

DECISION TREES - Ponto 4: A ESCOLHA DO MELHOR SPLIT COM ALGORITMOS CART & C4.5

#ALGORITMO CART

VAMOS CONSIDERAR OS SEQUENTES VALORES QUE CORRESPONDEM A UMA DE NOSSAS VARIÁVEIS PREDITORAS

[7, 13, 21, 30, 35, 38, 42]

O ALGORITMO CART VAI TESTAR TODAS AS POSSIBILIDADES. POR EXEMPLO, SEGUNDO A PRIMEIRA POSSIBILIDADE, ELE VAI DEIXAR O 7 DE UM LADO E TODOS O RESTO DO OUTRO, E ASSIM POR DENTRÃO. A PARTIR DISSO, ELE VAI VER PARA CADA UM DOS SPLITS QUAL AQUELE QUE MINIMIZA A FUNÇÃO DE CUSTO

VAMOS IMAGINAR QUE NOSSO VARIÁVEL TARGET SEJA DE REGRESSÃO, COMO AS VENDAS DE UMA LOJA. ELE VAI CONSIDERAR QUE PARA TODOS OS DADOS DA ESQUERDA, A OTIMIZAÇÃO DE VENDAS É 7, ENQUANTO QUE OS DA DIREITA SÃO FEITO UMA MÉDIA PARA CALCULAR O VALOR PREVISTO. A SUBDIVISÃO QUE TIVER A MENOR FUNÇÃO DE CUSTO SERÁ ESCOLHIDA PARA SER O PONTO ONDE VAI OCORRER O SPLIT.

#ALGORITMO C4.5

ADICIONAMENTO DO ALGORITMO ID3 (QUE TRABALHA SÓ COM DADOS DE CLASSIFICAÇÃO), MAS QUE CONSEGUE TRABALHAR COM ALGORITMOS DE REGRESSÃO. ESSE ALGORITMO

__/__/__

S T Q Q S S D

Faz algo muito parecido com o CART, porém,
em vez de usar a função de custo como
tomada de decisão, ele vai utilizar o ganho
de informação.

OBS. No PYTHON, o algoritmo padrão é
o CART.