



Machine Learning

Árvore de decisão: princípios

Prof. Hugo de Paula

Estrutura da árvore de decisão

- cada nó é um atributo da base de dados.
- nós folha são atributos-classe (rótulo/label),
- cada ramo ligando um nó-filho a um nó-pai é etiquetado com o valor do atributo do nó-pai.
- um atributo que aparece num nó não pode aparecer em seus nós descendentes.

Algoritmos de indução da árvore

- ID3 (final dos anos 1970) - Iterative Dichotomiser
- C45 (sucessor do ID3)
- CART (1984) - Classification and Regression Trees
- J48

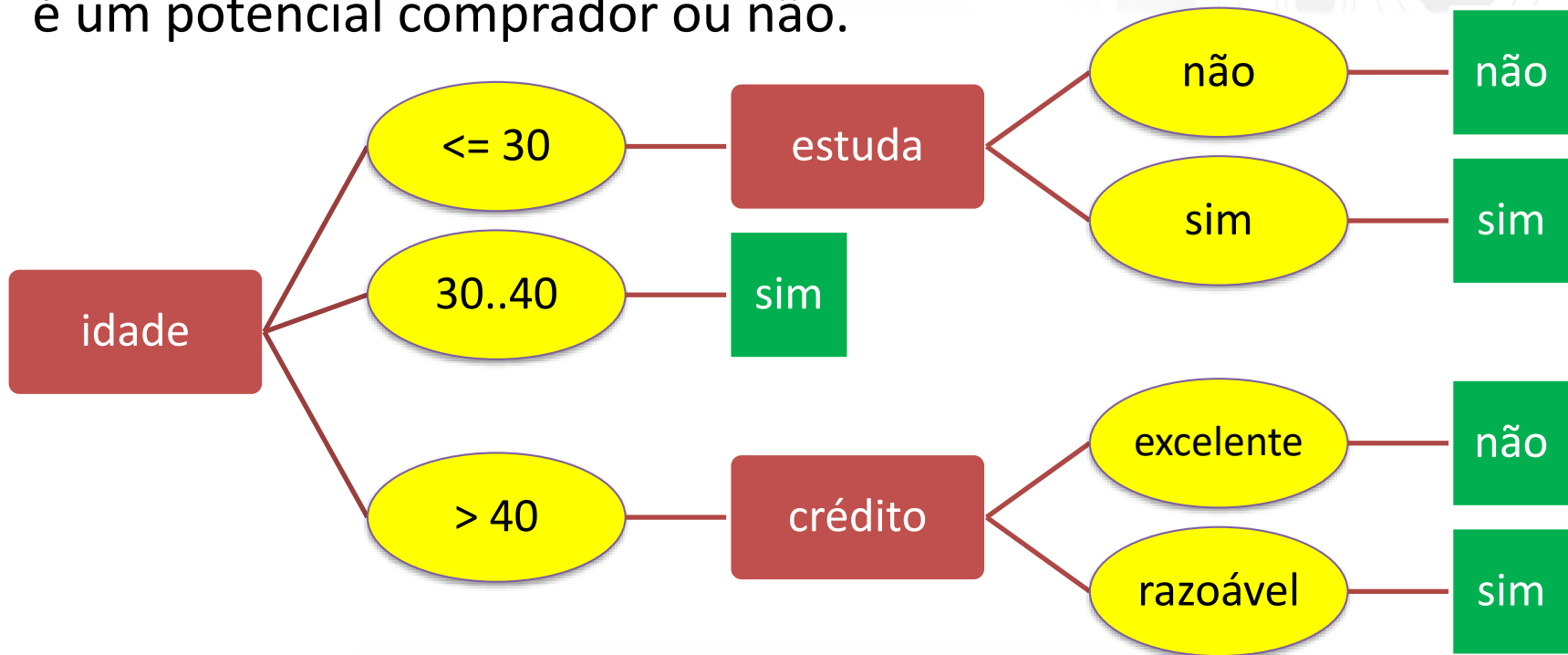
idade	renda	estuda	crédito	compra computador
<=30	alta	não	razoável	não
<=30	alta	não	excelente	não
31...40	alta	não	razoável	sim
>40	média	não	razoável	sim
>40	baixa	sim	razoável	sim
>40	baixa	sim	excelente	não
31...40	baixa	sim	excelente	sim
<=30	média	não	razoável	não
<=30	baixa	sim	razoável	sim
>40	média	sim	razoável	sim
<=30	média	sim	excelente	sim
31...40	média	não	excelente	sim
31...40	alta	sim	razoável	sim
>40	média	não	excelente	não

Exemplo de Quinlan's ID3

– Base de
treinamento

Árvore de decisão: exemplo

Uma possível árvore de decisão criada pelo algoritmo: o usuário é um potencial comprador ou não.



Indução de árvore de decisão

Tipos de dados

- ID3: dados categóricos.
- C4.5: dados contínuos, suporta omissões).

Indução de árvore de decisão

Parâmetros de entrada

- base de dados (B).
- lista de atributos candidatos (CAND).
- um atributo-classe (rótulo): sempre categórico.

Indução de árvore de decisão

Métodos de seleção de atributos

- Ganho de informação (ID3).
- Taxa de ganho (C4.5, J48).
- Índice GINI - impureza (CART).

Aviso legal

O material presente nesta apresentação foi produzido a partir de informações próprias e coletadas de documentos obtidos publicamente a partir da Internet. Este material contém ilustrações adquiridas de bancos de imagens de origem privada ou pública, não possuindo a intenção de violar qualquer direito pertencente à terceiros e sendo voltado para fins acadêmicos ou meramente ilustrativos. Portanto, os textos, fotografias, imagens, logomarcas e sons presentes nesta apresentação se encontram protegidos por direitos autorais ou outros direitos de propriedade intelectual.

Ao usar este material, o usuário deverá respeitar todos os direitos de propriedade intelectual e industrial, os decorrentes da proteção de marcas registradas da mesma, bem como todos os direitos referentes a terceiros que por ventura estejam, ou estiveram, de alguma forma disponíveis nos slides. O simples acesso a este conteúdo não confere ao usuário qualquer direito de uso dos nomes, títulos, palavras, frases, marcas, dentre outras, que nele estejam, ou estiveram, disponíveis.

É vedada sua utilização para finalidades comerciais, publicitárias ou qualquer outra que contrarie a realidade para o qual foi concebido. Sendo que é proibida sua reprodução, distribuição, transmissão, exibição, publicação ou divulgação, total ou parcial, dos textos, figuras, gráficos e demais conteúdos descritos anteriormente, que compõem o presente material, sem prévia e expressa autorização de seu titular, sendo permitida somente a impressão de cópias para uso acadêmico e arquivo pessoal, sem que sejam separadas as partes, permitindo dar o fiel e real entendimento de seu conteúdo e objetivo. Em hipótese alguma o usuário adquirirá quaisquer direitos sobre os mesmos.

O usuário assume toda e qualquer responsabilidade, de caráter civil e/ou criminal, pela utilização indevida das informações, textos, gráficos, marcas, enfim, todo e qualquer direito de propriedade intelectual ou industrial deste material.



PUC Minas
Virtual

© PUC Minas • Todos os direitos reservados, de acordo com o art. 184 do Código Penal e com a lei 9.610 de 19 de fevereiro de 1998.
Proibidas a reprodução, a distribuição, a difusão, a execução pública, a locação e quaisquer outras modalidades de utilização sem a devida autorização da Pontifícia Universidade Católica de Minas Gerais.