



Pontifícia Universidade Católica de Minas Gerais
Instituto de Ciências Exatas e Informática

Disciplina: Inteligência Artificial

Atividade: Lista 2 - IA

Prof.: Cristiane Neri Nobre

Nome: Davi Cândido de Almeida _857859

Estude os slides sobre Árvore de decisão que estão no CANVAS:

- ☒ ~~Parte 1 – Algoritmos de Indução de Árvore – Introdução a AD.pdf~~
- ☒ ~~Parte 2 – Algoritmos de Indução de Árvore – Como gerar – entendimento geral.pdf~~
- ☒ ~~Parte 3 – Algoritmos de Indução de Árvore – Como gerar – como calcular.pdf~~
- ☒ ~~Parte 4 – Algoritmos de Indução de Árvore – Vantagens e desvantagens.pdf~~
- ☒ ~~Parte 5 – Algoritmos de Indução de Árvore – métricas de avaliação.pdf~~

Questão 01

1) Como uma árvore de decisão é gerada? Qual o significado do atributo que está na raiz da árvore?

Uma árvore de decisão é gerada através de um cálculo sobre seus atributos de forma a encontrar os mais relevantes para o resultado/decisão em questão, o atributo dedicado a raiz da árvore é geralmente o atributo mais significativo, ou seja, o que melhor classifica um grande número de dados, ou seja de menor entropia.

2) O que você pode fazer com uma árvore de decisão gerada a partir de uma base de dados?

Uma árvore gerada a partir de uma base de dados pode ser usada para classificar e prever resultados a partir de informações logicamente relacionadas, classificando categorias ou prevendo valores, bem como sua geração permite uma análise e interpretação dos dados mais relevantes sobre uma tomada de decisão específica.

3) Quais as vantagens e desvantagens de um algoritmo de árvore de decisão?

Vantagens: Simplicidade, é interpretável, menor preparação dos dados, podendo ser utilizado tanto para classificação quanto regressão, outras características vantajosas de seu uso é justamente sua flexibilidade e capacidade de selecionar atributos relevantes à tomada de decisão, além de no geral ter uma implementação bastante eficiente (linear ao número de exemplos)

Desvantagens: Overfitting (ou superajuste) podendo ficar muito profundas e perdendo poder de generalização, Instabilidade (pequenas mudanças nos dados exigem a geração de árvores muito diferentes), menor desempenho em problemas complexos, problemas quanto a escolha de atributos mais relevantes (Nem sempre um atributo estatisticamente dito mais relevante conduzirá com a realidade do problema) e a própria limitação da árvore ao lidar com relações mais complexas como interações mais sutis entre variáveis, valores ausentes ou atributos contínuos

4) Como avaliamos a qualidade de uma árvore?

A qualidade de uma árvore é medida pelo seu número de nodos, bem como seu número de folhas, atributos e profundidade, por exemplo uma árvore com menor profundidade tende a ter um menor número de regras para classificar um determinado dado

Agora para avaliar um modelo em específico, ou seja a qualidade levando em conta os resultados da aplicação da árvore em si, se é utilizado métricas para a análise de cada tipo de aprendizado:

- Classificação: Avalia-se pela matriz de confusão
- Agrupamento: Avalia-se através de métricas como silhouette index
- Regressão: Métricas como Mean squared error (MSE), Mean absolute error (MAE), Normalized MSE and MAE
- Associação: Utiliza-se as métricas suporte, confiança, lift, dentre outras

5) Como podemos obter as regras a partir de uma árvore de decisão?

Para se obter uma regra basta observar os possíveis caminhos em uma árvore, cada caminho representa uma regra, onde cada nó se é possível verificar a condição do atributo

Questão 02

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

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

Fórmulas:

$$- \text{Ganho(atributo)} = \text{Entropia(Classe)} - \text{Entropia (Atributo)}$$

A heurística usada na função ESCOLHER – ATRIBUTO é simplesmente escolher o atributo com o maior ganho.

$$- \text{Entropia}(s) = - \sum_{i=1}^c p_i \times \log_2(p_i)$$

Legenda:

Pi: Refere – se à probabilidade de um dado elemento pertencer à classe i.

Log2(pi): Refere – se ao logaritmo de base 2 da probabilidade pi

Sinal negativo (–): Garante que a entropia seja um valor positivo

Cálculos de ganho sobre os atributos base de dados ‘Restaurante.csv’:

Entropia da Classe:

$$\text{Entropia(Classe)} = - \left(\frac{6}{12} \times \log_2\left(\frac{6}{12}\right) \right) - \left(\frac{6}{12} \log_2\left(\frac{6}{12}\right) \right)$$

$$\text{Entropia (Classe)} = -(1/2 * \log_2(1/2)) - (1/2 * \log_2(1/2))$$

$$\text{Entropia (Classe)} = -(1/2 * -1) - (1/2 * -1)$$

$$\text{Entropia (Classe)} = -(-1/2) - (-1/2)$$

$$\text{Entropia (Classe)} = 1/2 + 1/2$$

$$\text{Entropia (Classe)} = 1$$

• Alternativo

$$\text{Entropia (Alternativo)} = 1 - (4/5 * I(2/5,3/5) + 6/12 * I(3/6,3/6))$$

$$\text{Entropia (Alternativo)} = 1 - (6/12 * 1 + 6/12 * 1)$$

$$\text{Entropia (Alternativo)} = 1 - (6/12 + 6/12)$$

$$\text{Entropia (Alternativo)} = 1 - 1$$

$$\text{Entropia (Alternativo)} = 0$$

• Bar

$$\text{Entropia (Bar)} = 1 - (6/12 * I(3/6,3/6) + 6/12 * I(3/6,3/6))$$

$$\text{Entropia (Bar)} = 1 - (6/12 * 1 + 6/12 * 1)$$

$$\text{Entropia (Bar)} = 1 - (6/12 + 6/12)$$

$$\text{Entropia (Bar)} = 1 - 1$$

$$\text{Entropia (Bar)} = 0$$

- **SexSab**

$$\text{Entropia (SexSab)} = 1 - (5/12 * I(2/5,3/5) + 7/12 * I(4/7,3/7))$$

$$I(2/5,3/5) = -(2/5 * \log_2(2/5) + 3/5 * \log_2(3/5))$$

$$I(2/5,3/5) = 0,9709505945$$

$$I(4/7,3/7) = - (4/7 * \log_2(4/7) + 3/7 * \log_2(3/7))$$

$$I(4/7,3/7) = 0,985228136$$

$$\text{Entropia (SexSab)} = 1 - (5/12 * (0,9709505945) + 7/12 * (0,985228136))$$

$$\text{Entropia (SexSab)} = 1 - (0,4045627477 + 0,5747164127)$$

$$\text{Entropia (SexSab)} = 0,0207208396$$

- **Fome**

$$\text{Entropia (Fome)} = 1 - (7/12 * I(5/7,2/7) + 5/12 * I(1/5,4/5))$$

$$I(5/7,2/7) = -(5/7 * \log_2(5/7) + 2/7 * \log_2(2/7))$$

$$I(5/7,2/7) = 0,86312005686$$

$$I(1/5,4/5) = - (1/5 * \log_2(1/5) + 4/5 * \log_2(4/5))$$

$$I(1/5,4/5) = 0,7219280949$$

$$\text{Entropia (Fome)} = 1 - (7/12 * (0,86312005686) + 5/12 * (0,7219280949))$$

$$\text{Entropia (Fome)} = 1 - (0,5034869983 + 0,3008033729)$$

$$\text{Entropia (Fome)} = 0,1957096288$$

- **Cliente**

$$\text{Entropia (Cliente)} = 1 - (2/12 * I(0,1) + 4/12 * I(1,0) + 6/12 * I(2/6,4/6))$$

$$I(0,1) = 0$$

$$I(1,0) = 0$$

$$I(2/6,4/6) = - (2/6 * \log_2(2/6) + 4/6 * \log_2(4/6))$$

$$I(2/6,4/6) = 0,9182958341$$

$$\text{Entropia (Cliente)} = 1 - (2/12 * 0 + 4/12 * 0 + 6/12 * 0,9182958341)$$

$$\text{Entropia (Cliente)} = 1 - (0,459147917)$$

$$\text{Entropia (Cliente)} = 0,540852083$$

- **Preço**

$$\text{Entropia (Preço)} = 1 - (7/12 * I(3/7,4/7) + 2/12 * I(1,0) + 3/12 * I(1/3,2/3))$$

$$I(3/7,4/7) = -(3/7 * \log_2(3/7) + 4/7 * \lg_2(4/7))$$

$$I(3/7,4/7) = 0,985228136$$

$$I(1,0) = 0$$

$$I(1/3,2/3) = - (1/3 * \log_2(1/3) + 2/3 * \log_2(2/3))$$

$$I(1/3,2/3) = 0,9182958341$$

$$\text{Entropia (Preço)} = 1 - (7/12 * (0,985228136) + 2/12 * (0) + 3/12 * (0,9182958341))$$

$$\text{Entropia (Preço)} = \mathbf{0,195709629}$$

- **Chuva**

$$\text{Entropia (Chuva)} = 1 - (5/12 * I(3/5,2/5) + 7/12 * I(3/7,4/7))$$

$$I(3/5,2/5) = -(3/5 * \log_2(3/5) + 2/5 * \lg_2(2/5))$$

$$I(2/5,3/5) = 0,9709505945$$

$$I(3/7,4/7) = -(3/7 * \log_2(3/7) + 4/7 * \lg_2(4/7))$$

$$I(3/7,4/7) = 0,985228136$$

$$\text{Entropia (Chuva)} = 1 - (5/12 * 0,9709505945 + 7/12 * 0,985228136)$$

$$\text{Entropia (Chuva)} = 1 - 0,97927916$$

$$\text{Entropia (Chuva)} = \mathbf{0,02072084}$$

- **Res**

$$\text{Entropia (Res)} = 1 - (5/12 * I(3/5,2/5) + 7/12 * I(3/7,4/7))$$

$$I(3/5,2/5) = -(3/5 * \log_2(3/5) + 2/5 * \lg_2(2/5))$$

$$I(2/5,3/5) = 0,9709505945$$

$$I(3/7,4/7) = -(3/7 * \log_2(3/7) + 4/7 * \lg_2(4/7))$$

$$I(3/7,4/7) = 0,985228136$$

$$\text{Entropia (Res)} = 1 - (5/12 * 0,9709505945 + 7/12 * 0,985228136)$$

$$\text{Entropia (Res)} = 1 - 0,97927916$$

$$\text{Entropia (Res)} = \mathbf{0,02072084}$$

- **Tipo**

$$\text{Entropia (Tipo)} = 1 - (2/12 * I(1/2,1/2) + 2/12 * I(1/2,1/2) + 4/12 * I(1/2,1/2) + 4/12 * I(1/2,1/2))$$

$$\text{Entropia (Tipo)} = 1 - (2/12 * 1 + 2/12 * 1 + 4/12 * 1 + 4/12 * 1)$$

$$\text{Entropia (Tipo)} = 1 - (1)$$

$$\text{Entropia (Tipo)} = 0$$

- **Tempo**

$$\text{Entropia (Tempo)} = 1 - (6/12 * I(4/6,2/6) + 2/12 * I(1/2,1/2) + 2/12 * I(1/2,1/2) + 2/12 * I(0,1))$$

$$\text{Entropia (Tempo)} = 1 - (6/12 * I(4/6,2/6) + 2/12 * 1 + 2/12 * 1 + 2/12 * 0)$$

$$\text{Entropia (Tempo)} = 1 - (6/12 * I(4/6,2/6) + 2/12 + 2/12)$$

$$I(4/6,2/6) = - (4/6 * \log_2(4/6) + 2/6 * \log_2(2/6))$$

$$I(4/6,2/6) = 0,9182958341$$

$$\text{Entropia (Tempo)} = 1 - (0,79248125)$$

$$\text{Entropia (Tempo)} = 0,20751875$$

Logo de acordo com os cálculos:

- Entropia (Alternativo) = 0
- Entropia (Bar) = 0
- Entropia (SexSab) = 0,0207208396
- Entropia (Fome) = 0,1957096288
- **Entropia (Cliente) = 0,540852083 ← Escolhido**
- Entropia (Preco) = 0,195709629
- Entropia (Chuva) = 0,02072084
- Entropia (Res) = 0,02072084
- Entropia (Tipo) = 0
- Entropia (Tempo) = 0,20751875

Portanto o atributo que será usado na raiz da árvore, e que dará um maior grau de ganho, ou seja melhor contribuirá para a classificação dos atributos nessa etapa é o atributo **Cliente**.

2) Que atributo estará no segundo nível da árvore? Faça os cálculos e apresente a árvore gerada até o segundo nível da árvore.

Para a árvore no segundo nível, ou seja com os registros onde clientes seja nenhum ou alguns retirados da contagem:

Entropia da Classe:

$$Entropia(Classe) = - \left(\frac{2}{6} \times \log_2\left(\frac{2}{6}\right) \right) - \left(\frac{4}{6} \log_2\left(\frac{4}{6}\right) \right)$$

$$Entropia (Classe) = -(1/3 * \log_2(1/3)) - (2/3 * \log_2(2/3))$$

$$\mathbf{Entropia (Classe) = 0,918295834}$$

- **Alternativo**

$$Entropia (Alternativo) = 0,918295834 - (5/6 * I(2/5,3/5) + 1/5 * I(0,1))$$

$$Entropia (Alternativo) = 0,918295834 - (5/6 * I(2/5,3/5) + 1/5 * 0)$$

$$Entropia (Alternativo) = 0,918295834 - (5/6 * I(2/5,3/5))$$

$$I(2/5,3/5) = -(2/5 * \log_2(2/5) + 3/5 * \log_2(3/5))$$

$$I(2/5,3/5) = 0,9709505945$$

$$Entropia (Alternativo) = 0,918295834 - (5/6 * 0,9709505945)$$

$$Entropia (Alternativo) = 0,918295834 - (0,809125495)$$

$$\mathbf{Entropia (Alternativo) = 0,109170339}$$

- **Bar**

$$Entropia (Bar) = 0,918295834 - (3/6 * I(1/3,2/3) + 3/6 * I(1/3,1/3))$$

$$I(1/3,2/3) = -(1/3 * \log_2(1/3) + 2/3 * \log_2(2/3))$$

$$I(1/3,2/3) = 0,918295834$$

$$Entropia (Bar) = 0,918295834 - (3/6 * 0,918295834 + 3/6 * 0,918295834)$$

$$Entropia (Bar) = 0,918295834 - 0,918295834$$

$$\mathbf{Entropia (Bar) = 0}$$

- **SexSab**

$$\text{Entropia (SexSab)} = 0,918295834 - (5/6 * I(2/5,3/5) + 1/5 * I(0,1))$$

$$\text{Entropia (SexSab)} = 0,918295834 - (5/6 * I(2/5,3/5) + 1/5 * 0)$$

$$\text{Entropia (SexSab)} = 0,918295834 - (5/6 * I(2/5,3/5))$$

$$I(2/5,3/5) = -(2/5 * \log_2(2/5) + 3/5 * \log_2(3/5))$$

$$I(2/5,3/5) = 0,9709505945$$

$$\text{Entropia (SexSab)} = 0,918295834 - (5/6 * 0,9709505945)$$

$$\text{Entropia (SexSab)} = 0,918295834 - (0,809125495)$$

$$\text{Entropia (SexSab)} = \mathbf{0,109170339}$$

- **Fome**

$$\text{Entropia (Fome)} = 0,918295834 - (4/6 * I(2/4,2/4) + 2/6 * I(0,1))$$

$$\text{Entropia (Fome)} = 0,918295834 - (4/6 * 1 + 2/6 * 0)$$

$$\text{Entropia (Fome)} = 0,918295834 - (0,666666667)$$

$$\text{Entropia (Fome)} = \mathbf{0,251629167}$$

- **Cliente (Já utilizado na raiz)**

- **Preço**

$$\text{Entropia (Preço)} = 0,918295834 - (4/6 * I(2/4,2/4) + 2/6 * I(0,1))$$

$$\text{Entropia (Preço)} = 0,918295834 - (4/6 * 1 + 2/6 * 0)$$

$$\text{Entropia (Preço)} = 0,918295834 - (0,666666667)$$

$$\text{Entropia (Preço)} = \mathbf{0,251629167}$$

- **Chuva**

$$\text{Entropia (Chuva)} = 0,918295834 - (2/6 * I(1/2,1/2) + 4/6 * I(1/4,3/4))$$

$$\text{Entropia (Chuva)} = 0,918295834 - (2/6 * 1 + 4/6 * I(1/4,3/4))$$

$$I(1/4,3/4) = -(1/4 * \log_2(1/4) + 3/4 * \log_2(3/4))$$

$$I(1/4,3/4) = 0,811278124$$

$$\text{Entropia (Chuva)} = 0,918295834 - (2/6 + 4/6 * 0,811278124)$$

$$\text{Entropia (Chuva)} = 0,918295834 - 0,874185416$$

$$\text{Entropia (Chuva)} = \mathbf{0,044110418}$$

- **Res**

$$\text{Entropia (Res)} = 0,918295834 - (2/6 * I(0,1) + 4/6 * I(2/4,2/4))$$

$$\text{Entropia (Res)} = 0,918295834 - (2/6 * 0 + 4/6 * 1))$$

$$\text{Entropia (Res)} = 0,918295834 - (0,666666667)$$

$$\text{Entropia (Res)} = \mathbf{0,251629167}$$

- **Tipo**

$$\text{Entropia (Tipo)} = 0,918295834 - (2/6 * I(1/2,1/2) + 1/6 * I(0,1) + 2/6 * I(1/2,1/2) + 1/6 * I(1/2,1/2))$$

$$\text{Entropia (Tipo)} = 0,918295834 - (2/6 * 1 + 2/6 * 0 + 2/6 * 1 + 1/6 * 1)$$

$$\text{Entropia (Tipo)} = 0,918295834 - (0,833333333)$$

$$\text{Entropia (Tipo)} = \mathbf{0,084962501}$$

- **Tempo**

$$\text{Entropia (Tempo)} = 0,918295834 - (2/6 * I(1/2,1/2) + 2/6 * I(1/2,1/2) + 2/6 * I(0,1))$$

$$\text{Entropia (Tempo)} = 0,918295834 - (2/6 * 1 + 2/6 * 1 + 2/6 * 0)$$

$$\text{Entropia (Tempo)} = 0,918295834 - (0,666666667)$$

$$\text{Entropia (Tempo)} = \mathbf{0,251629167}$$

Logo de acordo com os cálculos:

- Entropia (Alternativo) = 0,109170339
- Entropia (Bar) = 0
- Entropia (SexSab) = 0,109170339
- **Entropia (Fome) = 0,251629167**
- Entropia (Cliente) (Já utilizado na raiz)
- **Entropia (Preco) = 0,251629167**
- Entropia (Chuva) = 0,044110418
- **Entropia (Res) = 0,251629167**
- Entropia (Tipo) = 0,084962501
- **Entropia (Tempo) = 0,251629167**

Portanto o atributo que será usado na raiz da árvore, e que dará um maior grau de ganho, ou seja melhor contribuirá para a classificação dos atributos nessa etapa é os atributos **Fome,Preco, Res, Tempo empataram no calculo de ganho, ou seja, um será escolhido aleatoriamente para o próximo ramo de decisões .**

Estrutura da árvore:

Raiz: Cliente

├─ Nenhum → Classe = Não

├─ Alguns → Classe = Não

└─ Cheio

└─ [próximo nó: Fome OU Preço OU Res OU Tempo]