge Analysis

Version 3.4.10

(c) 1999 - 2007  
University of Waikato  
New Zealand



GUI

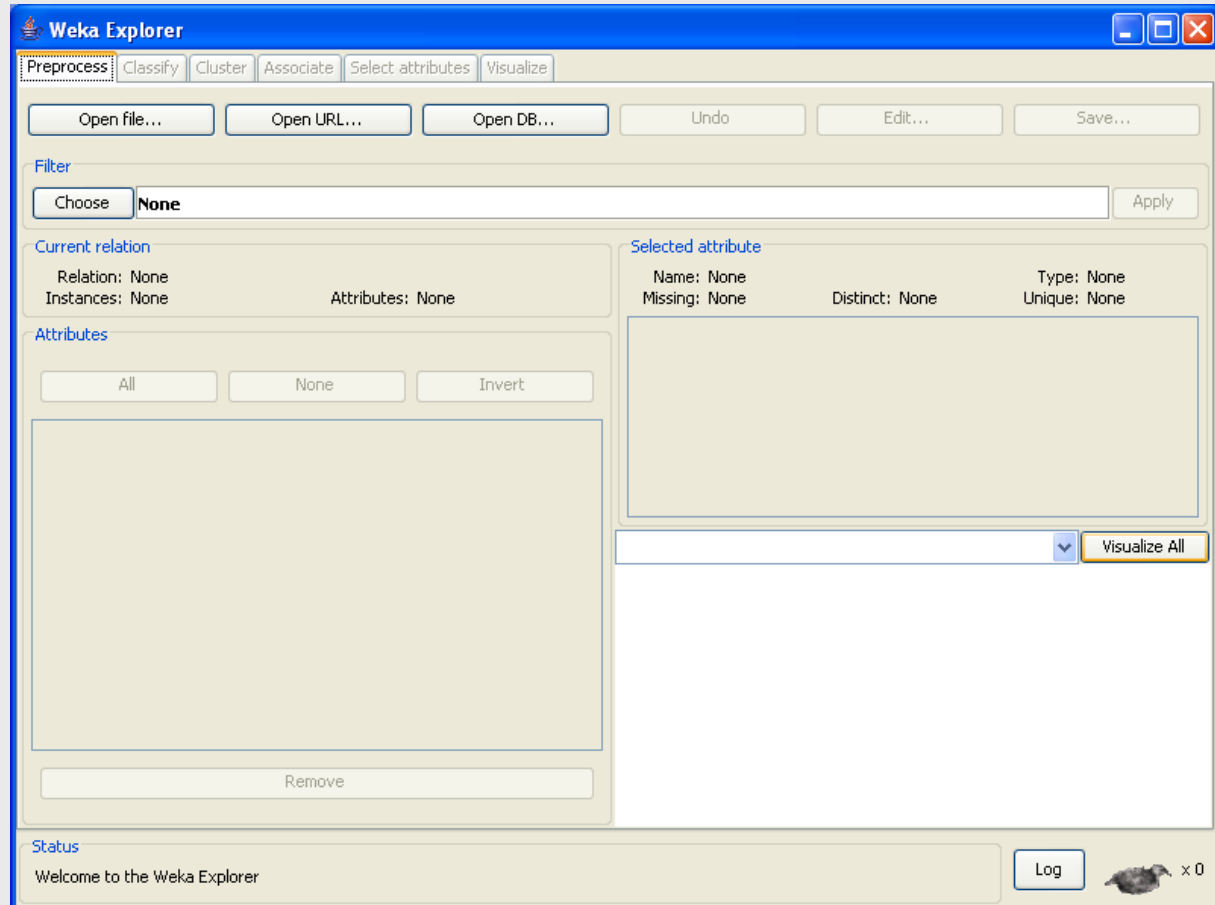
Simple CLI

Explorer

Experimenter

KnowledgeFlow

# WEKA



# Redes Neurais no Weka

- Arquivos possuem extensão .ARFF
- Interface gráfica chamada de *Explorer* é de fácil utilização pelo usuário.
- *Explorer* fornece uma gama de alternativas para investigar o conjunto de dados.
- Na janela de entrada do Weka *Explorer* existem 6 ícones principais no topo: a) pre-processamento; b) classificação; c) agrupamento; d) associação; e) seleção de atributos; f) visualização.

# Redes Neurais no Weka

## Base de Dados e Arquivo . ARFF

Microsoft Excel - weather.csv [Read-Only]

	A	B	C	D	E
1	outlook	temperatu	humidity	windy	play
2					
3	sunny	85	85	FALSE	no
4	sunny	80	90	TRUE	no
5	overcast	83	86	FALSE	yes
6	rainy	70	96	FALSE	yes
7	rainy	68	80	FALSE	yes
8	rainy	65	70	TRUE	no
9	overcast	64	65	TRUE	yes
10	sunny	72	95	FALSE	no
11	sunny	69	70	FALSE	yes
12	rainy	75	80	FALSE	yes
13	sunny	75	70	TRUE	yes
14	overcast	72	90	TRUE	yes
15	overcast	81	75	FALSE	yes
16	rainy	71	91	TRUE	no
17					
18					
19					
20					

weather

Microsoft Word - weather.csv

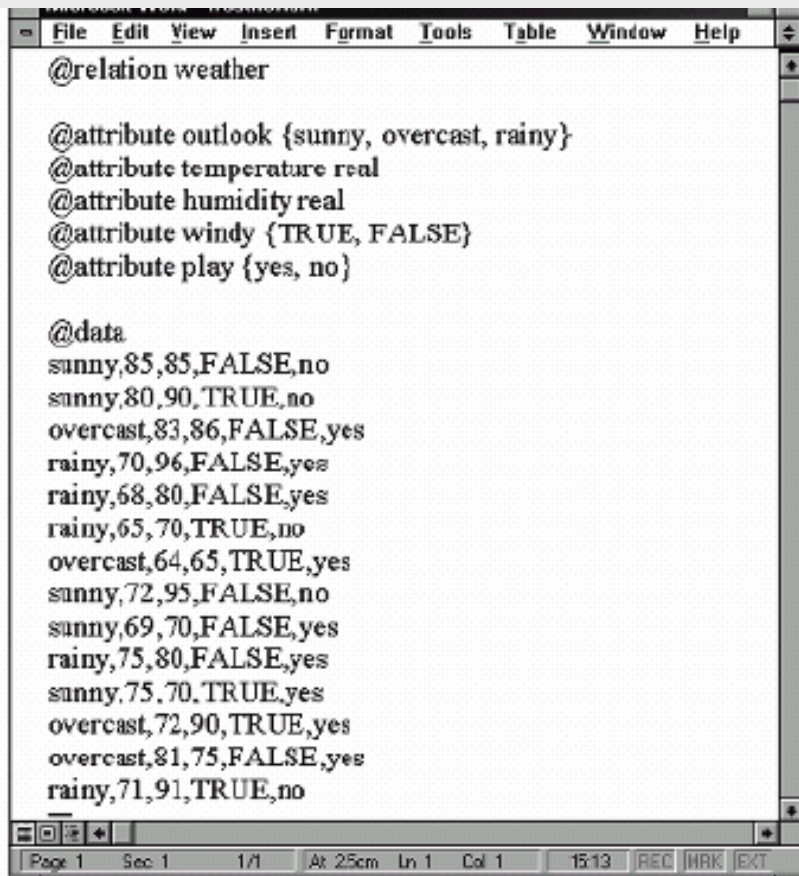
outlook,temperature,humidity,windy,play

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

Page 1 Sec 1 1/1 At 25cm Ln 1 Col 1 15:13 REC WRK EXT

# Redes Neurais no Weka

## Base de Dados e Arquivo . ARFF



The screenshot shows a text editor window with the following ARFF code:

```
@relation weather

@attribute outlook {sunny, overcast, rainy}
@attribute temperature real
@attribute humidity real
@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
```

The window has a menu bar with File, Edit, View, Insert, Format, Tools, Table, Window, and Help. The status bar at the bottom shows Page 1, Sec 1, 1/1, At 25cm, Ln 1, Col 1, 15:13, and buttons for REC, WRK, and EXT.

(c)

**Figure 10.2** Weather data: (a) spreadsheet, (b) CSV format, and (c) ARFF.

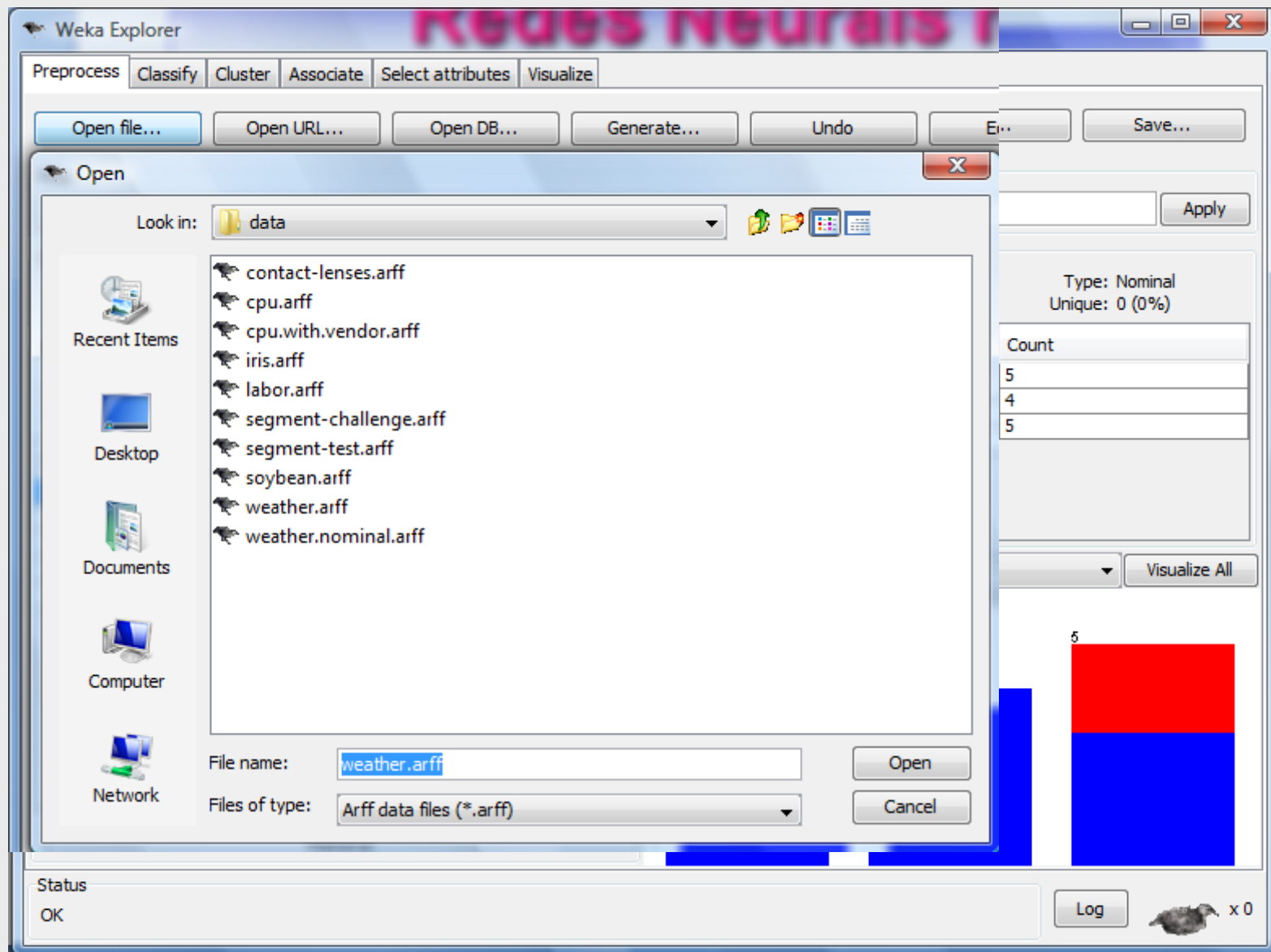


# Redes Neurais no Weka

## a) **Pré-processamento:**

- É onde se abre o arquivo de dados: open file – data – nome do arquivo em .ARFF
- Current relation: fornece o número de exemplos e atributos. (para o arquivo Weather tem-se 14 exemplos e 5 atributos).
- Pode-se selecionar o atributo e interpretar o seu histograma.

# Redes Neurais no Weka



# WEKA

**Weka Explorer**

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

Open file... Open URL... Open DB... Undo Edit... Save...

Filter: Choose **None** Apply

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

Attributes: All None Invert

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

Remove

Selected attribute: Name: outlook Missing: 0 (0%) Distinct: 3 Type: Nominal Unique: 0 (0%)

Label	Count
sunny	5
overcast	4
rainy	5

Class: play (Nom) Visualize All

Status: OK Log x 0

# Redes Neurais no Weka

## Caso: exercício em aula (arquivo *Iris*)

- 150 exemplos;
- 5 atributos: sepallength; sepalwidth; petallength; petawidth; class.
- classificação: Iris-setosa;  
Iris-versicolor;  
Iris-virginica.
- Neste exemplo existe 1 classe linearmente separável e 2 classes não linearmente separáveis.



# Redes Neurais no Weka

**Weka Explorer**

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

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose **None** Apply

Current relation  
Relation: iris  
Instances: 150  
Attributes: 5

Attributes

All | None | Invert | Pattern

No.	Name
1	<input checked="" type="checkbox"/> sepallength
2	<input type="checkbox"/> sepalwidth
3	<input type="checkbox"/> petallength
4	<input type="checkbox"/> petalwidth
5	<input type="checkbox"/> class

Remove

Selected attribute  
Name: sepallength  
Missing: 0 (0%)  
Distinct: 35  
Type: Numeric  
Unique: 9 (6%)

Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

Class: class (Nom) Visualize All

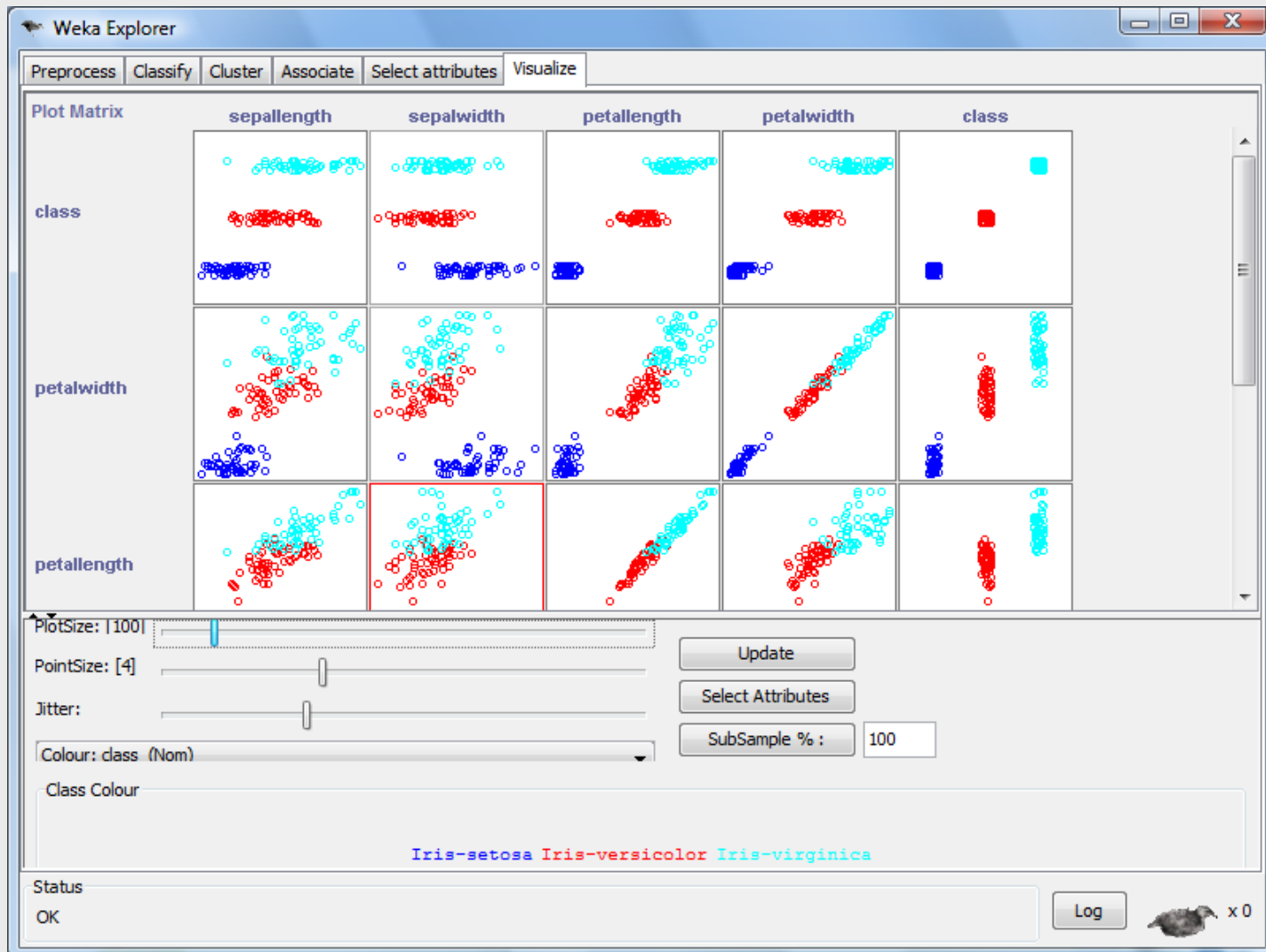
The histogram displays the distribution of the 'sepallength' attribute. The x-axis ranges from 4.3 to 7.9. The y-axis represents frequency. The distribution is multimodal, with three main clusters of data points colored blue, red, and cyan. The blue cluster is centered around 4.8, the red cluster around 5.5, and the cyan cluster around 6.5. The frequencies for each cluster are: Blue (16), Red (30), Cyan (34), Red (28), Cyan (25), Red (10), Cyan (7).

Status: OK

Log x 0

# Redes Neurais no Weka

- Ícone principal visualize:



# Redes Neurais no Weka

Ícone principal *visualize* (permite visualizar o conjunto de dados em si e não os resultados de uma classificação e de clusterização):

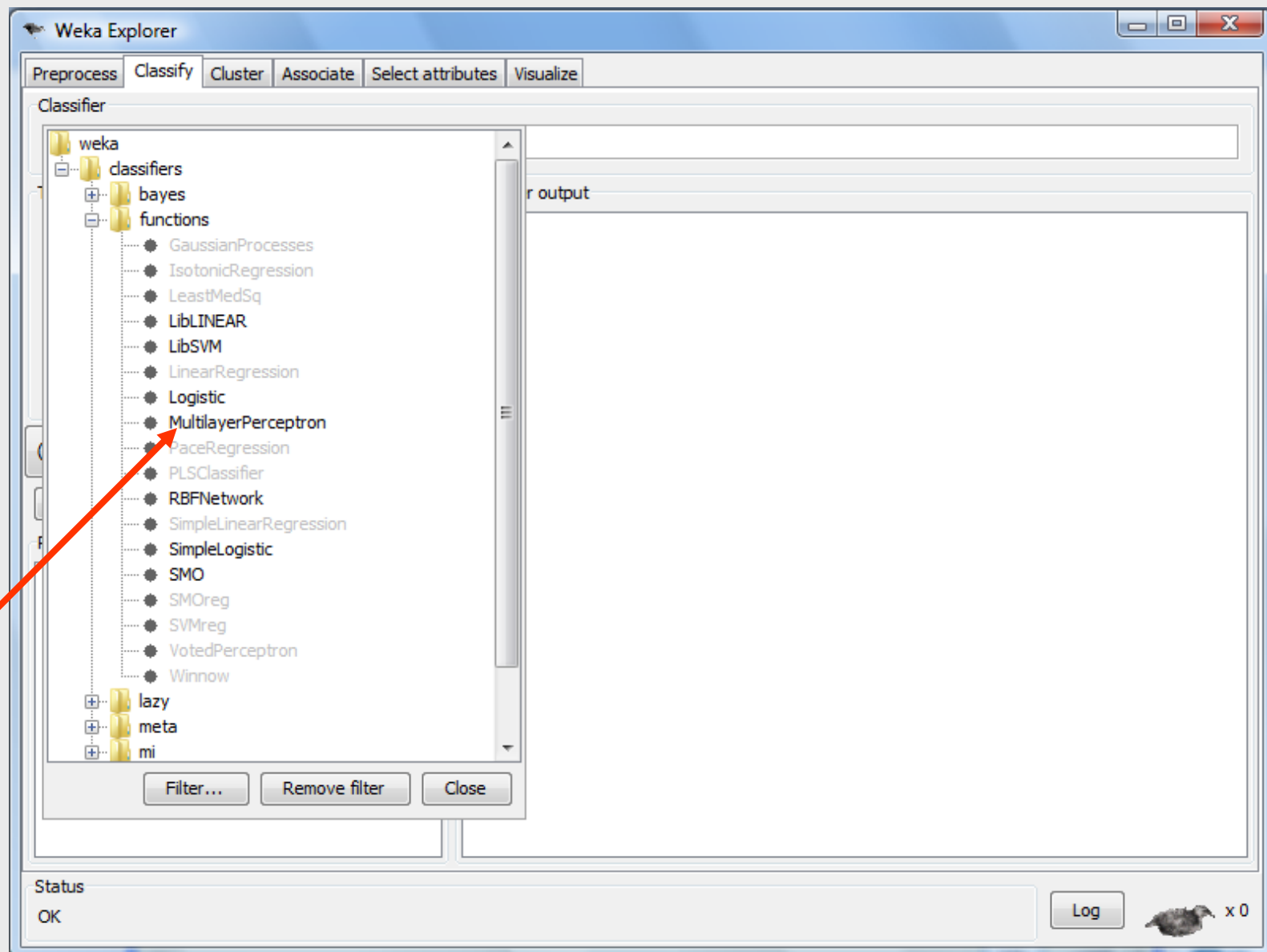
- Plot size (altera o tamanho da caixa)
- Point size (tamanho dos pontos)
- Jitter (mexe na dispersão dos pontos)

## Atenção:

No caso de classificação (exemplo *Iris*), as classes já são conhecidas, podendo-se utilizar um algoritmo de treinamento supervisionado - *Multilayer Perceptron*.

(*classify – choose – functions – multilayer perceptron*)

# Redes Neurais no Weka





# Redes Neurais no Weka

The image shows two overlapping windows from the Weka software. The background window is 'Weka Explorer', and the foreground window is 'weka.gui.GenericObjectEditor'.

**Weka Explorer (Background):**

- Buttons: Preprocess, Classify, Cluster, Associate, Select attributes, Visualize
- Classifier: Choose **MultilayerPerceptron** -L 0.3 -M 0.2 -N 500 -V 0
- Test options:
  - ☐ Use training set
  - ☐ Supplied test set (Set...)
  - ☒ Cross validation Folds: 10
  - ☐ Percentage split %: 66
  - More options...
- Classifiers: (Nom) class
- Buttons: Start, Stop
- Result list (right-click for options):
- Status: OK

**weka.gui.GenericObjectEditor (Foreground):**

weka.classifiers.functions.MultilayerPerceptron

About: A Classifier that uses backpropagation to classify instances. [More] [Capabilities]

GUI: False

autoBuild: True

debug: False

decay: False

hiddenLayers: a

learningRate: 0.3

momentum: 0.2

nominalToBinaryFilter: True

normalizeAttributes: True

normalizeNumericClass: True

randomSeed: 0

reset: True

trainingTime: 500

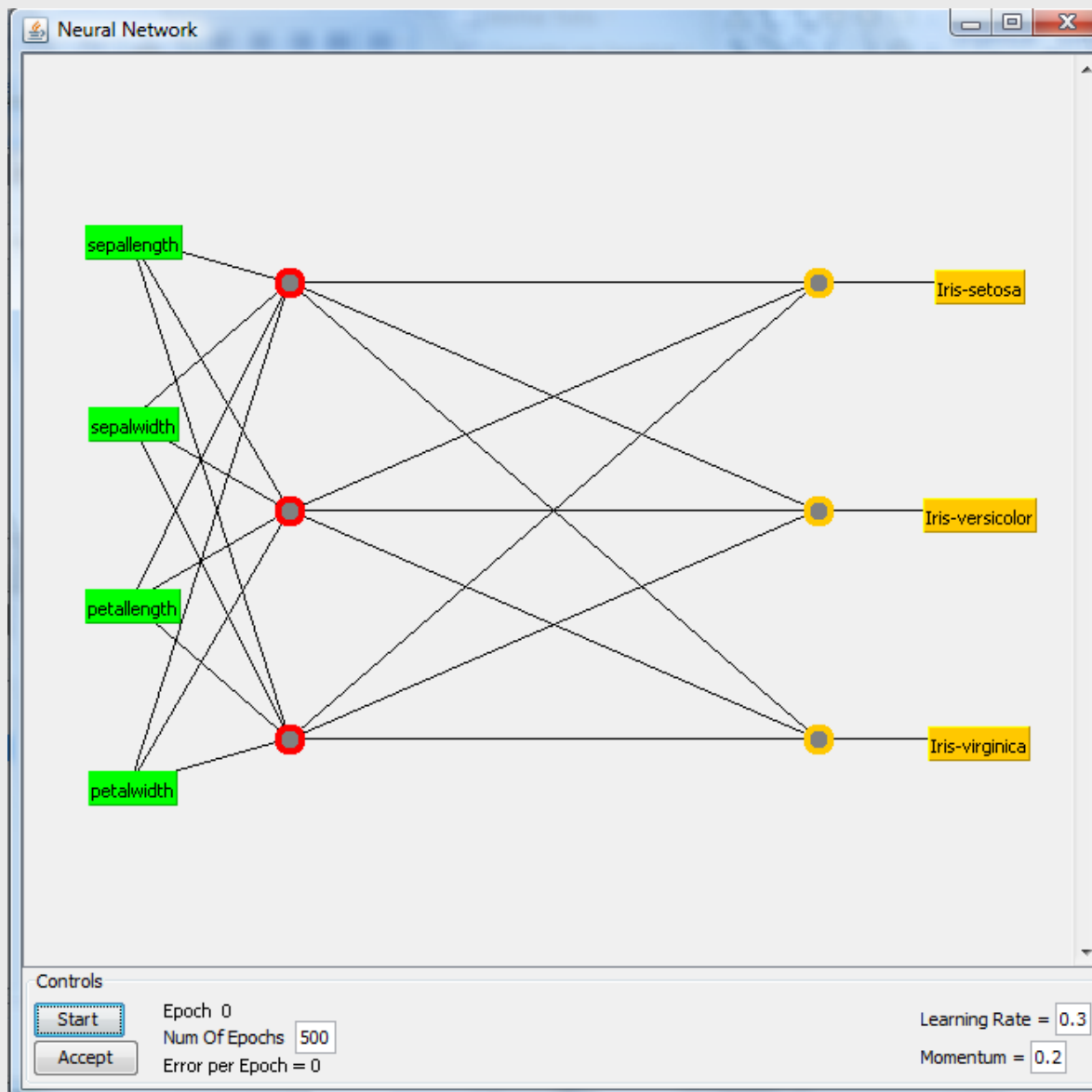
validationSetSize: 0

validationThreshold: 20

[Open...] [Save...] [OK] [Cancel]

An orange arrow points from the 'Cross validation' option in the Weka Explorer 'Test options' section to the 'MultilayerPerceptron' classifier in the 'weka.gui.GenericObjectEditor' window.

# Redes Neurais no Weka



# Redes Neurais no Weka

## Validação cruzada: Exemplo

Por exemplo, para 150 observações, faz-se 10 grupos de 15. Pega-se 9 para treinar a rede e 1 para validar o modelo. A idéia do procedimento consiste em criar grupos e combiná-los várias vezes.

I, II, III, IV.

Treinamento *versus* validação

I, II, III

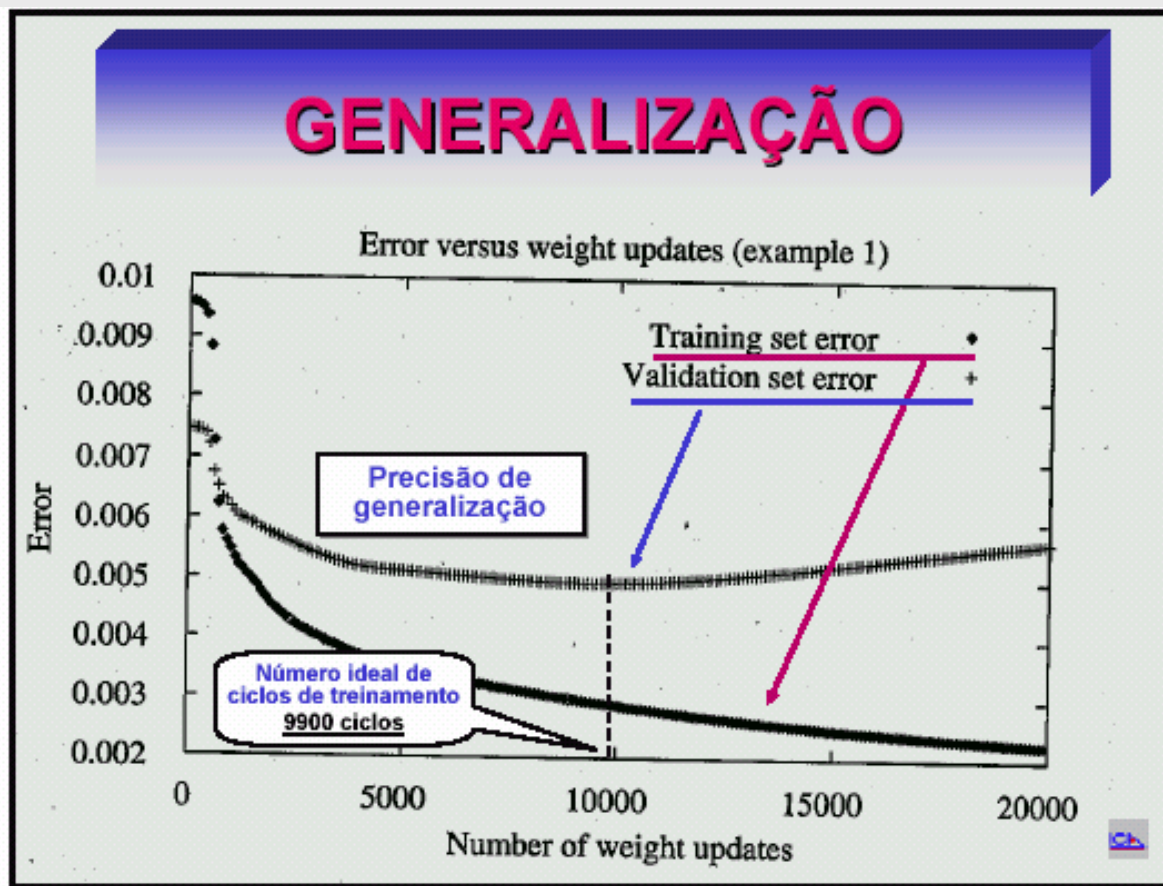
IV

I, II IV

III

# Redes Neurais no Weka

## Early Stopping: exemplo



# Redes Neurais no Weka

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 Set...
- ☐ Cross-validation Folds 10
- ☐ Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

14:21:10 - functions.MultilayerPerceptron

Classifier output

Correctly Classified Instances 148 98.6667 %  
Incorrectly Classified Instances 2 1.3333 %  
Kappa statistic 0.98  
Mean absolute error 0.0248  
Root mean squared error 0.0911  
Relative absolute error 5.5779 %  
Root relative squared error 19.3291 %  
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	Iris-setosa
	0.98	0.01	0.98	0.98	0.98	0.999	Iris-versicolor
	0.98	0.01	0.98	0.98	0.98	0.999	Iris-virginica
Weighted Avg.	0.987	0.007	0.987	0.987	0.987	0.999	

=== Confusion Matrix ===

```
a b c <-- classified as
50 0 0 | a = Iris-setosa
0 49 1 | b = Iris-versicolor
0 1 49 | c = Iris-virginica
```

Status: OK

Log x 0

# Redes Neurais no Weka

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

☐ Percentage split %

(Nom) class

Result list (right-click for options)

- 14:21:10 - functions.MultilayerPerceptron
- 14:31:39 - functions.MultilayerPerceptron

Classifier output

Correctly Classified Instances 146 97.3333 %

Incorrectly Classified Instances 4 2.6667 %

Kappa statistic 0.96

Mean absolute error 0.0327

Root mean squared error 0.1291

Relative absolute error 7.3555 %

Root relative squared error 27.3796 %

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	Iris-setosa
	0.96	0.02	0.96	0.96	0.96	0.996	Iris-versicolor
	0.96	0.02	0.96	0.96	0.96	0.996	Iris-virginica
Weighted Avg.	0.973	0.013	0.973	0.973	0.973	0.998	

=== Confusion Matrix ===

a	b	c	<-- classified as
50	0	0	a = Iris-setosa
0	48	2	b = Iris-versicolor
0	2	48	c = Iris-virginica

Status

OK

 x 0

# Redes Neurais no WEKA

- **Introdução ao WEKA**
  - **Base *Benchmark***
- **Estudo de Casos**
  - **Análise de Crédito Bancário**

# Análise de Crédito Bancário

- A base de dados contém **2077 exemplos** de créditos concedidos.
- Possui **11 atributos** de entrada e **2 classes de saída**.
- A saída indica se o **cliente pagou** o empréstimo (=1) ou **se não pagou** (=0).

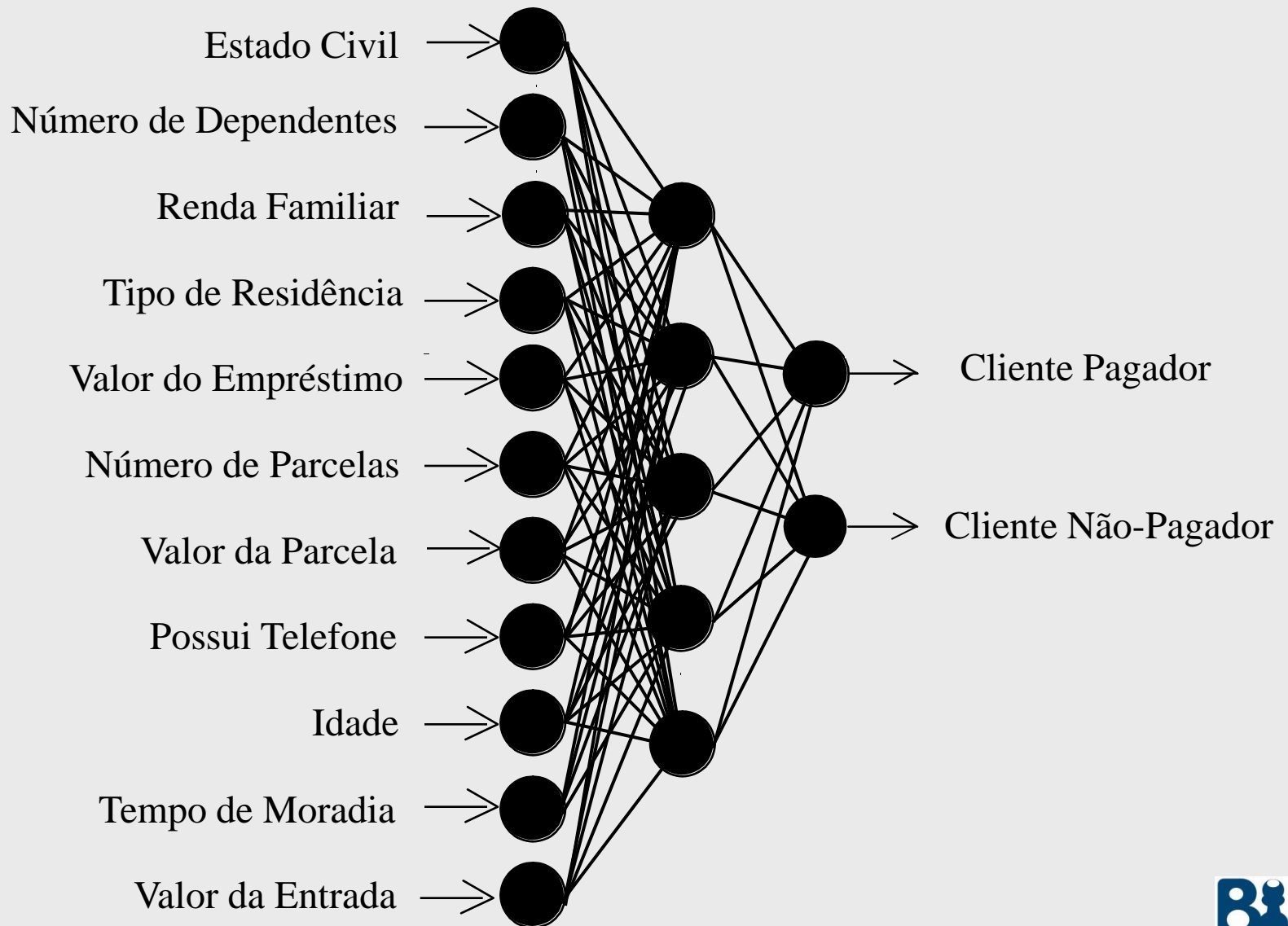


# Análise de Crédito Bancário

- Os 11 atributos de entrada são:

- |           |   |
|-----------|---|
| – ESTC    | Estado civil  |
| – NDEP    | Número de dependentes                                 |
| – RENDA   | Renda familiar  |
| – TIPOR   | Tipo de Residência (0=alugada, 1=própria, 2=parentes) |
| – VBEM    | O valor do bem a ser adquirido                        |
| – NPARC   | O número de parcelas a serem pagas                    |
| – VPARC   | O valor de cada parcelas                              |
| – TEL     | Possui telefone (0=não e 1=sim)                       |
| – IDADE   | Quantos anos a pessoa tem                             |
| – RESMS   | Tempo de moradia (em mês)                             |
| – ENTRADA | O valor da primeira parcela no dia da compra          |

# Análise de Crédito Bancário



# Análise de Crédito Bancário

- Amostra da Base de Dados:

ESTC	NDEP	RENDA	TIPOR	VBEM	NPARC	VPARC	TEL	IDADE	RESMS	ENTRADA	CLASSE_1	CLASSE_2
0	0	350	1	468	10	65	0	33	6	0	0	1
0	2	580	1	349	10	51	0	30	6	0	1	0
0	1	500	1	828	18	89	0	29	6	0	0	1
0	0	2000	1	309	6	66	0	29	6	0	1	0
1	0	411	1	390	12	55	1	40	0	0	1	0
0	0	500	1	614	10	81	0	59	6	0	1	0
0	2	500	1	347	4	101	0	44	6	0	1	0
0	0	800	1	359	6	76	0	51	6	0	1	0
0	0	300	1	309	6	66	0	30	6	0	0	1
0	0	554	1	429	10	60	0	46	6	0	1	0
0	0	332	1	593	4	100	0	44	215	198	1	0
1	0	342	1	453	10	59	0	52	6	0	1	0
0	2	632	1	365	6	78	0	49	6	0	1	0
0	0	600	1	587	10	89	0	44	60	0	1	0
0	0	800	1	495	15	59	0	50	6	0	0	1
1	0	480	1	460	3	135	0	59	6	100	1	0
0	0	1300	1	349	10	51	0	45	6	0	1	0
0	2	868	1	567	10	78	0	39	6	0	1	0
0	0	300	1	399	10	55	0	29	6	0	1	0

# Análise de Crédito Bancário

- A partir da base de dados original, foram criadas:
  - 3 bases para treinamento, escolhidas aleatoriamente a partir da base original, com 1500 exemplos.
  - 3 bases para teste com 577 exemplos.

# Análise de Crédito Bancário

- Abra o arquivo *treino01.arff* no WEKA.
- Verifique se todos os exemplos (*instâncias*) foram corretamente carregadas.
- Analise os histogramas dos atributos de entrada. Que comentário pode ser feito a respeito do número de dependentes (NDEP)?

# Análise de Crédito Bancário

- Treine e teste (com o respectivo conjunto de teste) cada uma das bases de dados.
- Faça variações nos seguintes parâmetros:
  - Normalização
  - Número de processadores na camada escondida
  - Número de épocas de treinamento
  - Taxa de aprendizado
  - Termo de momento
  - Validação cruzada

# Análise de Crédito Bancário

- No caso do atributo **Número de Dependentes**, deve-se alterar os valores para:
  - 0 dependentes
  - 1 ou mais dependentes.
- Utilize o filtro *MergeTwoValues*.