

Lista 1

Inteligência Artificial

Iyan Lucas Duarte Marques¹

¹Instituto de Ciências Exatas e Informática - Pontifícia Universidade Católica Minas Gerais (PUC-MG)

1. Questão 01

Usando a base de restaurantes os parâmetros default do algoritmo J48, preencha os valores das seguintes métricas para esta base de dados a partir da árvore gerada pelo Weka

KDD é um processo conceitualizado por Fayyad com o objetivo de identificar padrões [validos] que possam ser utilizáveis na compreensão de dados.

1.1. Processos

- **Análise da base:** Consiste na busca a compreensão da base de dados, qual a finalidade, atributos, campos, etc.
- **Seleção e adição:** Consiste em selecionar um conjunto ou subconjunto de dados, dos quais os mesmos farão parte da análise pelo algoritmo.
- **Pre-processamento:** Consiste em validar a qualidade e a usabilidade do dado. Sujeito à correção ou retirada dos dados inválidos do conjunto de dados usado.
- **Transformação:** Consiste em aplicar técnicas de transformação como: normalização, agregação, criação de novos atributos, redução e sintetização dos dados. Aqui os dados ficam disponíveis agrupados em um mesmo local para a aplicação dos modelos de análise.
- **Mineração de dados:** Consiste em construir modelos ou aplicar técnicas de mineração de dados. Essas técnicas têm por objetivo (1) verificar uma hipótese, (2) descobrir novos padrões de forma autônoma. Além disso, a descoberta pode ser dividida em: preditiva e descritiva. Esses modelos geralmente são aplicados e refeitos inúmeras vezes dependendo do objetivo do projeto.
- **Avaliação e interpretação:** Consiste em avaliar o desempenho do modelo, aplicando em cima de dados que não foram utilizados na fase de treinamento ou mineração. A validação pode ser feita de diversas formas, algumas delas são: utilizar medidas estatísticas, passar pela avaliação dos profissionais de negócio.
- **Visualização e integração:** Consiste na modelagem e *plot* dos resultados obtidos nas etapas anteriores para visualização do cientista.

2. Questão 02

Quais são os principais problemas de aprendizado de máquina existentes? Explique e forneça exemplos.

- **Classificação:** É um dos tipos de problemas mais utilizados que prevê ou descreve uma classe. O atributo de classificação é nominal.

Exemplo:

Chuva

Sim

Não

Sim

- **Regressão:** Tal qual o problema de classificação porém seus atributos de classe são numéricos.

Exemplo:

Medalhas

5

7

1

- **Agrupamento (clusterização):** O problema busca agrupar as instâncias de acordo com os atributos de entrada. O problema não leva em consideração o atributo de classificação.

Exemplo: Agrupar pessoas com interesses em comum ao planejar uma festa.

- **Regras de associação:** O problema busca semelhanças e infere associações entre os elementos.

Exemplo: Se uma pessoa escolheu x, também escolheu y.

3. Questão 03

Quais são os principais métodos de aprendizado existentes? Explique e forneça exemplos.

3.1. Supervisionado

O indutor recebe conjunto de exemplos com uma entrada e um rótulo. As técnicas utilizadas são:

- Redes Neurais do tipo Multilayer Perceptron
- Máquinas de Vetores Suporte
- Árvores de Decisão

3.2. Não Supervisionado

O indutor recebe conjunto de exemplos somente com uma entrada e tenta encontrar agrupamentos.

- Redes Neurais do tipo mapas auto-organizáveis
- Algoritmo k-médias

3.3. Semi Supervisionado

O indutor recebe conjunto de exemplos com uma entrada. Os atributos podem ser ou não ser rotulados, cabe ao indutor decidir o melhor modelo.

- SVM

3.4. Reforço

O indutor concentra-se na maximização das recompensas do resultado.

- Redes Neurais

3.5. Deep Learning

Rede neural com muitas camadas. Para cada reconhecendo situações com ordem de complexidade maiores

- Redes Neurais

4. Questão 04

Considerando-se a base de dados sobre “Esperar ou não pelo restaurante” (verificar base de dados disponibilizada no CANVAS), pede-se:

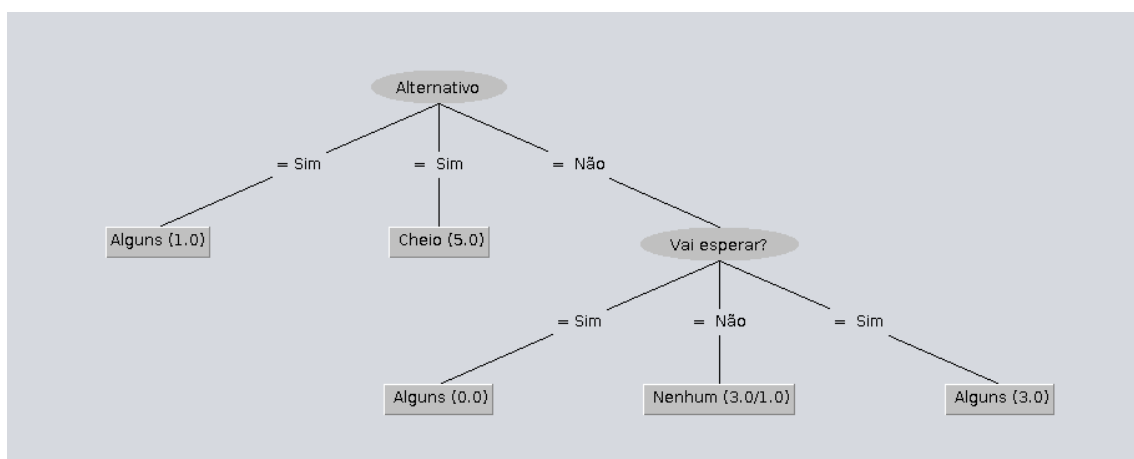
4.1. Calcular o ganho de informação de cada atributo. Que atributo é a raiz da árvore?

O atributo raiz da árvore é a coluna CLIENTE.

4.2. Que atributo estará no segundo nível da árvore. Faça os cálculos e apresente a árvore gerada.

O 2º atributo será a coluna SEX/SAB

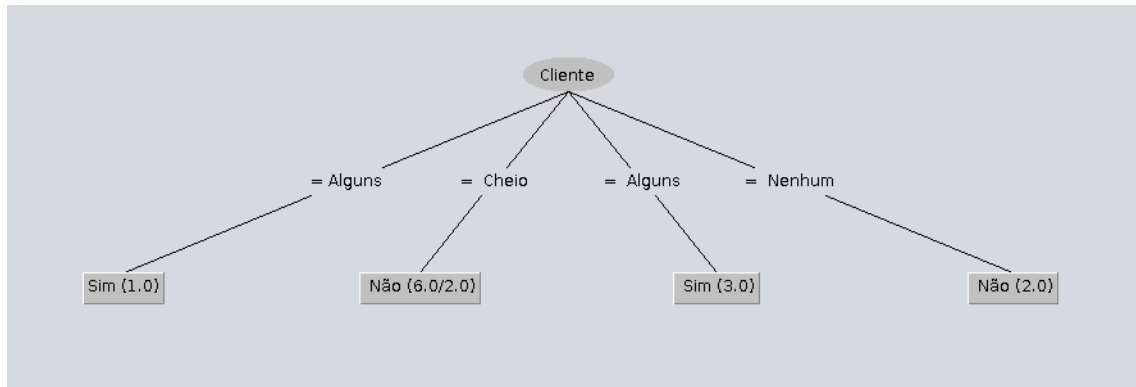
4.3. Quais as regras obtidas a partir desta árvore? Qual a cobertura de cada regra?



4.4. Rodar o algoritmo J48 do WEKA ou C4.5 de algum outro framework de Machine Learning para verificar a árvore gerada.

5. Questão 05

5.1. Utilizando-se a base de dados “Esperar ou não pelo restaurante”, altere os parâmetros indicados na figura abaixo. Mande gerar a árvore novamente e veja o que acontece. Explique. PS: Para alterar algum parâmetro da função, clique sobre o nome da função, ao lado da função “Choose”.



5.2. Investigue o significado dos parâmetros ConfidenceFactor e NimNumObj A opção More explica cada parâmetro.

confidenceFactor – The confidence factor used for pruning (smaller values incur more pruning).

minNumObj – The minimum number of instances per leaf.