

Sistemas de Representação de Conhecimento e Raciocínio

Exercício 3

Relatório de Desenvolvimento

Grupo 18

Cesário Miguel Pereira Pernetá - A73883

Luís Miguel Bravo Ferraz - A70824

Rui Pedro Barbosa Rodrigues - A74572

Tiago Miguel Fraga Santos - A74092

12 de Maio de 2018

Resumo

Neste relatório será abordada a resolução e verificação do terceiro trabalho prático da unidade curricular de Sistemas de Representação de Conhecimento e Raciocínio. Será possível encontrar uma detalhada explicação dos predicados elaborados, bem como a apresentação dos scripts utilizados, que visam dar resposta aos requisitos propostos neste trabalho prático.

Conteúdo

| | | |
|----------|--|-----------|
| 1 | Introdução | 2 |
| 2 | Análise do Problema | 3 |
| 3 | Análise e Tratamento dos Dados para a representação do conhecimento do problema | 6 |
| 3.1 | DataSet - Vinho Tinto | 7 |
| 3.1.1 | Script - WEKA | 7 |
| 3.1.2 | Script - R | 8 |
| 3.2 | DataSet - Vinho Branco | 9 |
| 3.2.1 | Script - WEKA | 9 |
| 3.2.2 | Script - R | 10 |
| 3.3 | DataSet - Vinho Tinto e Branco distinguidos pelo tipo (12 atributos) | 11 |
| 3.3.1 | Script - WEKA | 11 |
| 3.3.2 | Script - R | 12 |
| 3.4 | DataSet - Vinho Tinto e Branco (22 atributos) | 13 |
| 3.4.1 | Script - WEKA | 13 |
| 3.4.2 | Script - R | 14 |
| 4 | Topologias, aprendizagem e testes das RNA's | 16 |
| 4.1 | Opção A - 2 RNA's abastecidas com 2 DataSets | 16 |
| 4.1.1 | Script em R - Vinho Tinto | 16 |
| 4.1.2 | Script em R - Vinho Branco | 18 |
| 4.1.3 | Conclusões e Resultados Obtidos | 20 |
| 4.2 | Opção B - 1 RNA abastecida por um DataSet de 12 atributos | 21 |
| 4.2.1 | Script em R | 21 |
| 4.2.2 | Conclusões e resultados obtidos | 22 |
| 4.3 | Opção C - 1 RNA abastecida por um DataSet de 22 atributos | 22 |
| 4.3.1 | Script em R | 22 |
| 4.3.2 | Conclusões e resultados obtidos | 24 |
| 5 | Conclusão | 25 |

Capítulo 1

Introdução

Com a realização deste trabalho o objetivo da equipa docente era a motivação dos alunos para a utilização de sistemas não simbólicos na representação de conhecimento e no desenvolvimento de mecanismos de raciocínio, nomeadamente, Redes Neuronais Artificiais (RNAs) para a resolução de problemas. Neste terceiro e último exercício pretende-se que seja realizado um estudo que envolva a identificação da qualidade de vinhos (branco e tinto) através do estudo de vários atributos pelos quais, estes são constituídos. No próximo capítulo vamos apresentar os *datasets* e fazer uma análise pormenorizada dos mesmos e do problema em questão.

Capítulo 2

Análise do Problema

De forma a iniciar o trabalho prático os professores disponibilizaram dados que descrevem a qualidade de vinhos através da análise de onze factores. De forma a perceber se uma dada entrada de informação de um vinho tem um selo de qualidade correto face aos seus atributos é preciso fazer uma análise correta dos dados fornecidos, pois são estes que vamos utilizar para efetuar a aprendizagem da rede e respetivos testes. Os ficheiros são compostos por 11 atributos sendo eles:

- **fixed acidity** - acidez fixa;
- **volatile acidity** - acidez volátil;
- **citric acid** - acidez citrica;
- **residual sugar** - açúcar residual;
- **chlorides** - cloretos;
- **free sulfur dioxide** - dióxido livre de enxofre;
- **total sulfur dioxide** - dióxido total de enxofre;
- **density** - densidade;
- **pH**;
- **sulphates** - sulfatos;
- **alcohol** - álcool;

Os dados dos vinhos foram distribuídos por dois ficheiros, um referente ao vinho tinto (*winequality-red.csv*), e outro ao vinho branco (*winequality-white.csv*). De forma a procedermos a uma correta análise dos dados e efetuar trabalho sólido e competente, decidimos dividir o trabalho por duas etapas, sendo elas:

- Análise e Tratamento dos Dados para a representação do conhecimento do problema;
- Topologias, aprendizagem e testes das RNA's

No primeiro ponto, pretendemos fazer uma análise aos dados fornecidos de forma a estabelecer os principais atributos para definir a qualidade dos vinhos, e com o auxílio da ferramenta **WEKA**, iremos verificar as estatísticas de cada atributo para saber de que forma influenciam a decisão final. Após a análise decidiremos se existe algum atributo a alterar ou se podemos passar para a fase seguinte com os dados inalteráveis. De seguida, iremos organizar os dados de maneiras diferentes para fornecer a diferentes redes neuronais, com o intuito que no fim possamos avaliar os resultados consoante diferentes neurónios de entrada para diferentes redes. Após conversa com o professor, decidimos que iríamos avançar com três opções. A primeira opção irá ser composta por duas redes neuronais com 11 neurónios cada uma. Cada neurónio é caracterizado pelos 11 atributos da lista em cima. A segunda opção pretendemos criar uma rede neuronal com 12 neurónios de entrada. Além dos atributos já apresentados, o segundo neurónio é caracterizado por um novo atributo:

- **type** - tipo do vinho(0 - Tinto; 1- Branco);

A terceira opção é composta por uma rede neuronal com 22 neurónios de entrada. Cada neurónio é caracterizado pelos seguintes atributos:

- **fixed_acidity_red** - acidez fixa tinto;
- **volatile_acidity_red** - acidez volátil tinto;
- **citric_acid_red** - acidez cítrica tinto;
- **residual_sugar_red** - açúcar residual tinto;
- **chlorides_red** - cloretos tinto;
- **free_sulfur_dioxide_red** - dióxido livre de enxofre tinto;
- **total_sulfur_dioxide_red** - dióxido total de enxofre tinto;
- **density_red** - densidade tinto;
- **pH_red**; - pH tinto;
- **sulphates_red** - sulfatos tinto;
- **alcohol_red** - álcool tinto;
- **fixed_acidity_white** - acidez fixa branco;
- **volatile_acidity_white** - acidez volátil branco;
- **citric_acid_white** - acidez cítrica branco;
- **residual_sugar_white** - açúcar residual branco;
- **chlorides_white** - cloretos branco;
- **free_sulfur_dioxide_white** - dióxido livre de enxofre branco;
- **total_sulfur_dioxide_white** - dióxido total de enxofre branco;
- **density_white** - densidade branco;
- **pH_white**; - pH branco;
- **sulphates_white** - sulfatos branco;
- **alcohol_white** - álcool branco;

De forma a conseguir, estes dois ultimos *Datasets* tivemos de juntar os dois *datasets* fornecidos pelo prof, e fazer as respetivas alterações como acabaram de ser descritas, quer seja ao nível de classificar o vinho conforme o seu tipo ou alterar o nome dos atributos como neste ultimo caso.

Por fim, no terceiro ponto descrito em cima, após ter definido todos os neuronios de entrada para cada uma das redes neuronais, iremos criar cada uma das redes, designando para cada uma a sua topologia, as suas formas de aprendizagem bem como os respetivos testes com o intuito de exercitar as redes e tirar conclusões sobre os resultados obtidos.

Capítulo 3

Análise e Tratamento dos Dados para a representação do conhecimento do problema

Neste primeiro capítulo, vamos fazer a análise dos atributos mais importantes com o intuito de perceber quais deles é que influenciam a aprendizagem das redes neuronais em maior escala. Para o fazer, utilizamos a ferramenta **WEKA** e alguns comandos interpretados pelo **RStudio** que iremos apresentar em cada secção dos *Datasets*. Antes da análise detalhada, referimos que em todas as seleções no software *WEKA*, utilizamos o *attribute evaluator* **CfsSubsetEval** e para o *search method* utilizamos o **BestFirst**, isto porque, foram os únicos parâmetros que nos forneceram informações relativas sobre os atributos principais para cada *DataSet*. De notar que o *WEKA* oferece muitos mais *attribute evaluators* e *search methods*. A grande maioria não podiam ser utilizados com os nossos dados. O *search method* *BestFirst* e o *GreedyStepwise* devolviam sempre a mesma seleção de atributos, no entanto, o *GreedyStepwise* não oferece informação sobre os nodos. O *attribute evaluator* *WrapperSubsetEval* não fazia seleção de qualquer atributo. Utilizando o *search method* *Ranker* com qualquer um dos possíveis *attribute evaluators* em vez de obtermos uma seleção de atributos tal como os *search methods* *BestFirst* e *GreedyStepwise*, obtemos uma classificação de todos os atributos. Os comandos na linguagem **R** precessados no *RStudio* foram os seguintes:

```
# -----
#           For Red and White Wines DataSet
# -----
# Formula to calculate quality of the wines
formula_A <- quality ~
    fixed_acidity + volatile_acidity +
    citric_acid + residual_sugar +
    chlorides + free_sulfur_dioxide +
    total_sulfur_dioxide + density + pH + sulphates + alcohol

# See the most important attributes
selecao_A <- regsubsets(formula_A,data,method="backward")
summary(selecao_A)

# -----
#           For Dataset with 12 attributes
# -----

# Formula to calculate quality of the wines
formula_A <- quality ~
```



```

        fixed_acidity + volatile_acidity +
        citric_acid + residual_sugar + chlorides +
        free_sulfur_dioxide + total_sulfur_dioxide +
        density + pH + sulphates + alcohol + type

# See the most important atributes
selecao_A <- regsubsets(formula_A,data,method="backward")
summary(selecao_A)

# -----
#           For Dataset with 22 atributes
# -----

# Formula to calculate Quality
formula_A <- quality ~
fixed_acidity_red + volatile_acidity_red +
citric_acid_red + residual_sugar_red + chlorides_red +
free_sulfur_dioxide_red + total_sulfur_dioxide_red +
density_red + pH_red + sulphates_red + alcohol_red +
fixed_acidity_white + volatile_acidity_white +
citric_acid_white + residual_sugar_white +
chlorides_white + free_sulfur_dioxide_white +
total_sulfur_dioxide_white + density_white +
pH_white + sulphates_white + alcohol_white

# See the most important atributes
selecao_A <- regsubsets(formula_A,data,method="backward")
summary(selecao_A)

```

Com o auxilio da função *regsubsets* a funcionar através do metodo *backward* conseguimos visualizar com a função *summary* a avaliação dos atributos mais importantes para os quatro *DataSets* que avaliamos.

3.1 DataSet - Vinho Tinto

3.1.1 Script - WEKA

```

=== Run information ===

Evaluator:   weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search:      weka.attributeSelection.BestFirst -D 1 -N 5
Relation:    winequality-red
Instances:    1599
Attributes:   12
              fixed_acidity
              volatile_acidity
              citric_acid
              residual_sugar
              chlorides
              free_sulfur_dioxide
              total_sulfur_dioxide
              density
              pH
              sulphates

```

```

        alcohol
        quality
Evaluation mode:    evaluate on all training data

```

=== Attribute Selection on all input data ===

```

Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 62
  Merit of best subset found:    0.548

```

```

Attribute Subset Evaluator (supervised, Class (numeric): 12 quality):
  CFS Subset Evaluator
  Including locally predictive attributes

```

```

Selected attributes: 2,3,10,11 : 4
                    volatile_acidity
                    citric_acid
                    sulphates
                    alcohol

```

No primeiro *DataSet* analisado, que contem as informações sobre o vinho tinto, como podemos ver em cima, através da análise do script fornecido pelo *WEKA*, os principais atributos que irão ser utilizados nas formulas para aprendizagem das redes neuronais sao: o **volatile_acidity**, o **citric_acid**, o **sulphates** e o **alcohol**.

3.1.2 Script - R

```

Subset selection object
Call: regsubsets.formula(formula_Red, data_red, method = "backward")
11 Variables (and intercept)

```

| | | Forced in | Forced out |
|----------------------|--|-----------|------------|
| fixed_acidity | | FALSE | FALSE |
| volatile_acidity | | FALSE | FALSE |
| citric_acid | | FALSE | FALSE |
| residual_sugar | | FALSE | FALSE |
| chlorides | | FALSE | FALSE |
| free_sulfur_dioxide | | FALSE | FALSE |
| total_sulfur_dioxide | | FALSE | FALSE |
| density | | FALSE | FALSE |
| pH | | FALSE | FALSE |
| sulphates | | FALSE | FALSE |
| alcohol | | FALSE | FALSE |

```

1 subsets of each size up to 8
Selection Algorithm: backward

```

| | fixed_acidity | volatile_acidity | citric_acid | residual_sugar | chlorides | free_sulfur_dioxide |
|---------|---------------|------------------|-------------|----------------|-----------|---------------------|
| 1 (1) | " " | " " | " " | " " | " " | " " |
| 2 (1) | " " | "*" | " " | " " | " " | " " |
| 3 (1) | " " | "*" | " " | " " | " " | " " |
| 4 (1) | " " | "*" | " " | " " | " " | " " |

```

5 ( 1 ) " "          "*"          " "          " "          "*"          " "
6 ( 1 ) " "          "*"          " "          " "          "*"          " "
7 ( 1 ) " "          "*"          " "          " "          "*"          "*"
8 ( 1 ) " "          "*"          "*"          " "          "*"          "*"

total_sulfur_dioxide density pH sulphates alcohol
1 ( 1 ) " "          " "          " " " "          "*"
2 ( 1 ) " "          " "          " " " "          "*"
3 ( 1 ) " "          " "          " " "*"          "*"
4 ( 1 ) "*"          " "          " " "*"          "*"
5 ( 1 ) "*"          " "          " " "*"          "*"
6 ( 1 ) "*"          " "          "*" "*"          "*"
7 ( 1 ) "*"          " "          "*" "*"          "*"
8 ( 1 ) "*"          " "          "*" "*"          "*"

```

O scrip fornecido pelo *RStudio* vem confirmar os atributos fornecidos em cima pelo *WEKA* com a exceção que atribui mais influencia ao atributo **chlorides** em vez do **citric_acid**.

Em suma, de forma a carregar os dados de aprendizagem para a rede neuronal que os irá processar, vamos testar as duas versões com duas formulas distintas, uma com os atributos fornecidos pelo *WEKA* e a outra com estes ultimos atributos.

3.2 DataSet - Vinho Branco

3.2.1 Script - WEKA

```

=== Run information ===

Evaluator:   weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search:      weka.attributeSelection.BestFirst -D 1 -N 5
Relation:    winequality-white
Instances:   4898
Attributes:   12
              fixed_acidity
              volatile_acidity
              citric_acid
              residual_sugar
              chlorides
              free_sulfur_dioxide
              total_sulfur_dioxide
              density
              pH
              sulphates
              alcohol
              quality
Evaluation mode:  evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward

```

```

Stale search after 5 node expansions
Total number of subsets evaluated: 70
Merit of best subset found:    0.452

```

```

Attribute Subset Evaluator (supervised, Class (numeric): 12 quality):
CFS Subset Evaluator
Including locally predictive attributes

```

```

Selected attributes: 2,5,9,10,11 : 5
    volatile_acidity
    chlorides
    pH
    sulphates
    alcohol

```

No *DataSet* analisado, que contem as informações sobre o vinho branco, como podemos ver em cima, através da análise do script fornecido pelo *WEKA*, os principais atributos que irão ser utilizados nas formulas para aprendizagem das redes neuronais sao: o **volatile_acidity**, o **chlorides**, o **pH**, o **sulphates** e o **alcohol**.

3.2.2 Script - R

```

Subset selection object
Call: regsubsets.formula(formula_A, data, method = "backward")
11 Variables (and intercept)

```

| | Fixed in | Forced in | Forced out |
|----------------------|----------|-----------|------------|
| fixed_acidity | FALSE | FALSE | FALSE |
| volatile_acidity | FALSE | FALSE | FALSE |
| citric_acid | FALSE | FALSE | FALSE |
| residual_sugar | FALSE | FALSE | FALSE |
| chlorides | FALSE | FALSE | FALSE |
| free_sulfur_dioxide | FALSE | FALSE | FALSE |
| total_sulfur_dioxide | FALSE | FALSE | FALSE |
| density | FALSE | FALSE | FALSE |
| pH | FALSE | FALSE | FALSE |
| sulphates | FALSE | FALSE | FALSE |
| alcohol | FALSE | FALSE | FALSE |

```

1 subsets of each size up to 8
Selection Algorithm: backward

```

| | fixed_acidity | volatile_acidity | citric_acid | residual_sugar | chlorides | free_sulfur_dioxide |
|---------|---------------|------------------|-------------|----------------|-----------|---------------------|
| 1 (1) | " " | " " | " " | " " | " " | " " |
| 2 (1) | " " | "*" | " " | " " | " " | " " |
| 3 (1) | " " | "*" | " " | "*" | " " | " " |
| 4 (1) | " " | "*" | " " | "*" | " " | " " |
| 5 (1) | " " | "*" | " " | "*" | " " | " " |
| 6 (1) | " " | "*" | " " | "*" | " " | " " |
| 7 (1) | " " | "*" | " " | "*" | " " | "*" |
| 8 (1) | "*" | "*" | " " | "*" | " " | "*" |

| | total_sulfur_dioxide | density | pH | sulphates | alcohol |
|---------|----------------------|---------|-----|-----------|---------|
| 1 (1) | " " | " " | " " | " " | "*" |
| 2 (1) | " " | " " | " " | " " | "*" |
| 3 (1) | " " | " " | " " | " " | "*" |
| 4 (1) | " " | "*" | " " | " " | "*" |
| 5 (1) | " " | "*" | "*" | " " | "*" |

| | | | | | |
|---|-------|-----|-----|-------|-----|
| 6 | (1) | " " | "*" | "*" " | "*" |
| 7 | (1) | " " | "*" | "*" " | "*" |
| 8 | (1) | " " | "*" | "*" " | "*" |

O scrip fornecido pelo *RStudio* vem confirmar os atributos fornecidos em cima pelo *WEKA* com a exceção que atribui mais influencia aos atributos **residual_sugar** e **density** em vez do **chlorides** e do **sulphates**.

Mais uma vez, de forma a carregar os dados de aprendizagem para a rede neuronal que os irá processar, vamos testar as duas versões com duas formulas distintas, uma com os atributos fornecidos pelo *WEKA* e a outra com estes ultimos atributos.

3.3 DataSet - Vinho Tinto e Branco distinguidos pelo tipo (12 atributos)

3.3.1 Script - WEKA

```
=== Run information ===

Evaluator:    weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search:       weka.attributeSelection.BestFirst -D 1 -N 5
Relation:     winequality-all-type
Instances:    6497
Attributes:    13
               fixed_acidity
               volatile_acidity
               citric_acid
               residual_sugar
               chlorides
               free_sulfur_dioxide
               total_sulfur_dioxide
               density
               pH
               sulphates
               alcohol
               type
               quality
Evaluation mode:    evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 75
  Merit of best subset found:    0.353

Attribute Subset Evaluator (supervised, Class (numeric): 13 quality):
  CFS Subset Evaluator
  Including locally predictive attributes

Selected attributes: 3,11,12 : 3
```

```

citric_acid
alcohol
type

```

No *DataSet* analisado, que contem as informações sobre o vinho tinto e vinho branco distinguidos por um atributo novo(**type**), como podemos ver em cima, através da análise do script fornecido pelo *WEKA*, os principais atributos que irão ser utilizados nas formulas para aprendizagem das redes neuronais sao: o **citric_acid**, o **alcohol** e o **type**.

3.3.2 Script - R

```

Subset selection object
Call: regsubsets.formula(formula_A, data, method = "backward")
12 Variables (and intercept)

```

| | | Forced in | Forced out |
|----------------------|--|-----------|------------|
| fixed_acidity | | FALSE | FALSE |
| volatile_acidity | | FALSE | FALSE |
| citric_acid | | FALSE | FALSE |
| residual_sugar | | FALSE | FALSE |
| chlorides | | FALSE | FALSE |
| free_sulfur_dioxide | | FALSE | FALSE |
| total_sulfur_dioxide | | FALSE | FALSE |
| density | | FALSE | FALSE |
| pH | | FALSE | FALSE |
| sulphates | | FALSE | FALSE |
| alcohol | | FALSE | FALSE |
| type | | FALSE | FALSE |

```

1 subsets of each size up to 8
Selection Algorithm: backward

```

| | fixed_acidity | volatile_acidity | citric_acid | residual_sugar | chlorides | free_sulfur_dioxide |
|---------|---------------|------------------|-------------|----------------|-----------|---------------------|
| 1 (1) | " " | " " | " " | " " | " " | " " |
| 2 (1) | " " | " " | " " | " " | " " | " " |
| 3 (1) | " " | " " | " " | " " | " " | " " |
| 4 (1) | " " | " " | " " | " " | " " | "* |
| 5 (1) | " " | " " | " " | " " | " " | "* |
| 6 (1) | " " | "* | " " | " " | " " | "* |
| 7 (1) | " " | "* | " " | " " | "* | "* |
| 8 (1) | " " | "* | "* | " " | "* | "* |

| | total_sulfur_dioxide | density | pH | sulphates | alcohol | type |
|---------|----------------------|---------|-----|-----------|---------|------|
| 1 (1) | " " | " " | " " | " " | "* | " " |
| 2 (1) | " " | " " | " " | " " | "* | "* |
| 3 (1) | "* | " " | " " | " " | "* | "* |
| 4 (1) | "* | " " | " " | " " | "* | "* |
| 5 (1) | "* | " " | " " | "* | "* | "* |
| 6 (1) | "* | " " | " " | "* | "* | "* |
| 7 (1) | "* | " " | " " | "* | "* | "* |
| 8 (1) | "* | " " | " " | "* | "* | "* |

No script do *RStudio*, verificamos que os atributos **alcohol** e **type** aparecem em ambos os casos. Neste segundo caso, em vez do protagonismo dado ao **citric_acid** é atribuido um papel mais relevante ao **free_sulfur_dioxide** e ao **sulphates**.

Como nos casos anteriores, de forma a carregar os dados de aprendizagem para a rede neuronal que os irá processar, vamos testar as duas versões com duas formulas distintas, uma com os atributos fornecidos pelo *WEKA* e a outra com estes ultimos atributos.

3.4 DataSet - Vinho Tinto e Branco (22 atributos)

3.4.1 Script - WEKA

=== Run information ===

```
Evaluator:    weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search:       weka.attributeSelection.BestFirst -D 1 -N 5
Relation:     winequality-all-22
Instances:    6497
Attributes:    23
               fixed_acidity_red
               volatile_acidity_red
               citric_acid_red
               residual_sugar_red
               chlorides_red
               free_sulfur_dioxide_red
               total_sulfur_dioxide_red
               density_red
               pH_red
               sulphates_red
               alcohol_red
               fixed_acidity_white
               volatile_acidity_white
               citric_acid_white
               residual_sugar_white
               chlorides_white
               free_sulfur_dioxide_white
               total_sulfur_dioxide_white
               density_white
               pH_white
               sulphates_white
               alcohol_white
               quality
Evaluation mode:    evaluate on all training data
```

=== Attribute Selection on all input data ===

```
Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 122
  Merit of best subset found:    0.204
```

```
Attribute Subset Evaluator (supervised, Class (numeric): 23 quality):
  CFS Subset Evaluator
  Including locally predictive attributes
```

```
Selected attributes: 22 : 1
                     alcohol_white
```

No *DataSet* analisado, que contem as informações sobre o vinho tinto e vinho branco distinguidos por 22 atributos, como podemos ver em cima, através da análise do script fornecido pelo *WEKA*, só nos é fornecido um atributo que é o **alcohol_white**. Desde logo, estranhamos este resultado e decidimos verificar na linguagem **R** este resultado.

3.4.2 Script - R

```
Subset selection object
Call: regsubsets.formula(formula_A, data, method = "backward")
22 Variables (and intercept)

              Forced in Forced out
fixed_acidity_red      FALSE      FALSE
volatile_acidity_red   FALSE      FALSE
citric_acid_red        FALSE      FALSE
residual_sugar_red     FALSE      FALSE
chlorides_red          FALSE      FALSE
free_sulfur_dioxide_red FALSE      FALSE
total_sulfur_dioxide_red FALSE      FALSE
density_red           FALSE      FALSE
pH_red                FALSE      FALSE
sulphates_red          FALSE      FALSE
alcohol_red            FALSE      FALSE
fixed_acidity_white    FALSE      FALSE
volatile_acidity_white  FALSE      FALSE
citric_acid_white      FALSE      FALSE
residual_sugar_white   FALSE      FALSE
chlorides_white        FALSE      FALSE
free_sulfur_dioxide_white FALSE      FALSE
total_sulfur_dioxide_white FALSE      FALSE
density_white          FALSE      FALSE
pH_white               FALSE      FALSE
sulphates_white        FALSE      FALSE
alcohol_white          FALSE      FALSE

1 subsets of each size up to 8
Selection Algorithm: backward

      fixed_acidity_red volatile_acidity_red citric_acid_red residual_sugar_red chlorides_red
1 ( 1 ) " "          " "                " "                " "                " "
2 ( 1 ) " "          " "                " "                " "                " "
3 ( 1 ) " "          " "                " "                " "                " "
4 ( 1 ) " "          " "                " "                " "                " "
5 ( 1 ) " "          " "                " "                " "                " "
6 ( 1 ) " "          " "                " "                " "                " "
7 ( 1 ) " "          " "                " "                " "                " "
8 ( 1 ) " "          " "                " "                " "                " "

      free_sulfur_dioxide_red total_sulfur_dioxide_red density_red pH_red sulphates_red
1 ( 1 ) " "                " "                " "                " "                " "
2 ( 1 ) " "                " "                " "                " "                " "
3 ( 1 ) " "                " "                " "                " "                " "
4 ( 1 ) " "                " "                " "                " "                "*"
5 ( 1 ) " "                " "                " "                " "                "*"
6 ( 1 ) " "                " "                " "                " "                "*"
7 ( 1 ) " "                " "                " "                " "                "*"
8 ( 1 ) " "                " "                " "                " "                "*"

      alcohol_red fixed_acidity_white volatile_acidity_white citric_acid_white
1 ( 1 ) " "          " "                " "                " "
```



```

2 ( 1 ) "*"      " "      " "      " "
3 ( 1 ) "*"      " "      " "      " "
4 ( 1 ) "*"      " "      " "      " "
5 ( 1 ) "*"      " "      " "      " "
6 ( 1 ) "*"      " "      " "      " "
7 ( 1 ) "*"      " "      " "      " "
8 ( 1 ) "*"      " "      " "      " "

      residual_sugar_white chlorides_white free_sulfur_dioxide_white total_sulfur_dioxide_white
1 ( 1 ) " "      " "      " "      " "
2 ( 1 ) " "      " "      " "      " "
3 ( 1 ) " "      " "      "*"      " "
4 ( 1 ) " "      " "      "*"      " "
5 ( 1 ) "*"      " "      "*"      " "
6 ( 1 ) "*"      " "      "*"      "*"
7 ( 1 ) "*"      " "      "*"      "*"
8 ( 1 ) "*"      " "      "*"      "*"

      density_white pH_white sulphates_white alcohol_white
1 ( 1 ) " "      " "      " "      "*"
2 ( 1 ) " "      " "      " "      "*"
3 ( 1 ) " "      " "      " "      "*"
4 ( 1 ) " "      " "      " "      "*"
5 ( 1 ) " "      " "      " "      "*"
6 ( 1 ) " "      " "      " "      "*"
7 ( 1 ) " "      " "      "*"      "*"
8 ( 1 ) " "      "*"      "*"      "*"

```

Neste caso, verificamos que os principais atributos são o: **sulphates_red**, o **alcohol_red**, o **free_sulfur_dioxide_white** e o **alcohol_white**. Após esta análise, decidimos então, carregar os dados de aprendizagem para a rede neuronal que os irá processar, com uma fórmula composta por estes atributos.

Capítulo 4

Topologias, aprendizagem e testes das RNA's

Depois do estudo dos atributos, com as fórmulas criadas, falta-nos agora identificar a topologia de rede mais adequada. Decidimos utilizar o algoritmo construtivo para isto. De seguida encontram-se *scripts* utilizados de forma a obter os resultados de alguns dos muitos testes realizados. O valor do threshold utilizado era sempre o mais baixo possível, de modo a conseguir realizar o treino da rede neuronal. Utilizamos o *root-mean-square deviation* (RMSE) para avaliação da qualidade da rede neuronal. O valor da RMSE apresentado nas tabelas é sempre o mais baixo que obtivemos nos testes, à exceção de *outliers* que foram excluídos, pois nunca conseguíamos replica-los, isto devido à natureza das redes neuronais.

4.1 Opção A - 2 RNA's abastecidas com 2 DataSets

4.1.1 Script em R - Vinho Tinto

```
# Load needed packages
library(neuralnet)
library(hydroGOF)
library(arules)
library(leaps)

# Read file to memory
data <- read.csv("~/Desktop/SRCR/Trabalhos/Local/TP3/Dados/winequality-red.csv", header=TRUE, sep=";",

# Mix data
#data <- data[sample(nrow(data)), ]

# Training set
training <- data[1:1000,]

# Testing set
test <- data[1001:1599,]

# -----
#           Teste A           -> Weka attributes
# -----
```

```

# Formula to calculate quality of the wines
formula_A <- quality ~ fixed_acidity + volatile_acidity + citric_acid + residual_sugar + chlorides +

# See the most important atributes
selecao_A <- regsubsets(formula_A,data,method="backward")
summary(selecao_A)

# Formula to calculate quality of the wines with weka'a most important atributes
formula_Weka <- quality ~ volatile_acidity + citric_acid + sulphates + alcohol

# Train neural network to predict Quality of the wines
rna_Weka <- neuralnet(formula_Weka, training, hidden = c(6,4,2), lifesign = "full", linear.output =

# Plot neural network
plot(rna_Weka, rep = "best")

# Define input and test variables
test_Weka <- subset(test, select=c("volatile_acidity","citric_acid" , "sulphates" ,"alcohol"))

# Test the neural net with new cases
rna_Weka.results <- compute(rna_Weka, test_Weka)

# Print the results
results_Weka <- data.frame(current = test$quality, prediction = rna_Weka.results$net.result)

# Round the results
results_Weka$prediction <- round(results_Weka$prediction, digits = 0)

# Calculate RMSE
rmse(c(test$quality), c(results_Weka$prediction))

# Plot RMSE
plot(results_Weka$current, results_Weka$prediction)

# -----
#           Teste B           -> With R atributtes
# -----

# Formula to calculate quality of the wines with weka'a most important atributes
formula_R <- quality ~ volatile_acidity + chlorides + sulphates + alcohol

# Train neural network to predict Quality
rna_R <- neuralnet(formula_R, training, hidden = c(5,3), lifesign = "full", linear.output = TRUE, th

# Plot neural network
plot(rna_R, rep = "best")

# Define input and test variables
test_R <- subset(test, select = c("volatile_acidity","chlorides","sulphates","alcohol"))

# Test the neural net with new cases
rna_R.results <- compute(rna_R, test_R)

```

```

# Print the results
results_R <- data.frame(current = test$quality, prediction = rna_R.results$net.result)

# Round the results
results_R$prediction <- round(results_R$prediction, digits = 0)

# Calculate RMSE
rmse(c(test$quality), c(results_R$prediction))

# Plot RMSE
plot(results_R$current, results_R$prediction)

```

4.1.2 Script em R - Vinho Branco

```

# Load needed packages
library(neuralnet)
library(hydroGOF)
library(arules)
library(leaps)
# Read file to memory
data <- read.csv("~/Desktop/SRCR/Trabalhos/Local/TP3/Dados/winequality-white.csv", header=TRUE, sep=";")

# Mix data
#data <- data[sample(nrow(data)), ]

# Training set
training <- data[1:3000,]

# Testing set
test <- data[3001:4898,]

# -----
#           Teste A           -> All the atributes
# -----

# Formula to calculate quality of the wines
formula_A <- quality ~ fixed_acidity + volatile_acidity + citric_acid + residual_sugar + chlorides +

# See the most important atributes
selecao_A <- regsubsets(formula_A,data,method="backward")
summary(selecao_A)

# Formula to calculate quality of the wines with weka'a most important atributes
formula_Weka <- quality ~ volatile_acidity + chlorides + pH + sulphates + alcohol

# Train neural network to predict Quality of the wines
rna_Weka <- neuralnet(formula_Weka, training, hidden = c(6,4,2), lifesign = "full", linear.output =

# Plot neural network
plot(rna_Weka, rep = "best")

```

```

# Define input and test variables
test_Weka <- subset(test, select=c("volatile_acidity","chlorides" , "pH" , "sulphates" ,"alcohol"))

# Test the neural net with new cases
rna_Weka.results <- compute(rna_Weka, test_Weka)

# Print the results
results_Weka <- data.frame(current = test$quality, prediction = rna_Weka.results$net.result)

# Round the results
results_Weka$prediction <- round(results_Weka$prediction, digits = 0)

# Calculate RMSE
rmse(c(test$quality), c(results_Weka$prediction))

# Plot RMSE
plot(results_Weka$current, results_Weka$prediction)

# -----
#           Teste B           -> With R atributtes
# -----

# Formula to calculate quality of the wines with weka'a most important atributes
formula_R <- quality ~ volatile_acidity + residual_sugar + density + alcohol

# Train neural network to predict Quality
rna_R <- neuralnet(formula_R, training, hidden = c(5,3), lifesign = "full", linear.output = TRUE, th

# Plot neural network
plot(rna_R, rep = "best")

# Define input and test variables
test_R <- subset(test, select = c("volatile_acidity","residual_sugar","density","alcohol"))

# Test the neural net with new cases
rna_R.results <- compute(rna_R, test_R)

# Print the results
results_R <- data.frame(current = test$quality, prediction = rna_R.results$net.result)

# Round the results
results_R$prediction <- round(results_R$prediction, digits = 0)

# Calculate RMSE
rmse(c(test$quality), c(results_R$prediction))

# Plot RMSE
plot(results_R$current, results_R$prediction)

```

4.1.3 Conclusões e Resultados Obtidos

Previsão da Qualidade do Vinho Tinto

| Vinho Tinto - WEKA | | |
|--------------------|-----------|--------------|
| Camada Intermédia | Threshold | RMSE |
| 1 | 0,001 | 0.7100518451 |
| 2 | 0,001 | 0.7041493998 |
| 3 | 0,05 | 0.7100518451 |
| 4 | 0,05 | 0.7017745216 |
| 5 | 0,1 | .7205544146 |

Figura 4.1: Valor do RMSE obtido para os dados do weka.

| Vinho Tinto - R | | |
|-------------------|-----------|--------------|
| Camada Intermédia | Threshold | RMSE |
| 1 | 0,001 | 0.7159056279 |
| 2 | 0,001 | 0.7286184242 |
| 3 | 0,001 | 0.7123991303 |
| 4 | 0,005 | 0.7217119325 |
| 5 | 0,05 | 0.7217119325 |

Figura 4.2: Valor do RMSE obtido para os dados do R.

Previsão da Qualidade do Vinho Branco

| Vinho Branco - WEKA | | |
|---------------------|-----------|--------------|
| Camada Intermédia | Threshold | RMSE |
| 1 | 0,01 | 0.8240970407 |
| 2 | 0,2 | 0.7756842046 |
| 3 | 0,5 | 0.7898193095 |
| 4 | 0,6 | 0.7837926222 |
| 5 | 0,6 | 0.7954692566 |

Figura 4.3: Valor do RMSE obtido para os dados do weka.

| Vinho Branco - R | | |
|-------------------|-----------|--------------|
| Camada Intermédia | Threshold | RMSE |
| 1 | 0,1 | 0.7705731175 |
| 2 | 0,2 | 0.7657719921 |
| 3 | 0,2 | 0.7794110338 |
| 4 | 0,2 | 0.7800867262 |
| 5 | 0,5 | 0.7854713404 |

Figura 4.4: Valor do RMSE obtido para os dados do R.

4.2 Opção B - 1 RNA abastecida por um DataSet de 12 atributos

4.2.1 Script em R

```
# Load needed packages
library(neuralnet)
library(hydroGOF)
library(arules)
library(leaps)

# Read file to memory
data <- read.csv("~/Desktop/SRCR/Trabalhos/Local/TP3/Dados/winequality-all-type.csv", header=TRUE, sep=";")

# Mix data
#data <- data[sample(nrow(data)), ]

# Training set
training <- data[1:5000,]

# Testing set
test <- data[5001:6497,]

# -----
#           Teste A           -> Weka attributes
# -----

# Formula to calculate quality of the wines
formula_A <- quality ~ fixed_acidity + volatile_acidity + citric_acid + residual_sugar + chlorides + ash + potassium + sulphates + total_sulphates + free_sulphates + total_acidity + total_alkalinity

# See the most important attributes
selecao_A <- regsubsets(formula_A, data, method="backward")
summary(selecao_A)

# Formula to calculate quality of the wines with weka's most important attributes
formula_Weka <- quality ~ citric_acid + alcohol + type

# Train neural network to predict Quality of the wines
rna_Weka <- neuralnet(formula_Weka, training, hidden = c(6,4,2), lifesign = "full", linear.output = 1)

# Plot neural network
plot(rna_Weka, rep = "best")

# Define input and test variables
test_Weka <- subset(test, select=c("citric_acid" , "alcohol" , "type"))

# Test the neural net with new cases
rna_Weka.results <- compute(rna_Weka, test_Weka)

# Print the results
results_Weka <- data.frame(current = test$quality, prediction = rna_Weka.results$net.result)

# Round the results
results_Weka$prediction <- round(results_Weka$prediction, digits = 0)
```

```

# Calculate RMSE
rmse(c(test$quality), c(results_Weka$prediction))

# Plot RMSE
plot(results_Weka$current, results_Weka$prediction)

# -----
#           Teste B           -> With R atributtes
# -----

# Formula to calculate quality of the wines with weka'a most important atributes
formula_R <- quality ~ free_sulfur_dioxide + sulphates + alcohol + type

# Train neural network to predict Quality
rna_R <- neuralnet(formula_R, training, hidden = c(5,3), lifesign = "full", linear.output = TRUE, th

# Plot neural network
plot(rna_R, rep = "best")

# Define input and test variables
test_R <- subset(test, select = c("free_sulfur_dioxide" , "sulphates","alcohol" , "type"))

# Test the neural net with new cases
rna_R.results <- compute(rna_R, test_R)

# Print the results
results_R <- data.frame(current = test$quality, prediction = rna_R.results$net.result)

# Round the results
results_R$prediction <- round(results_R$prediction, digits = 0)

# Calculate RMSE
rmse(c(test$quality), c(results_R$prediction))

# Plot RMSE
plot(results_R$current, results_R$prediction)

```

4.2.2 Conclusões e resultados obtidos

| Vinhos distinguidos por tipo - WEKA | | |
|-------------------------------------|-----------|--------------|
| Camada Intermédia | Threshold | RMSE |
| 1 | 0,1 | 0.8509494089 |
| 2 | 0,1 | 0.8458314618 |
| 3 | 0,1 | 0.8575960249 |
| 4 | 0,1 | 0.8290806653 |
| 5 | 0,5 | 0.820577049 |

Figura 4.5: Valor do RMSE obtido para os dados do weka.

| Vinhos distinguidos por tipo - R | | |
|----------------------------------|-----------|--------------|
| Camada Intermédia | Threshold | RMSE |
| 1 | 0,01 | 0.8672779305 |
| 2 | 0,1 | 0.8665073591 |
| 3 | 0,1 | 0.8450413341 |
| 4 | 0,5 | 0.8414765786 |
| 5 | 0,5 | 0.8478035593 |

Figura 4.6: Valor do RMSE obtido para os dados do R.

4.3 Opção C - 1 RNA abastecida por um DataSet de 22 atributos

4.3.1 Script em R

```
# Load needed packages
library(neuralnet)
library(hydroGOF)
library(arules)
library(leaps)

# Read file to memory
data <- read.csv("~/Desktop/SRCR/Trabalhos/Local/TP3/Dados/winequality-all-22.csv", header=TRUE, sep=";")

# Mix data
#data <- data[sample(nrow(data)), ]

# Training set
training <- data[1:3000,]

# Testing set
test <- data[3001:4898,]

# -----
# Teste A -> With R attributes
# -----

# Formula to calculate Quality
formula_A <- quality ~ fixed_acidity_red + volatile_acidity_red + citric_acid_red + residual_sugar_red

# See the most important attributes
selecao_A <- regsubsets(formula_A, data, method="backward")
summary(selecao_A)

# Formula to calculate quality of the wines with weka's most important attributes
formula_R <- quality ~ sulphates_red + alcohol_red + free_sulfur_dioxide_white + alcohol_white

# Train neural network to predict Quality
rna_R <- neuralnet(formula_R, training, hidden = c(5,3), lifesign = "full", linear.output = TRUE, threshold = 0.01)

# Plot neural network
plot(rna_R, rep = "best")
```

```

# Define input and test variables
test_R <- subset(test, select = c("sulphates_red" , "alcohol_red" , "free_sulfur_dioxide_white" , "a

# Test the neural net with new cases
rna_R.results <- compute(rna_R, test_R)

# Print the results
results_R <- data.frame(current = test$quality, prediction = rna_R.results$net.result)

# Round the results
results_R$prediction <- round(results_R$prediction, digits = 0)

# Calculate RMSE
rmse(c(test$quality), c(results_R$prediction))

# Plot RMSE
plot(results_R$current, results_R$prediction)

```

4.3.2 Conclusões e resultados obtidos

| Vinhos distinguidos por 22 atributos - R | | |
|--|-----------|--------------|
| Camada Intermédia | Threshold | RMSE |
| 1 | 0,1 | 0.8560367581 |
| 2 | 0,1 | 0.840682358 |
| 3 | 0,5 | 0.840682358 |
| 4 | 0,5 | 0.848591115 |
| 5 | 0,5 | 0.8474095069 |

Figura 4.7: Valor do RMSE obtido para os dados do R.

Capítulo 5

Conclusão

blblblblbl