**Exercícios: árvores de decisão**

1. Considere o seguinte conjunto de treino.



* 1. Qual é a entropia do conjunto de treino com relação à função objectivo da classificação?
  2. Qual é o ganho de informação de A2 com respeito aos exemplos de treino dados?

1. Considere o conjunto de dados a seguir e que a tarefa é construir uma árvore de decisão para classificar o atributo Y.



* 1. Qual é o ganho de informação correspondente aos atributos U, W e X?
  2. Construa a árvore de decisão utilizando o algoritmo ID3.

1. Considere o seguinte conjunto de treino.



* 1. Calcule a entropia do conjunto de treino e o ganho de informação relativo ao atributo A2.
  2. Apresente a árvore de decisão que seria produzida pelo algoritmo ID3.

1. Suponha que pretende construir uma árvore de decisão, através do algoritmo ID3, com base no conjunto de dados apresentado na tabela seguinte.



* 1. Qual dos atributos seria escolhido para dividir os dados?
  2. Construa a árvore de decisão.

1. Suponhamos que pretende utilizar o algoritmo ID3 para aprender uma função e que lhe são apresentados os seguintes exemplos e contra-exemplos.



* 1. Construa a correspondente árvore de decisão.

1. Para cada uma das funções booleanas seguintes, apresente uma árvore de decisão que a represente:
   1. 
   2. 
2. Suponhamos que se pretende classificar os dias quanto à possibilidade de jogar ou não e que possuímos os exemplos de treino mostrados na tabela abaixo. Os atributos importantes são o aspecto do dia, a temperatura, a humidade e o estado do vento.



* 1. Apresente a árvore de decisão produzida pelo algoritmo ID3.
  2. Qual seria a classificação correspondente ao vector de características mostrado a continuação?



1. Se pretende predizer se um aluno aprovará a disciplina C = Aprendizagem Automática na sua primeira convocatória, tendo em conta se aprovou ou não na primeira convocatória as disciplinas X1 = Álgebra Linear, X2 = Análise Matemática, X3 = Inteligência Artificial e X4 = Probabilidades e Estatística. A tabela seguinte apresenta 20 casos de treino; o valor 1 em qualquer das variáveis anteriores significa que passou na primeira convocatória e o valor 2 que o aluno necessitou mais de uma convocatória para aprovar.



* 1. Obter a árvore de decisão utilizando o algoritmo ID3.

1. A selecção de 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.



* 1. Apresente a árvore decisão que seria produzida pelo algoritmo ID3.
  2. Expresse-a através do correspondente conjunto de regras.

1. Utilize os dados mostrados na tabela a continuação para aprender a classificar segundo as classes indicadas no atributo "Queimadura".



* 1. Calcule a entropia do conjunto de treino.
  2. Calcule o ganho de informação relativo a cada um dos atributos. Que atributo seria escolhido pelo algoritmo ID3 para raiz da árvore de decisão?
  3. Eliminando da tabela o atributo "Nome", determine a árvore de decisão que seria produzida pelo algoritmo ID3.

1. Suponhamos que um hipotético serviço de trens desde Edimburgo até Oldfort está sujeito a atrasos frequentes. O serviço é prestado por três companhias diferentes (COT). Ao longo do ano tomou-se uma mostra aleatória do serviço prestado, obtendo-se os seguintes dados.



* 1. Determine o nodo raiz para classificar se um trem chega a tempo ou atrasado, utilizando o máximo ganho de informação. Mostre que o mesmo é constituído pela companhia operadora do trem (COT).
  2. O algoritmo ID3 cria os seguintes dois níveis da árvore. Suponha que a árvore foi podada aos referidos dois níveis. Determine a árvore de decisão final preenchendo os valores de classificação e os ratios de classificação em falta na árvore.
  3. De acordo a árvore de decisão anterior, como seriam classificados os exemplos a continuação.

