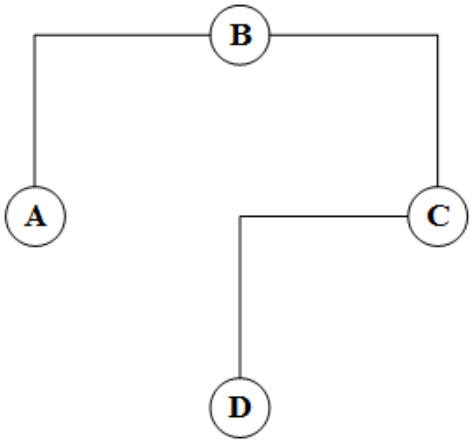
Seja A, B, C e D atributos de um conjunto de dados de exemplo.

Este conjunto de dados possui outros atributos, mas para clarificar as coisas usaremos apenas quatro.

Cada atributo deste é um valor discreto que após a análise tem sua frequência contabilizada e formar o seguinte:

$A = \{a_1, a_2\}$
$B = \{b_1, b_2, b_3, b_4, b_5\}$
$C = \{c_1, c_2, c_3\}$
$D = \{d_1, d_2\}$

Suponhamos que após a aplicação do algoritmo de arvore de decisão obtivemos a seguinte arvore

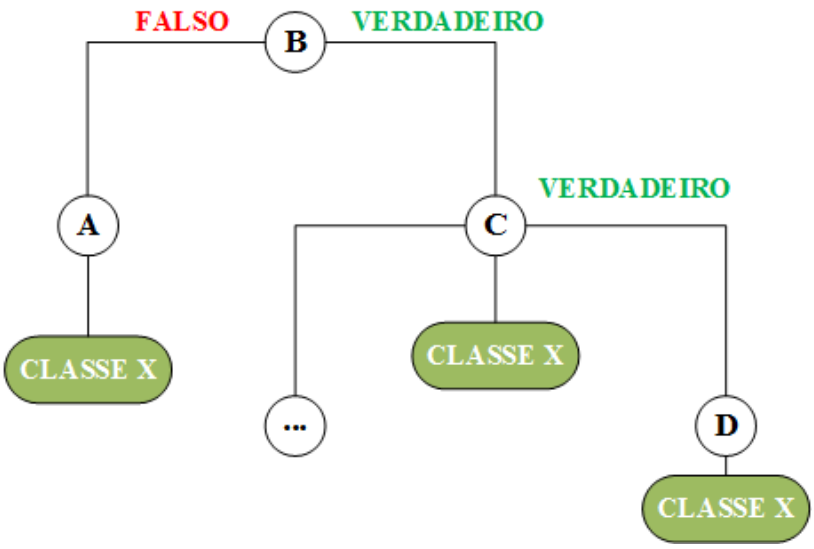


O atributo B foi escolhido como raiz, houve uma divisão, e assim sucessivamente com cada atributo.

Através do classificador formado determinamos em uma tarefa de classificação qual a classe o exemplo pertence.

A primeira dúvida que me surge é em relação a classificação.

O classificador em questão determina no final se final o exemplo pertence a uma determinada classe? Assim, um classificador não tem a liberdade de classificar em um outro, do ponto de vista matemático, o classificador verifica se um exemplo pertence a uma classe?



O que estou querendo dizer com a pergunta acima é se há espaço para o classificador para um **Senão**. Porque tenho visto que das arvores de decisão posso extrair as regras de decisão, mas elas comportam estruturas **Se...Então**. Na minha cabeça (de doido... diga-se de passagem) o algoritmo da Kati não cabe o Senão, mas de forma algorítmica eu poderia resolver isto criando classificadores diferenciados para cada classe.

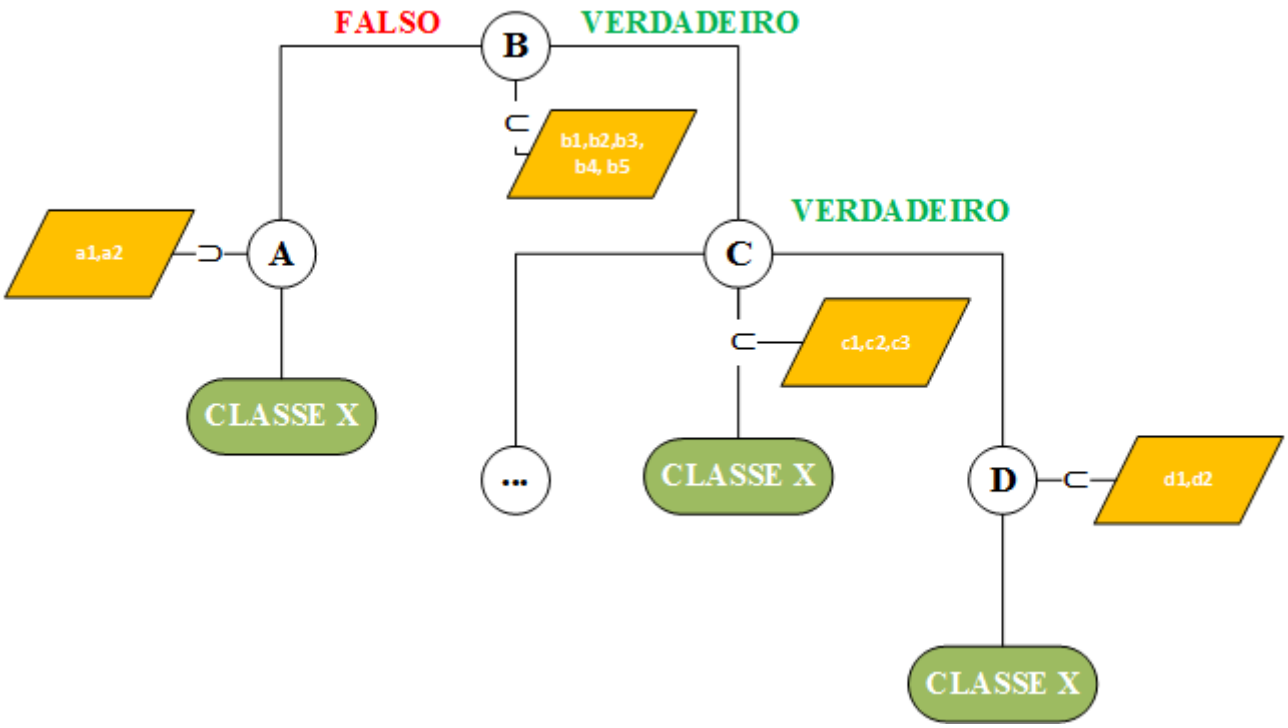
Observando a figura acima surge a segunda questão... A divisão em verdadeiro ou falso não está nada relacionada com um teste logico, mas sim com a probabilidade de dado um atributo a classe poder ser determina? Além de determinar a classe se busca o menor caminho para isto?

No livro da Kati temos uma frase muito massa que diz o seguinte "[...] não há um critério de divisão que seja sistematicamente o melhor de todos", e isso me fica claro porque o que tenho visto não é uma homogeneidade nestas questões e até por isto acabei voltando a elas.

Se pudéssemos representar os dados em um plano poderíamos dizer que cada atributo deste seria um ponto neste plano imaginário, e o classificadores vão ordenando estes pontos de acordo com critérios, no caso do C4.5 é a entropia e o ganho de informação.

Para cada atributo podemos ter dois tipos de dados, não igual a programação, só não encontrei termo melhor. Seria atributos discretos e atributos contínuos.

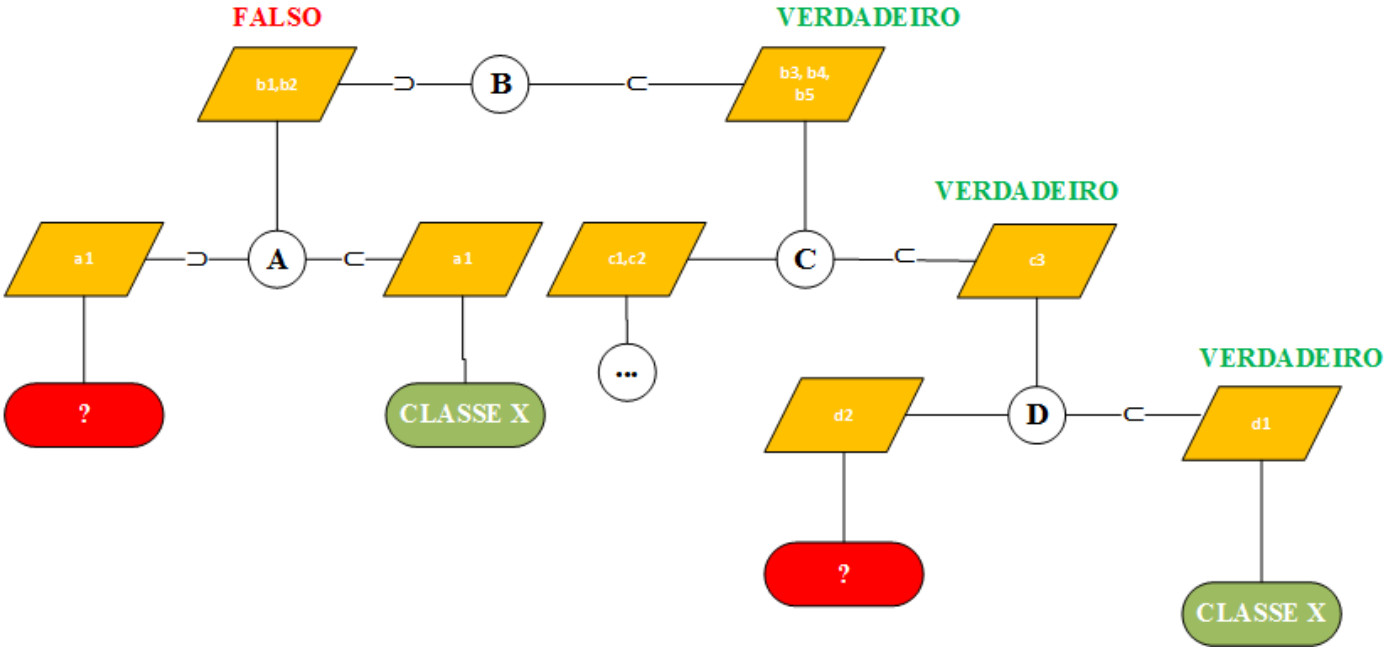
Minha terceira dúvida surge em relação aos atributos contínuos. Os testes de divisão para valores discretos envolvem um cálculo de ganho de informação, no final o teste condicional estabelecido em cada nó consiste em verificar se o valor do exemplo está contido no nó da arvore treinada.



Li e reli o texto de Kati e pelo que entendi a arvore é construída desta forma. O teste com os valores se baseia em verificar se está contido ou não.
Pois bem, tentando construir o algoritmo C4.5 "na unha" percebi que pode ocorrer de que em um determinado atributo para seu conjunto de valores um dos valores pode acabar sendo insignificante para a análise, pois o valor do atributo não determina a classe a alvo. Mesmo o cálculo de entropia e ganho de informação possibilitando escolher outro nó, o que acaba acontecendo e que este atributo é empurrado para um nó a frente ou por alguma outra situação ele acaba sendo elegido como nó. Esse acaso pode acontecer dado a significância dos demais valores do conjunto que pertence ao atributo.

É aqui que surge a questão... Seria errado construir a arvore de decisão considerando a quebra não somente no atributo de maior relevância, mas também em cima dos atributos ou conjuntos de atributos com maior relevância?

Se eu pudesse expressar graficamente seria algo mais ou menos assim,



Minha última dúvida se dá com atributos contínuos,
Seguindo a ideia da Kati a divisão separa os dados em um grupo menor que um valor e outro grupo maior ou igual a um valor.
Este valor pode ser a média do conjunto ou a moda.
Mas me passa como dúvida e se caso houver situações de exatidão no valor, por exemplo, o dado que chega pra ser classificado possui naquele atributo o valor igual a moda.
Da forma que está construída entendo intuitivamente que se isto ocorresse então o exemplo seria avançado mais um nó e avaliado outro atributo.
Então minha dúvida é se na estrutura que representa a arvore não caberia um protótipo de lidar com o caso de igualdade? Acho que isto pode acontecer até com atributos discretos...