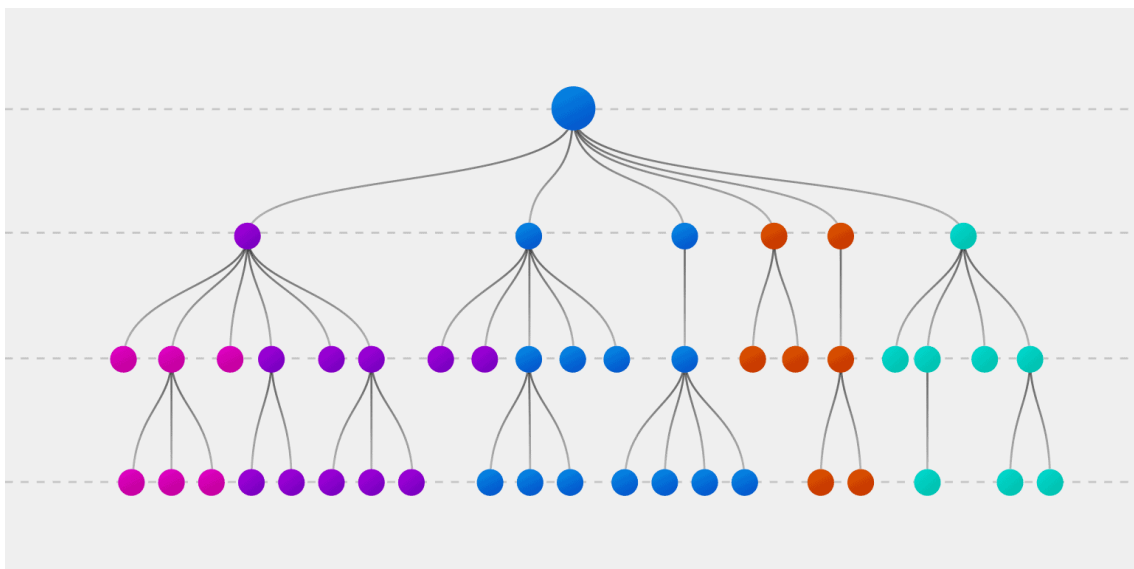




Universidade de Évora
Curso de Engenharia Informática
Aprendizagem Automática 2019/2020

Trabalho 1



Trabalho realizado por:

Dinis Matos nº42738

José Lopes nº37861

Introdução

Este trabalho foi realizado com o objetivo de gerar uma árvore de decisão usando o algoritmo ID3 e o método REP (Reduced Error Pruning).

Estrutura do trabalho

O programa começa com a abertura e leitura do input a ser analisado (ficheiro .csv) através da função "openData". Depois da leitura, vão ser divididos os dados para o treino e a geração da árvore de decisão.

A árvore de decisão vai ser composta por nodes que têm 5 elementos - "key" (guarda a chave do dicionário correspondente ao elemento que o node representa), "child" (guarda a informação de todos os filhos desse node), "father" (guarda a informação do pai desse node), "results" (guarda os resultados possíveis desse node) e "path" (variável que contém o caminho até esse mesmo node).

De seguida vai ser chamada a classe "DecisionTreeREPrune" - este método vai iniciar uma instância da árvore de decisão. Depois vai ser executada a função "fit" que vai treinar os dados e vai popular a árvore instanciada. Dentro desse método vai ser chamado o método "pruning" que vai fazer pruning na árvore se a variável global "PRUNE" assim o indicar.

Finalmente o método "score" vai ser executado, que vai calcular a exatidão obtida pela árvore gerada.

No ficheiro "ExtraFunctions.py" estão inseridas funções auxiliares que fazem diversas tarefas e calculo de contas ao longo da execução do programa. Funções como "errorReason", "gini" e "entropy" que calculam a Razão de Erro, o Índice de Gini e a Entropia, respetivamente. Existem também diversas outras funções que ajudam na iteração de arrays e de listas, e até mesmo na criação e geração da árvore de decisão.

Input/Output

Analisando os dados após 20 simulações aleatórias, eis a média da exatidão obtida para cada um dos inputs dados:

ENTROPIA

```
"weather.nominal - PRUNING OFF" - 61.25%
"weather.nominal - PRUNING ON" - 65%

"contact-lenses - PRUNING OFF" - 70%
"contact-lenses - PRUNING ON" - 75%

"vote - PRUNING OFF" - 95.74%
"vote - PRUNING ON" - 95.09%

"soybean - PRUNING OFF" - 86.81%
"soybean - PRUNING ON" - 84.22%
```

GINI

```
"weather.nominal - PRUNING OFF" - 67.5%
"weather.nominal - PRUNING ON" - 76.25%

"contact-lenses - PRUNING OFF" - 70.83%
"contact-lenses - PRUNING ON" - 73.33%

"vote - PRUNING OFF" - 94.14%
"vote - PRUNING ON" - 95.34%

"soybean - PRUNING OFF" - 83.51%
"soybean - PRUNING ON" - 81.31%
```

RAZÃO DE ERRO

```
"weather.nominal - PRUNING OFF" - 75%
"weather.nominal - PRUNING ON" - 78.75%

"contact-lenses - PRUNING OFF" - 67.5%
"contact-lenses - PRUNING ON" - 66.67%

"vote - PRUNING OFF" - 93.88%
"vote - PRUNING ON" - 95.17%

"soybean - PRUNING OFF" - 49.86%
"soybean - PRUNING ON" - 46.28%
```

É de notar que, nos inputs "weather.nominal" e "contact-lenses", o pruning foi benéfico pois aumentou a exatidão, no input "vote" não mudou assim tanto a exatidão e no input "soybean" piorou-a ligeiramente.