

## Exercícios sobre Árvores de Decisão

1. Para cada uma das funções booleanas seguintes apresente uma árvore de decisão que as representem:

a)  $A \wedge \neg B$

b)  $(A \wedge B) \vee (C \wedge D)$

2. Construa o conjunto completo de árvores de decisão a partir dos dados seguintes:

A1	A2	Y
0	1	0
1	0	1
0	0	0
1	1	1

3. Suponha que se pretende construir a árvore de decisão, através do algoritmo ID3, com base no conjunto de dados apresentados tabela seguinte

A1	A2	A3	Y
0	1	1	1
1	0	1	1
0	0	0	0
1	1	1	0

- a) Calcule a entropia média de cada um dos atributos  
 b) Qual dos atributos seria escolhido para dividir os dados?  
 c) Construa a árvore de decisão.
4. Considere o seguinte conjunto de exemplos de treino:

Exemplo	A1	A2	Classe
1	escuro	alto	-
2	claro	alto	+
3	claro	baixo	-
4	escuro	baixo	+
5	escuro	alto	-
6	claro	alto	+

- a) Calcule a informação esperada ou entropia deste conjunto de exemplos e o ganho de informação relativo ao atributo A2. A2 poderá ser raiz da árvore de decisão obtida pelo algoritmo ID3?  
 b) Apresente a árvore de decisão completa que seria produzida pelo algoritmo ID3.  
 c) Calcule a informação esperada ou entropia deste conjunto de exemplos e o ganho de informação relativo ao atributo A2. A2 poderá ser raiz da árvore de decisão obtida pelo algoritmo ID3?  
 d) Apresente a árvore de decisão completa que seria **produzida** pelo algoritmo ID3.

5. Considere o conjunto de dados seguinte:

A1	A2	Y
0	0	1
0	0	0
0	0	1
0	1	0
0	1	0
0	1	1
1	0	1
1	0	0
1	0	1
1	1	1
1	1	1
1	1	1

- É possível obter uma hipótese sem nenhum erro a partir destes dados?
  - Qual será a entropia média se escolhermos o atributo A1?
  - E se escolhermos A2?
  - Qual é a árvore obtida, tendo em conta que se as folhas não puderem ser uniformes, se escolhe como valor a classe mais comum?
  - Essa árvore poderia ainda ser simplificada?
6. Imagine que quer utilizar o algoritmo ID3 para aprender uma função e que lhe são apresentados um conjunto de exemplos e de contra-exemplos:

A1	A2	A3	A4	A5
0	1	1	0	0
1	0	1	0	0
1	1	1	0	1
0	0	0	1	1
1	0	0	1	0
0	1	0	1	0

- Construa a árvore de decisão
  - Como ficaria a árvore de decisão caso o processo de sub-divisão da árvore pare sempre que não se puder melhorar o ganho médio e se escolher para valor das folhas a classe mais comum?
7. Os candidatos a alunos de doutoramento na fictícia Universidade da Martinlândia baseia-se em quatro critérios: a nota de fim de curso, o ranking da universidade onde realizou o curso, o registo de publicações e as cartas de recomendação. Para simplificar a nota pode tomar três valores, que são 4.0, 3.7 e 3.5. A universidade pode ser classificada entre as 10 melhores (top-10), entre as 10 e as 20 melhores (top-20) e entre as 20 e 30 melhores (top-30). O registo de publicações é um atributo binário – o candidato publicou ou não; e as cartas de recomendação podem ser boas ou normais. Finalmente, os candidatos podem ser classificados como aceites (A) ou rejeitados (R). A tabela seguinte mostra um conjunto de exemplos de candidatos a doutoramento e a respectiva classificação.

Nota	Ranking	Publicou	Recomendação	Classe
4.0	top-10	sim	boa	A
4.0	top-10	não	boa	A
4.0	top-20	não	normal	A
3.7	top-10	sim	boa	A
3.7	top-20	não	boa	R
3.7	top-30	sim	boa	A
3.7	top-30	não	boa	R
3.7	top-10	não	boa	R
3.5	top-20	sim	normal	R
3.5	top-10	não	normal	R
3.5	top-30	sim	normal	R
3.5	top-30	não	boa	R

Apresente a árvore de decisão completa que seria produzida pelo algoritmo ID3.

8. Considere o problema de esperar ou não esperar por uma mesa num restaurante. O objectivo é aprender uma definição para o objectivo *Esperar*, sendo essa definição expressa sob a forma de uma árvore de decisão. Existem os seguintes atributos para descrever as situações exemplo:

*Alternativa*: existe ou não um restaurante na vizinhança que seja uma alternativa.

*Bar*: O restaurante tem ou não tem um bar.

*Sexta/Sábado*: Se esse dia é uma sexta ou um sábado.

*Fome*: estamos ou não com fome.

*Clientes*: Quantas pessoas estão no restaurante (nenhuma, algumas, cheio)

*Preço*: Três preços possíveis: (\$, \$\$, \$\$\$)

*Chuva*: Está a chover lá fora ou não.

*Reserva*: Foi feita uma reserva ou não.

*Tipo*: O tipo de restaurante: (francês, italiano, tailandês e hamburgueria)

*Estimativa do tempo de espera*: (0-10 minutos, 10-30, 30-60, >60).

Alter	Bar	Sexta	Fome	Clientes	Preço	Chuva	Reserva	Tipo	TmpEsp	Esperar
Sim	Não	Não	Sim	Alguns	\$\$\$	Não	Sim	Francês	0-10	Sim
Sim	Não	Não	Sim	Cheio	\$	Não	Não	Tailandês	30-60	Não
Não	Sim	Não	Não	Alguns	\$	Não	Não	Hamburg	0-10	Sim
Sim	Não	Sim	Sim	Cheio	\$	Não	Não	Tailandês	10-30	Sim
Sim	Não	Sim	Não	Cheio	\$\$\$	Não	Sim	Francês	>60	Não
Não	Sim	Não	Sim	Alguns	\$\$	Sim	Sim	Italiano	0-10	Sim
Não	Sim	Não	Não	Nenhuns	\$	Sim	Não	Hamburg	0-10	Não
Não	Não	Não	Sim	Alguns	\$\$	Sim	Sim	Tailandês	0-10	Sim
Não	Sim	Sim	Não	Cheio	\$	Sim	Não	Hamburg	>60	Não
Sim	Sim	Sim	Sim	Cheio	\$\$\$	Não	Sim	Italiano	10-30	Não
Não	Não	Não	Não	Nenhuns	\$	Não	Não	Tailandês	0-10	Não
Sim	Sim	Sim	Sim	Cheio	\$	Não	Não	Hamburg	30-60	Sim

- a) Utilize a ferramenta Weka, através do algoritmo ID3, para construir a árvore que representa estes a partir destes dados.
  - b) Assumindo que D1 e D2 são árvores de decisão representando funções booleanas, e que D2 é considerada uma elaboração de D1 se o algoritmo ID3 pode estender D1 em D2, indique se a frase seguinte é verdadeira ou falsa: Se a árvore D2 é uma elaboração de D1, então D1 é mais geral do que D2. Se considerar que é verdadeira, prove-o; se considerar que é falsa, apresente um contra-exemplo.
9. O algoritmo ID3 encontra apenas uma hipótese consistente enquanto o algoritmo do espaço de versões (também chamado algoritmo de eliminação de candidatos) encontra todas as hipóteses consistentes. Considere a correspondência entre esses 2 algoritmos:
- a) apresente os resultados obtidos por cada um destes dois algoritmos a partir dos exemplos de treino seguintes, para o conceito alvo fazDesporto:

Exemplos de treino:

	Céu	Temp.	Hum.	Prev	fazDesporto
0	sol	quente	normal	igual	sim
1	sol	quente	alta	igual	sim
2	chuva	frio	alta	muda	não
3	sol	quente	alta	muda	sim

- b) Qual é a relação entre a árvore de decisão aprendida e o espaço de versões obtido? A árvore de decisão é equivalente a algum dos membros do espaço de versões?
  - c) Suponha que quer definir um algoritmo de aprendizagem que, como o ID3, efectua uma procura no espaço das árvores de decisão e, como o espaço de versões, encontra todas as hipóteses consistentes com os dados. Isto é, pretende-se aplicar o algoritmo do espaço de versões à procura num espaço de hipóteses em que as hipóteses são árvores de decisão. Apresente os conjuntos S e G que resultam do 1º exemplo de treino dado. Mostre como é que S e G seriam refinados pelo segundo exemplo de treino (pode omitir árvores sintacticamente distintas que representem o mesmo conceito). Que dificuldades antevê na aplicação do espaço de versões a espaços de hipóteses de árvores de decisão?
10. Use os dados da tabela 1 como conjunto de treino para aprender a classificar, segundo as 3 classes indicadas no atributo “Queimadura”.
- a. Calcule a entropia (ou informação esperada) do conjunto de treino. Calcule o ganho de informação (“gain”) relativo a cada um dos atributos e indique o atributo que seria escolhido, de acordo com essa medida, para raiz da árvore de decisão pelo algoritmo ID3. Comente o resultado obtido.
  - b. Eliminando agora o atributo “Nome”, e continuando a usar a medida de ganho de informação, determine a árvore de decisão completa que seria produzida pelo ID3.

- c. Usando como medida a razão de ganhos (“gain ratio”) e considerando todos os atributos (incluindo “Nome”), qual seria o atributo escolhido para raiz da árvore de decisão?

Nome	Cabelo	Pele	LoçãoSolar	Queimadura
Emília	Castanho	Morena	Não	Sem
Sara	Louro	Morena	Não	ligeira
Diana	Louro	Morena	Sim	Sem
Andreia	Louro	Branca	Sim	Sem
Leonor	Louro	Branca	Não	grave
Emília	Ruivo	Branca	Sim	grave
Diana	Castanho	Branca	Não	Sem
Fernão	Ruivo	Morena	Não	ligeira
Carlos	Ruivo	Morena	Sim	ligeira
Joana	Castanho	Branca	Sim	Sem

Tabela 1: Dados relativos a queimaduras solares

11. Políbio, na sua apreciação das refeições em restaurantes, considera os seguintes atributos e respectivos valores possíveis:

Restaurante: {Copélia, Palma, Primavera}

Qualidade: {boa, má}

Preço: um inteiro

Refeição: {almoço, jantar, pequeno\_almoço}

Políbio usa afirmações em português para exprimir os seus processos de classificação das refeições como satisfatórias ou não satisfatórias, em lugar de usar árvores de decisão. Por exemplo, ele diria:

“Fico satisfeito com qualquer refeição de 10€ ou menos, mas não existe nenhuma refeição do restaurante Palma por 7€ ou menos que me agrade.”

Afirmarões como esta podem traduzir-se em diferentes árvores de decisão. Uma das árvores que poderíamos indicar para a frase acima é a seguinte, podendo haver outras com diferente número de nós, possivelmente testando os atributos por ordem diferente ou realizando diferentes testes:

Indique árvores de decisão, com número mínimo de nós, correspondentes a cada uma das seguintes afirmações:

- a) “Fico satisfeito com qualquer refeição de boa qualidade que não custe mais do que 10€ e satisfeito com refeições de má qualidade que não custem mais do que 5€.”

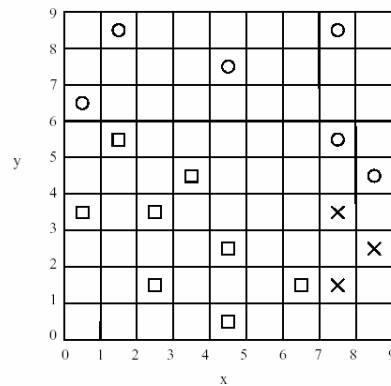
- b) Fico satisfeito com um pequeno almoço de 8€ ou menos ou com um jantar de 15€ ou menos. Nunca fico satisfeito com uma refeição de má qualidade nem se tenho de pagar por um almoço (já que almoço bem e de borla)” .

12. Imagine que temos o seguinte conjunto de dados, onde Y é o atributo alvo da classificação.

A	B	C	Y
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

Qual a árvore de decisão escolhida pelo algoritmo ID3? Não existirá outra árvore que envolva testar menos atributos (e possivelmente com menor profundidade) capaz de representar correctamente os dados? Qual a justificação que encontra para que o ID3 não devolva essa árvore?

13. Considere os dados apresentados na quadrícula abaixo, descritos por dois atributos (eixos x e y, com valores inteiros entre 0 e 8) e classificados em 3 classes, representadas por quadrados, cruzes ou círculos. Usando o algoritmo ID3 estendido para fazer testes binários sobre atributos numéricos, construa a árvore de decisão de forma a classificar correctamente os dados.



14. A tabela de dados seguinte baseia-se no livro do Tolkien “O senhor dos anéis”. A tabela apresenta informação sobre um conjunto de pessoas/entidades que aparecem nesse livro.

Nome	Raça	Peso	Senhor-do-anel	Classe
Frodo	Hobbit	Leve	Sim	Bom
Rosie	Hobbit	Leve	Não	Bom
Bilbo	Hobbit	Leve	Sim	Bom
Gollum	Hobbit	Leve	Sim	Mau
Faramir	Humano	Médio	Não	Bom
Aragorn	Humano	Médio	Não	Bom
Wormtongue	Humano	Médio	Não	Mau
Celeborn	Elves	Leve	Não	Bom
Galadriel	Elves	Leve	Sim	Bom
Sharku	Orc	Pesado	Não	Mau

Cada entidade possui três atributos (Raça, Peso, Senhor-anel) e é classificado como sendo Bom ou Mau (i.e., se quer ou não quer matar o Frodo). O atributo "Senhor-anel" indica se a pessoa/entidade alguma vez possuiu o anel mágico. Aplique o algoritmo ID3 aos dados na tabela, tendo em conta que os nós folhas são classificados com a classe maioritária. No entanto, introduza, uma pequena variação na qual se expandem os nós apenas quando resultar numa melhoria do ganho.

15. Considere o seguinte conjunto de dados, onde Y corresponde ao atributo classe. Vamos considerar formas de poda da árvore de decisão produzida pelo ID3 que não envolvem o uso de um conjunto de teste.

V	W	X	Y
0	0	0	0
0	1	0	1
1	0	0	1
1	1	0	0
1	1	1	0

- Apresente a árvore de decisão que seria construída pelo ID3, sem poda.
- Uma possível forma de podar a árvore consiste em, começando pelo nó raiz da árvore, podar a subárvore com origem num nó se o ganho de informação (ou outro critério dado) associado a esse nó for inferior a uma pequena quantidade  $\alpha$ . Este tipo de poda é chamada *poda descendente* ("top-down pruning"). Qual é a árvore de decisão retornada aplicando este tipo de poda com  $\alpha=0.0001$ ? Qual é o erro produzido por essa árvore podada para o conjunto de treino dado (% de exemplos mal classificados)?
- Outra forma possível de podar a árvore consiste em, começando pelos nós pais de folhas da árvore, podar subárvores com origem num nó se o ganho de informação (ou outro critério dado) for inferior a uma pequena quantidade  $\alpha$ . Segundo este método, nenhum antepassado de um nó com alto ganho de informação é podado. Este tipo de poda é chamada *poda ascendente* ("bottom-up pruning"). Qual é a árvore de decisão retornada aplicando este tipo de poda com  $\alpha=0.0001$ ? Qual é o erro produzido por essa árvore podada para o conjunto de treino dado (% de exemplos mal classificados)?
- discuta as possíveis vantagens e desvantagens destes dois tipos de poda, tendo em conta por exemplo a complexidade computacional envolvida e a precisão da classificação.