**Questão 1**

**[Treinamento](https://github.com/joaodest/inteligencia-artificial/blob/master/Lista%202/Lendo_e_tratando_arquivo_v2.ipynb)**

**[Teste do modelo](https://github.com/joaodest/inteligencia-artificial/blob/master/Lista%202/DecisionTree_com_gridsearch.ipynb)**

**Questão 2**

**Parte 1: Diferenças entre os algoritmos ID3 e C4.5**

Os algoritmos ID3 e C4.5 são ambos métodos de construção de árvores de decisão desenvolvidos por Ross Quinlan, mas o C4.5 foi projetado para superar algumas limitações do ID3. Principais diferenças:

**Manipulação de atributos**

* **ID3**: Trabalha apenas com atributos categóricos, ou seja, dados discretos.
* **C4.5**: Pode lidar tanto com atributos discretos quanto com atributos contínuos, realizando automaticamente a discretização dos dados numéricos ao determinar pontos de corte ideais.

**Critério de divisão**

* **ID3**: Utiliza o ganho de informação (information gain) para selecionar os atributos que melhor particionam os dados. Esse critério tende a favorecer atributos com muitos valores distintos, o que pode não ser ideal em todos os casos.
* **C4.5**: Emprega o gain ratio (razão de ganho), que normaliza o ganho de informação levando em conta a dispersão dos dados em cada divisão, evitando a preferência por atributos com muitos valores.

**Tratamento de valores faltantes**

* **ID3**: Não possui um mecanismo robusto para lidar com valores ausentes, o que pode limitar sua aplicação em conjuntos de dados reais.
* **C4.5**: Inclui estratégias para tratar dados faltantes, permitindo uma construção de árvore mais robusta mesmo quando há informações incompletas no conjunto de treinamento.

**Poda da árvore**

* **ID3**: Constrói a árvore até que os nós sejam perfeitamente classificados (ou até atingir um critério de parada simples), sem realizar poda. Isso pode resultar em árvores muito complexas e com alto risco de overfitting (sobreajuste).
* **C4.5**: Implementa técnicas de poda (pruning) após a construção inicial da árvore, removendo ramos que não contribuem significativamente para a capacidade preditiva e, assim, melhorando a generalização do modelo.

**Geração de regras**

* **ID3**: Foca na construção da árvore de decisão em si.
* **C4.5**: Além de construir a árvore, pode converter a estrutura em um conjunto de regras, o que pode facilitar a interpretação dos resultados e a aplicação do modelo em certos contextos.

**Parte 2: Como o C4.5 lida com atributos numéricos**

O algoritmo C4.5 lida com atributos numéricos dividindo-os em faixas, transformando-os em dados categóricos. Ele faz isso ao analisar os valores numéricos dos atributos no conjunto de treinamento, ordená-los e identificar pontos de corte ideais para separar os dados em dois grupos (menor ou igual a um valor específico e maior que esse valor). O ponto de corte escolhido é aquele que maximiza a razão de ganho (gain ratio), garantindo a melhor separação dos dados.

**Questão 3**

**Opção correta**: C) Iris\_Versicolor, íris\_Setosa, Iris\_Versicolor, Iris\_Virgínica

**Questão 4**

* I. VERDADEIRO
* II. FALSO
* III. VERDADEIRO (Nó com 5 amostras, que contém 3 da classe íris-virgínica, e considerando que o total de amostras dessa classe na raiz é 44, temos 3/44 x 100 = 6,8%)

**Opção correta**: d) I e III, apenas.

**Questão 5**

Tabela

O conteúdo gerado por IA pode estar incorreto.