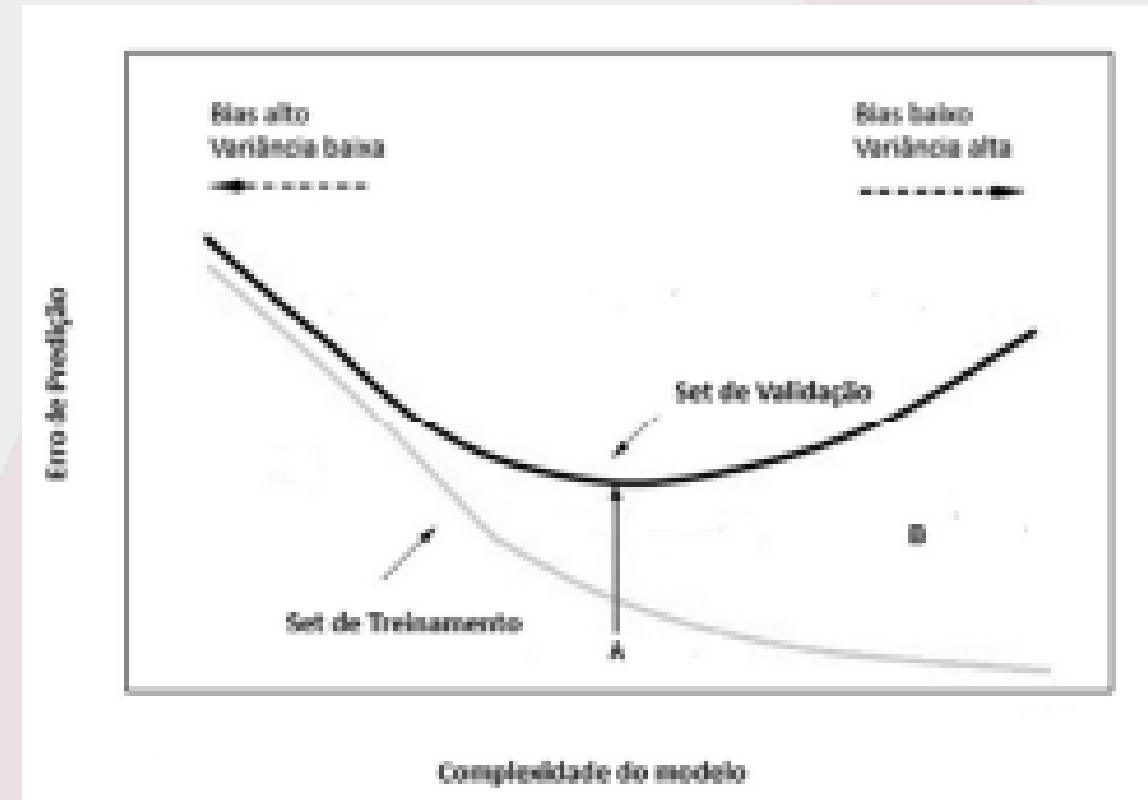


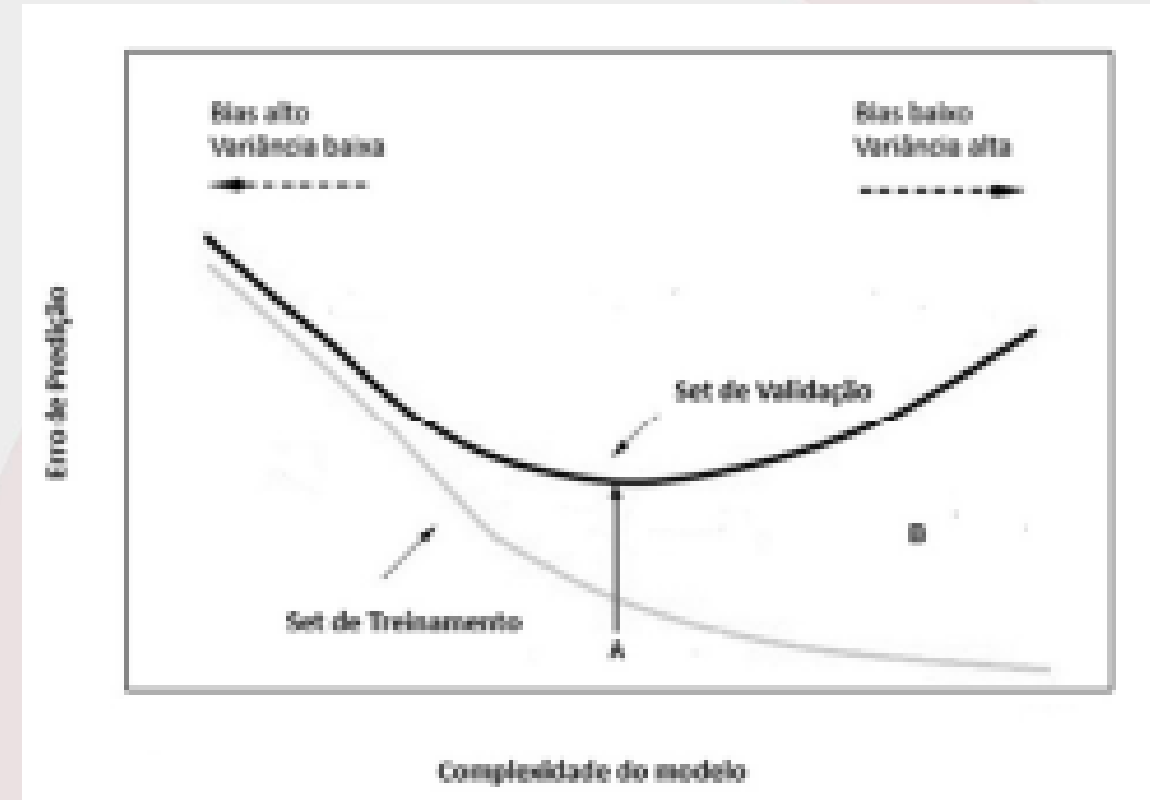
# Aprendizado de Máquina – Questões de Concurso

- Prova: CESPE / CEBRASPE - 2022 - Petrobras - Ciência de Dados
- Julgue o próximo item, considerando o gráfico precedente, que representa as regiões de overfitting e de underfitting, permitindo uma avaliação do relacionamento da complexidade do modelo de aprendizagem de máquina adotado, com o erro de predição.
- A região do gráfico entre as duas curvas, indicada pela letra B, mostra a região de erro de generalização para o modelo de aprendizado de máquina.



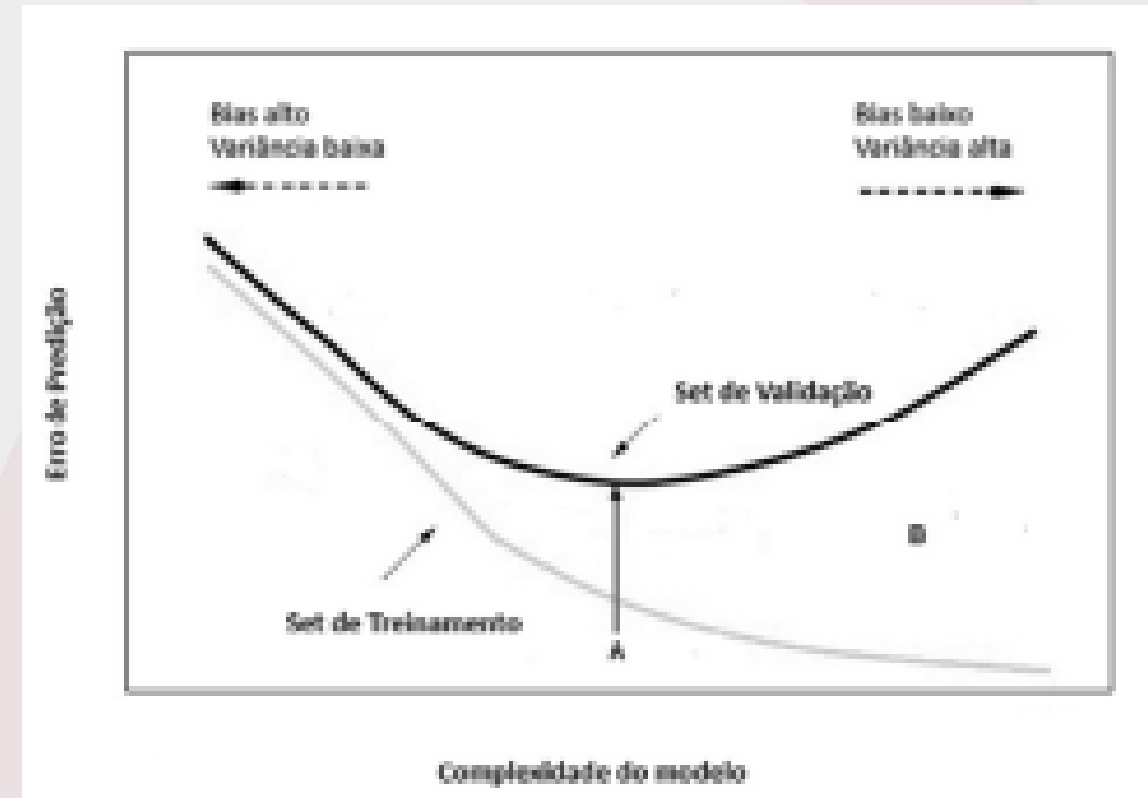
# Aprendizado de Máquina – Questões de Concurso

- Prova: CESPE / CEBRASPE - 2022 - Petrobras - Ciência de Dados
- Julgue o próximo item, considerando o gráfico precedente, que representa as regiões de overfitting e de underfitting, permitindo uma avaliação do relacionamento da complexidade do modelo de aprendizagem de máquina adotado, com o erro de predição.
- A região do gráfico entre as duas curvas, indicada pela letra B, mostra a região de erro de generalização para o modelo de aprendizado de máquina.



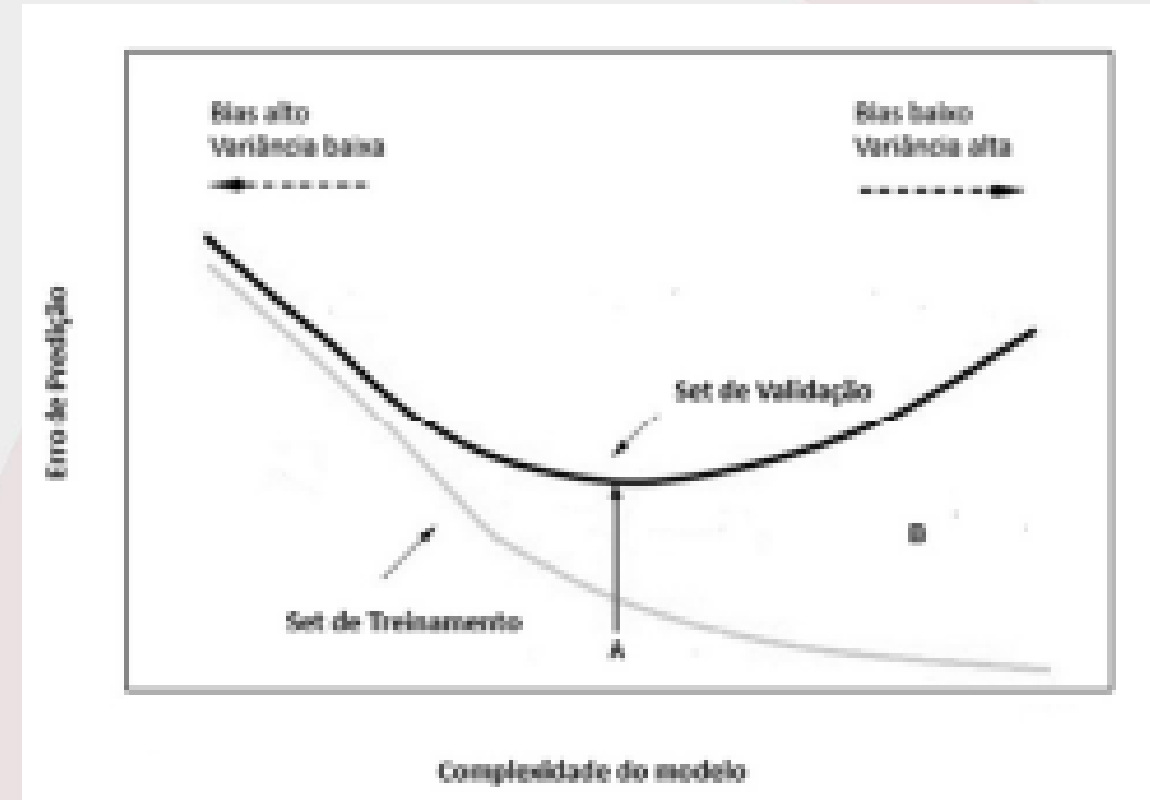
# Aprendizado de Máquina – Questões de Concurso

- Prova: CESPE / CEBRASPE - 2022 - Petrobras - Ciência de Dados
- Julgue o próximo item, considerando o gráfico precedente, que representa as regiões de overfitting e de underfitting, permitindo uma avaliação do relacionamento da complexidade do modelo de aprendizagem de máquina adotado, com o erro de predição.
- O Set de Treinamento é usado para qualificar o desempenho do modelo, enquanto o Set de Validação é utilizado para criar o modelo de aprendizado de máquina.



# Aprendizado de Máquina – Questões de Concurso

- Prova: CESPE / CEBRASPE - 2022 - Petrobras - Ciência de Dados
- Julgue o próximo item, considerando o gráfico precedente, que representa as regiões de overfitting e de underfitting, permitindo uma avaliação do relacionamento da complexidade do modelo de aprendizagem de máquina adotado, com o erro de predição.
- O Set de Treinamento é usado para qualificar o desempenho do modelo, enquanto o Set de Validação é utilizado para criar o modelo de aprendizado de máquina.



# Classificação

- Um problema de classificação tem um valor discreto como saída.

Idade	Gosta de abacaxi na pizza
42	1
65	1
50	1
76	1
96	1
50	1
91	0
58	1
25	1
23	1
75	1
46	0
87	0
96	0
45	0
32	1
63	0
21	1
26	1
93	0
68	1
96	0

# Classificação

- Atribuir um objeto/evento a uma categoria, pertencente a um conjunto finito de categorias)
  - Diagnóstico médico.
  - Detecção de fraude em cartões de crédito.
  - Detecção de vírus em redes de computadores.
  - Filtragem de spam em e-mails.
  - Bioinformática (sequências de DNA).
  - Reconhecimento de caracteres.
  - Reconhecimento de imagens.

# Classificação

- Exemplos de Classificadores:
  - Árvore de Decisão.
  - Random Forest.
  - k-Nearest Neighbors.
  - Naive Bayes.
  - Regressão Logística.
  - Análise Discriminante Linear.
  - SVM.
  - RNA.

# Classificação – Questões de Concurso

- Prova: CESPE - 2018 - Polícia Federal - Agente de Polícia Federal
- Julgue o item que segue, relativo a noções de mineração de dados, big data e aprendizado de máquina.
- Situação hipotética: Na ação de obtenção de informações por meio de aprendizado de máquina, verificou-se que o processo que estava sendo realizado consistia em examinar as características de determinado objeto e atribuir-lhe uma ou mais classes; verificou-se também que os algoritmos utilizados eram embasados em algoritmos de aprendizagem supervisionados. Assertiva: Nessa situação, a ação em realização está relacionada ao processo de classificação.



# Classificação – Questões de Concurso

- Prova: CESPE - 2018 - Polícia Federal - Agente de Polícia Federal
- Julgue o item que segue, relativo a noções de mineração de dados, big data e aprendizado de máquina.
- Situação hipotética: Na ação de obtenção de informações por meio de aprendizado de máquina, verificou-se que o processo que estava sendo realizado consistia em examinar as características de determinado objeto e atribuir-lhe uma ou mais classes; verificou-se também que os algoritmos utilizados eram embasados em algoritmos de aprendizagem supervisionados. Assertiva: Nessa situação, a ação em realização está relacionada ao processo de classificação.

# Classificação – Questões de Concurso

- Prova: CESGRANRIO - 2018 - Transpetro - Analista de Sistemas Júnior - SAP
- Classificação é uma importante tarefa utilizada na etapa de mineração de dados, que tem como uma de suas características básicas
- A construir seus modelos de enquadramento, a partir de um conjunto de dados contínuos.
- B poder ser implementada por algoritmos estáveis e de significativa eficácia, tais como C4.5, classificadores bayesianos ou K-Prototypes.
- C ser um método de aprendizado de máquina não supervisionado, observando o teorema NFL – No Free Lunch.
- D ter a sua eficácia avaliada por uma métrica denominada suporte, que indica quantas vezes um item de dado foi corretamente classificado.
- E ter como seu primeiro processo o aprendizado de uma função de mapeamento  $y = f(X)$ , que associa uma ocorrência de dados  $X$  em uma classe  $y$ .

# Classificação – Questões de Concurso

- Prova: CESGRANRIO - 2018 - Transpetro - Analista de Sistemas Júnior - SAP
- Classificação é uma importante tarefa utilizada na etapa de mineração de dados, que tem como uma de suas características básicas
- A construir seus modelos de enquadramento, a partir de um conjunto de dados contínuos.
- B poder ser implementada por algoritmos estáveis e de significativa eficácia, tais como C4.5, classificadores bayesianos ou K-Prototypes.
- C ser um método de aprendizado de máquina não supervisionado, observando o teorema NFL – No Free Lunch.
- D ter a sua eficácia avaliada por uma métrica denominada suporte, que indica quantas vezes um item de dado foi corretamente classificado.
- E ter como seu primeiro processo o aprendizado de uma função de mapeamento  $y = f(X)$ , que associa uma ocorrência de dados  $X$  em uma classe  $y$ .

# Classificação – Questões de Concurso

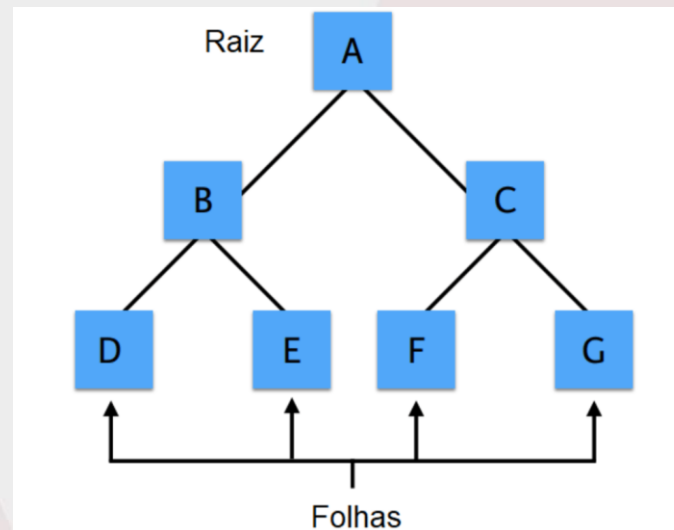
- Prova: INSTITUTO AOCP - 2020 - MJSP - Cientista de Dados - Big Data
- Dentre os métodos de mineração de dados, existem aqueles que são supervisionados e os não supervisionados. Assinale a alternativa que apresenta corretamente um dos métodos supervisionados mais comuns para a aplicação da mineração de dados que é voltado às tarefas frequentes do dia a dia.
- A Regras de associação.
- B Bubble sort.
- C Clusterização.
- D Classificação.
- E Formulação.

# Classificação – Questões de Concurso

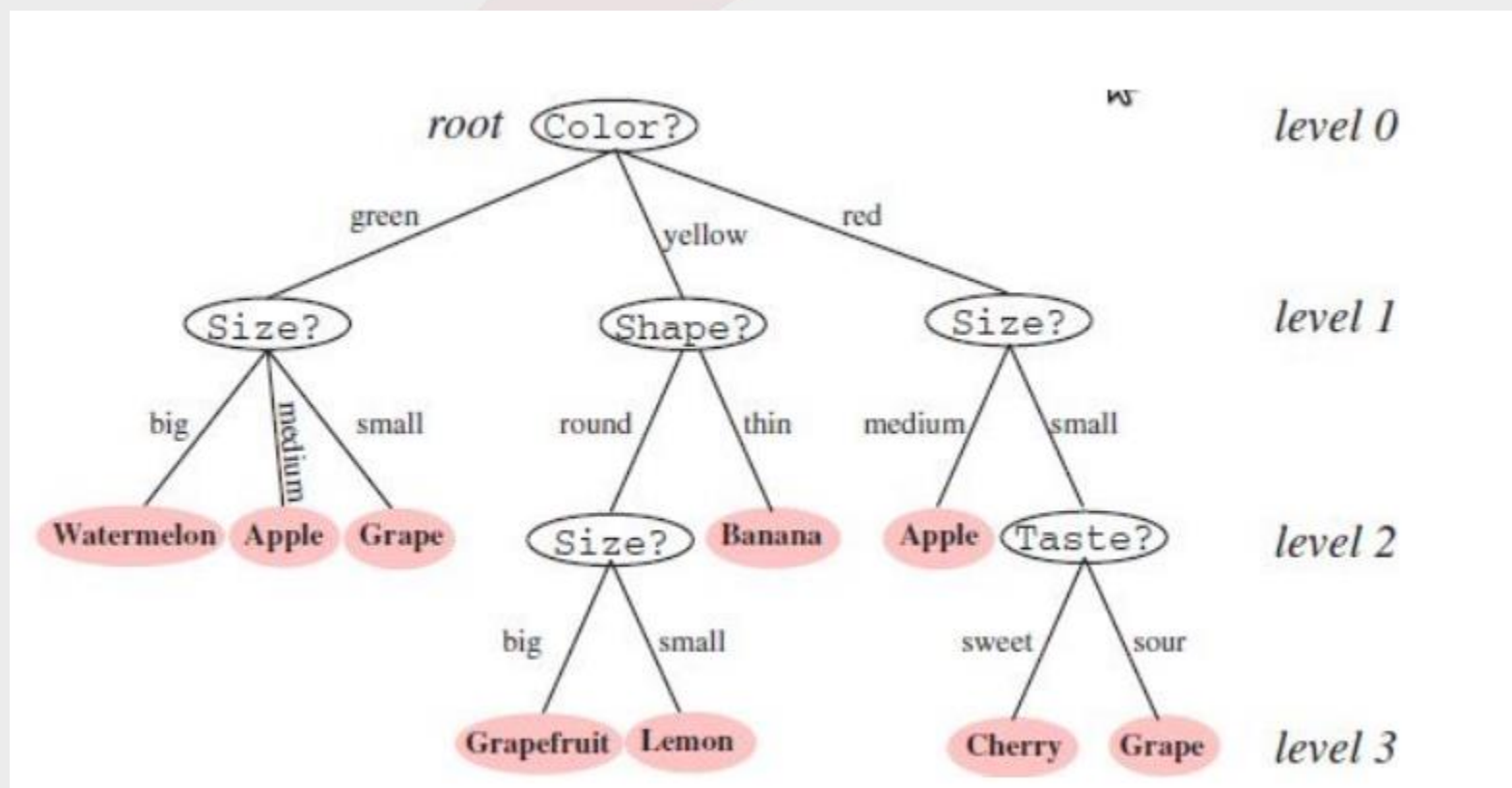
- Prova: INSTITUTO AOCP - 2020 - MJSP - Cientista de Dados - Big Data
- Dentre os métodos de mineração de dados, existem aqueles que são supervisionados e os não supervisionados. Assinale a alternativa que apresenta corretamente um dos métodos supervisionados mais comuns para a aplicação da mineração de dados que é voltado às tarefas frequentes do dia a dia.
- A Regras de associação.
- B Bubble sort.
- C Clusterização.
- D Classificação.
- E Formulação.

# Árvore de Decisão

- A árvore de decisão é uma técnica utilizada para Classificação e consiste em um mapa dos possíveis resultados de uma série de escolhas (*White model*).
- A árvore de decisões é formada por 3 partes principais:
  - Raiz: é a primeira decisão a ser tomada pelo usuário. A partir dela as pessoas são redirecionadas a outras decisões, de acordo com sua escolha nessa questão.
  - Nós: são todas as decisões apresentadas na árvore.
  - Folhas: são os resultados da árvore de decisões. É nela que a classificação é realizada.



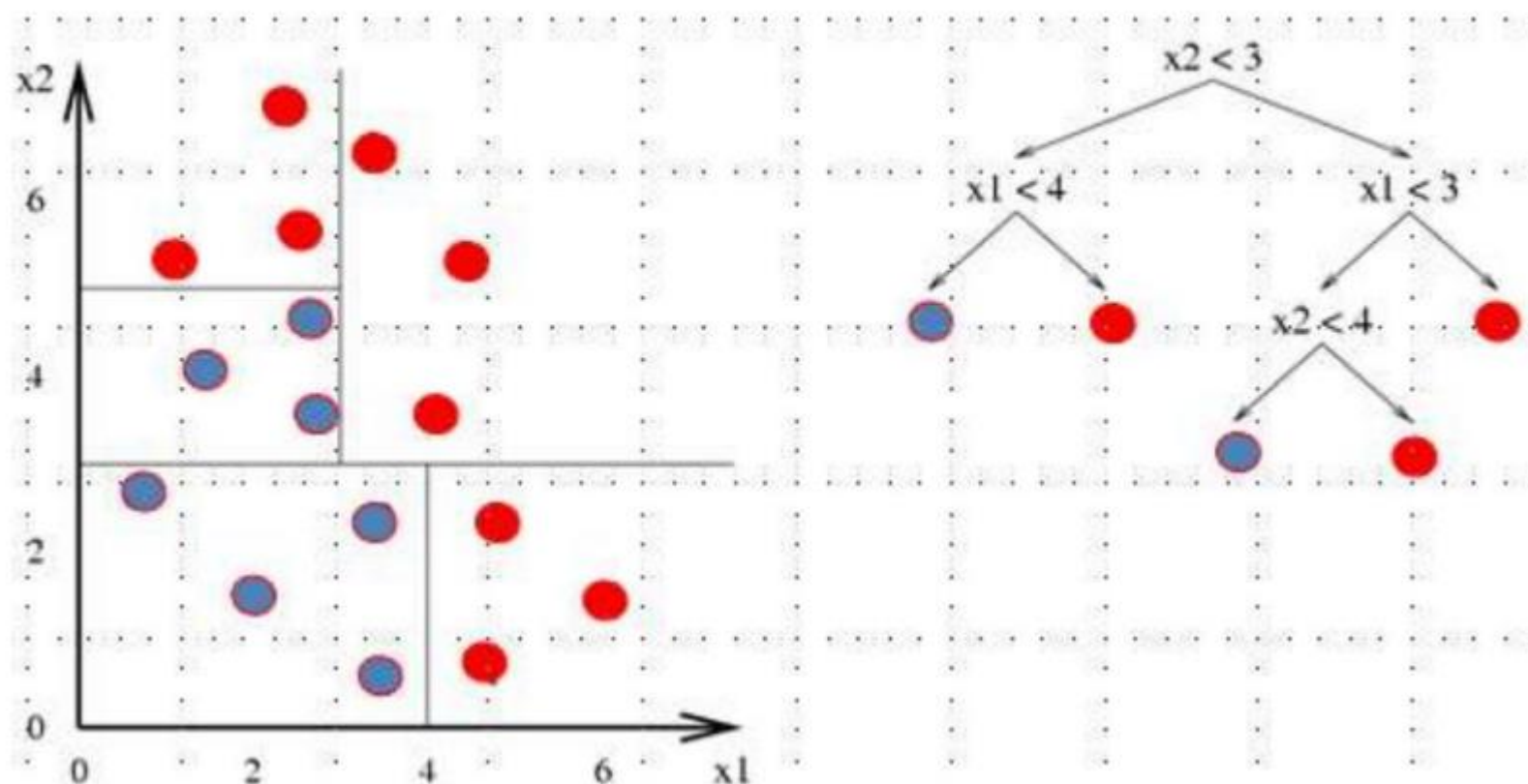
# Árvore de Decisão





# Árvore de Decisão

Árvores de decisão dividem o espaço de características em eixos paralelos retangulares e rotulam cada retângulo com uma classe





# Árvore de Decisão

- Podem ser reescritas como um conjunto de regras, por exemplo, em forma normal disjuntiva (DNF).

–  $\text{red} \wedge \text{circle} \rightarrow \text{pos}$   
 $\text{green} \rightarrow \text{pos}$   
 –  $\text{red} \wedge \text{circle} \rightarrow A$   
 $\text{blue} \rightarrow B$ ;  $\text{red} \wedge \text{square} \rightarrow B$   
 $\text{green} \rightarrow C$ ;  $\text{red} \wedge \text{triangle} \rightarrow C$

# Árvore de Decisão

- Atributos contínuos podem ser usados fazendo o nó dividir o domínio do atributo entre dois intervalos baseados em um limite (ex. tamanho  $< 3$  e tamanho  $\geq 3$ ).
- Árvores de classificação têm valores discretos nas folhas, árvores de regressão têm valores reais nas folhas.
- Algoritmos para encontrar árvores consistentes são eficientes e podem processar grandes quantidades de dados de treinamento.

# Árvore de Decisão

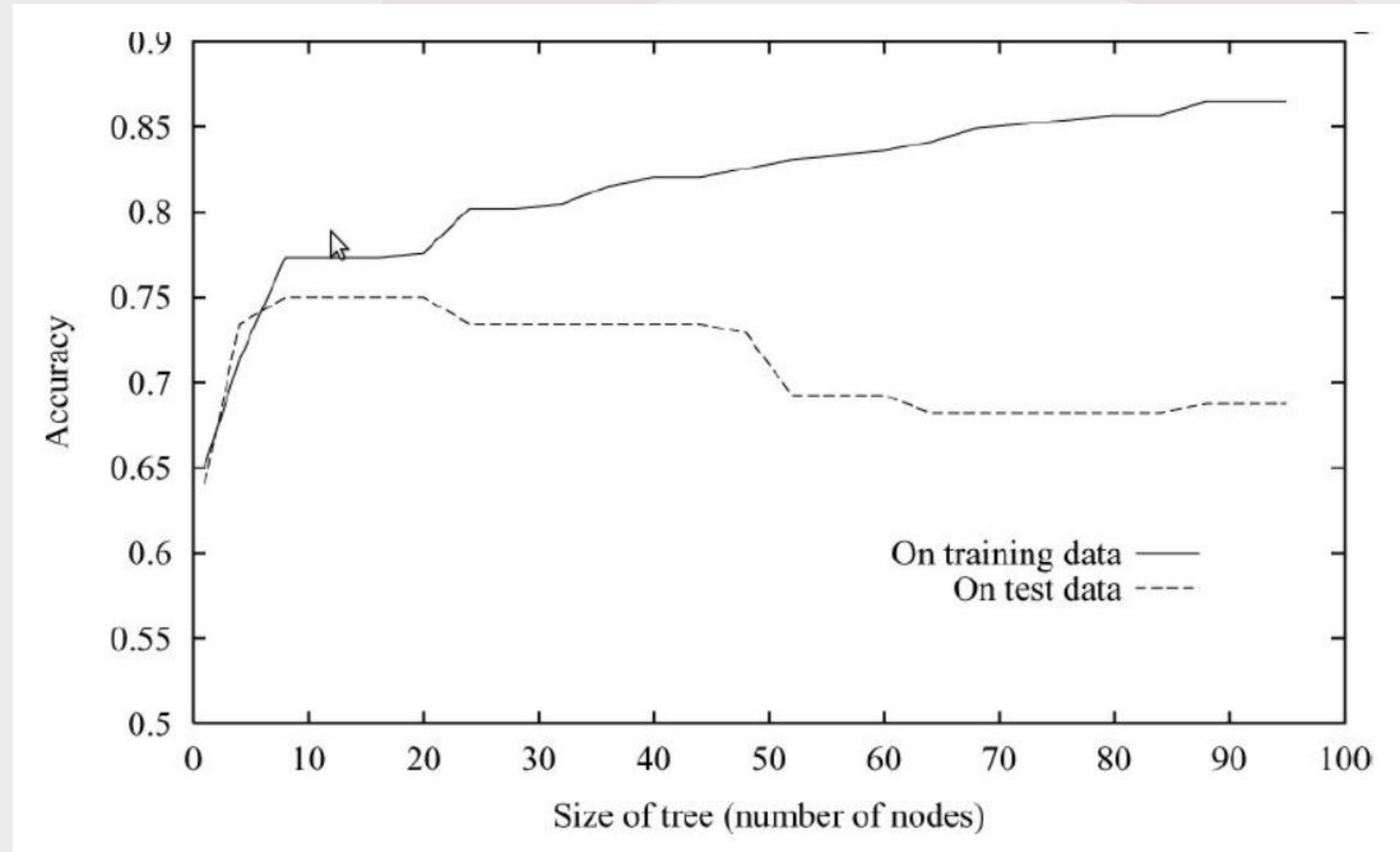
- Não necessita de manipulação de dados, como por exemplo, métodos de normalização.
- Alguns conceitos são de difícil aprendizagem em árvores de decisão, gerando árvores extremamente grandes, por exemplo, XOR.
- A maioria dos algoritmos de aprendizagem de árvores derivam do algoritmo ID3.
  - C4.5 e C5.0 são mais recentes.
  - O ID3 aprende a árvore usando uma estratégia top-down.

# Árvore de Decisão

- Algoritmo básico:
  - O melhor atributo é selecionado e usado como raiz da árvore.
  - Um descendente (sub-árvore) do nó raiz é então criado para cada valor possível deste atributo e os exemplos de treinamento são ordenados para o nó descendente apropriado.
  - O processo é repetido usando exemplos com cada nó descendente para selecionar o melhor atributo para avaliar naquele ponto da árvore.

# Árvore de Decisão

- *Overfitting:*



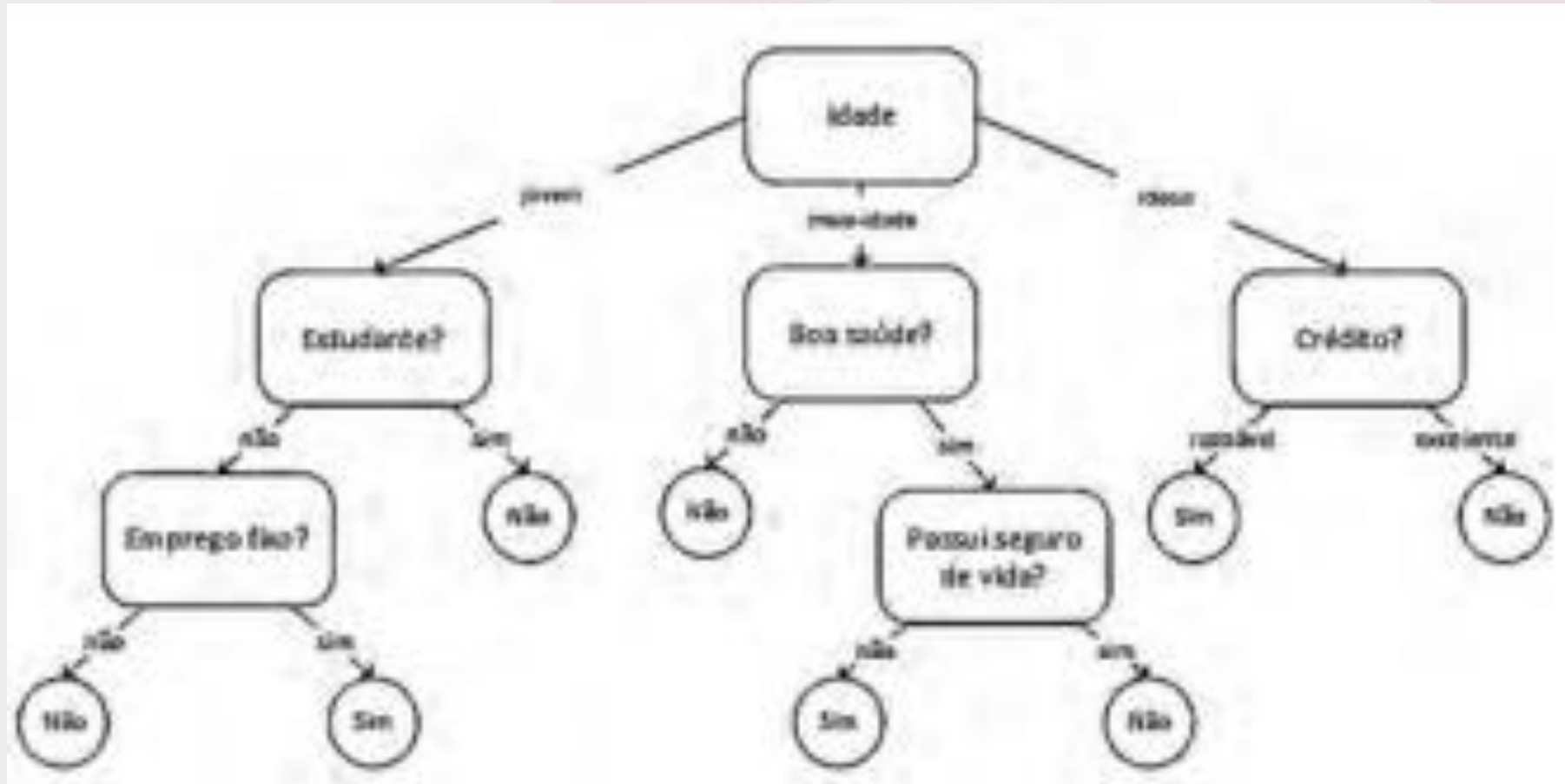
# Árvore de Decisão

- Para evitar *overfitting*:
  - Pré-poda: para de crescer a árvore quando não tem mais dados suficientes para fazer previsões confiáveis.
  - Pós-poda: constrói a árvore toda, depois são removidas sub-árvores com menos relevância.
    - Métodos para poda:
      - Validação cruzada: Reservar alguns dados de treinamento (conjunto de validação) para avaliar utilidade das sub-árvores.
      - Testes estatísticos: Usa o teste para determinar se a regularidade observada foi devida ao acaso.
      - Comprimento da descrição mínima (MDL): Determina se a complexidade adicional da hipótese é mais complexa que lembrar explicitamente as exceções resultantes da poda.

# Árvore de Decisão

- Prova: IADES - 2018 - SES-DF - Analista de Sistemas
- A árvore de decisão apresentada é utilizada para decidir se a compra de um notebook deve ser realizada ou não.
- De acordo com essa árvore de decisão, qual perfil resultaria em uma indicação positiva quanto à compra do computador?
- A Jovem estudante.
- B Jovem estudante sem emprego fixo.
- C Pessoa de meia-idade com saúde ruim.
- D Pessoa idosa com crédito razoável.
- E Pessoa de meia-idade com boa saúde com seguro de vida.

# Árvore de Decisão



(Figura ampliada na página 13)



# Árvore de Decisão

- Prova: IADES - 2018 - SES-DF - Analista de Sistemas
- A árvore de decisão apresentada é utilizada para decidir se a compra de um notebook deve ser realizada ou não.
- De acordo com essa árvore de decisão, qual perfil resultaria em uma indicação positiva quanto à compra do computador?
- A Jovem estudante.
- B Jovem estudante sem emprego fixo.
- C Pessoa de meia-idade com saúde ruim.
- D Pessoa idosa com crédito razoável.
- E Pessoa de meia-idade com boa saúde com seguro de vida.

# Árvore de Decisão

- Prova: CS-UFG - 2018 - UFG - Assistente em Administração
- Uma das ferramentas que auxilia no processo de tomada de decisão é a árvore de decisão. Ela se caracteriza por
- A ser uma técnica que permite a visualização gráfica das possibilidades ou alternativas de decisões.
- B gerar uma tabela ilustrativa das várias possibilidades de decisões e seus respectivos resultados.
- C idealizar uma matriz que compara várias alternativas com atribuições de pesos para diferentes critérios de decisões.
- D analisar prós e contras, vantagens e desvantagens sobre diferentes tomadas de decisões.

# Árvore de Decisão

- Prova: CS-UFG - 2018 - UFG - Assistente em Administração
- Uma das ferramentas que auxilia no processo de tomada de decisão é a árvore de decisão. Ela se caracteriza por
- A ser uma técnica que permite a visualização gráfica das possibilidades ou alternativas de decisões.
- B gerar uma tabela ilustrativa das várias possibilidades de decisões e seus respectivos resultados.
- C idealizar uma matriz que compara várias alternativas com atribuições de pesos para diferentes critérios de decisões.
- D analisar prós e contras, vantagens e desvantagens sobre diferentes tomadas de decisões.

# Árvore de Decisão

- Prova: CESPE / CEBRASPE - 2022 - Petrobras - Ciência de Dados
- Uma árvore de decisão representa um determinado número de caminhos possíveis de decisão e os resultados de cada um deles, apresentando muitos pontos positivos, ou seja, são fáceis de entender e interpretar. Elas têm processo de previsão completamente transparente e lidam facilmente com diversos atributos numéricos, assim como atributos categóricos, podendo até mesmo classificar dados sem atributos definidos.
- De acordo com os aspectos construtivos de uma árvore de decisão, julgue o item a seguir.
- A entropia de uma árvore de decisão aborda o aspecto da quantidade de informações que está associada às respostas que podem ser obtidas às perguntas formuladas, representando o grau de incerteza associado aos dados.

# Árvore de Decisão

- Prova: CESPE / CEBRASPE - 2022 - Petrobras - Ciência de Dados
- Uma árvore de decisão representa um determinado número de caminhos possíveis de decisão e os resultados de cada um deles, apresentando muitos pontos positivos, ou seja, são fáceis de entender e interpretar. Elas têm processo de previsão completamente transparente e lidam facilmente com diversos atributos numéricos, assim como atributos categóricos, podendo até mesmo classificar dados sem atributos definidos.
- De acordo com os aspectos construtivos de uma árvore de decisão, julgue o item a seguir.
- A entropia de uma árvore de decisão aborda o aspecto da quantidade de informações que está associada às respostas que podem ser obtidas às perguntas formuladas, representando o grau de incerteza associado aos dados.

# Árvore de Decisão

- Prova: CESPE / CEBRASPE - 2022 - TCE-RJ - Analista de Controle Externo
- Com referência aos conceitos e técnicas de mineração de dados, julgue o item seguinte.
- Em um modelo para um aprendizado supervisionado dos dados no formato de uma árvore de decisão, um algoritmo de construção da árvore busca minimizar a informação necessária para classificar os dados nas partições da árvore.

# Árvore de Decisão

- Prova: CESPE / CEBRASPE - 2022 - TCE-RJ - Analista de Controle Externo
- Com referência aos conceitos e técnicas de mineração de dados, julgue o item seguinte.
- Em um modelo para um aprendizado supervisionado dos dados no formato de uma árvore de decisão, um algoritmo de construção da árvore busca minimizar a informação necessária para classificar os dados nas partições da árvore.

# Árvore de Decisão

- Prova: CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados
- As árvores de decisão apresentam significativa independência dos dados de treinamento, o que garante forte estabilidade dos caminhos em caso de mudanças nos dados.



# Árvore de Decisão

- Prova: CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados
- As árvores de decisão apresentam significativa independência dos dados de treinamento, o que garante forte estabilidade dos caminhos em caso de mudanças nos dados.