

Relatório IA P2 2020-2021

Grupo al98 Lúcia Silva 92510 & Mariana Cintrão 93737

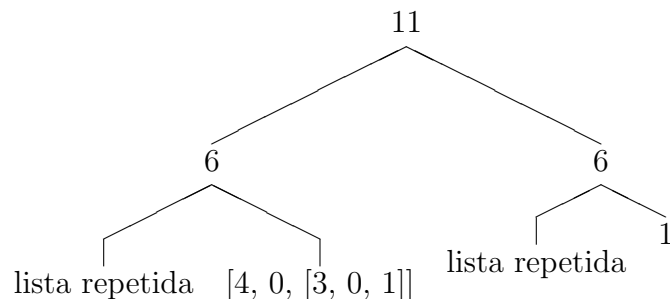
Soluções mais pequenas Ao observar os conjuntos de treino e as suas soluções verificámos dois tipos de repetições que ao serem otimizadas levariam a soluções mais pequenas. Os dois casos estão a seguir exemplificados:

- $[0, [1, [2, 0, 1], [2, 1, 0]], [1, [2, 0, 1], [2, 1, 0]]]$

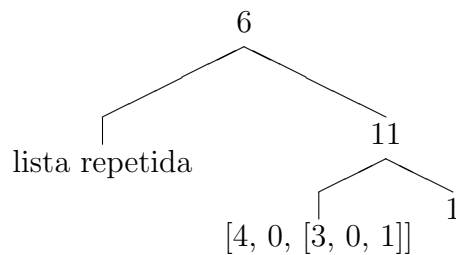
A feature 0 tem os mesmos resultados quer quando é 0 quer quando é 1 Assim, quando na chamada recursiva tentamos retornar um lista no formato $[feature, ramo-esquerdo, ramo-direito]$, em que ramo-esquerdo e ramo-direito são iguais, basta-nos apenas retornar um dos ramos para efetuar uma redução válida.

- $[11, [6, [1, 1, [3, 0, [4, 0, 1]]], [4, 0, [3, 0, 1]]], [6, [1, 1, [3, 0, [4, 0, 1]]], 1]]$

A árvore que se apresenta no exemplo é a seguinte:



Como os ramos à esquerda da feature 6 são iguais, os ramos à direita apenas dependem da feature 11 ser 0 ou 1. Pelo que, podemos reordenar esta árvore colocando a feature 6 na raiz, da seguinte forma:



Ruído Para lidar com o ruído utilizamos uma abordagem de *Top-Down Pruning* em que a árvore de decisão é também calculada recursivamente. No entanto, o aprofundamento dos ramos da árvore apenas é feito se a relevância dos nós assim o justificar. Se esta for insuficiente, o nó é substituído pela folha subjacente (se existir) ou pelo *majority vote* então calculado. A relevância é estimada a partir do valor do Ganho de Informação que consideramos insuficiente se for inferior a 0,05.

A situação mais problemática que encontramos foi o caso em que a função recursiva recebia os argumentos ilustrados em baixo:

D				Y
False	False	True	False	False
False	False	True	False	False
False	False	True	False	False
False	False	True	False	True

Como é evidente, para este ramo particular da árvore não existe um resultado único. Ao verificar que o *majority vote* é *False* devolvemos 0. Em caso de empate, o valor devolvido por *default* é 0.

Análise dos resultados Para o processamento de conjuntos de treino com ruído, a nossa solução pode induzir erros de pessimismo uma vez que, ao considerar um nó interior como irrelevante, ignora toda a árvore subjacente independentemente da sua relevância (que não chega a ser calculada). Uma abordagem que poderia ter sido tomada podia ser, por exemplo, *Bottom-Up Pruning*, que evitaria o descarte de sub-árvores relevantes. Não adotámos este método uma vez que a adaptação do algoritmo construído anteriormente nos pareceu menos intuitiva.