



Fixação de conhecimento

Discente

Igor Lima Rocha

Qual é o objetivo principal do Support Vector Machine (SVM) no contexto da classificação de dados?

Resposta: C) Maximizar a margem entre as duas classes de dados

O SVM busca encontrar um hiperplano que maximize a margem entre as duas classes de dados, proporcionando assim a melhor separação possível entre elas.

Qual das seguintes afirmações é verdadeira em relação às Máquinas de Vetores de Suporte (SVM)?

Resposta: E) SVMs podem ser usados com vários tipos de funções de kernel, como RBF e polinomial.

SVMs são versáteis e podem ser adaptados a diferentes tipos de dados e problemas utilizando várias funções de kernel, o que permite que o algoritmo opere eficazmente em espaços de alta dimensão e com diferentes tipos de dados.

O que é inferência em Redes Bayesianas?

Resposta: B) O processo de atualização de crenças com base em evidências observadas.

A inferência em Redes Bayesianas envolve atualizar a probabilidade de uma hipótese à medida que mais informações (evidências) se tornam disponíveis.

Qual é a finalidade da estrutura de uma Rede Bayesiana?

Resposta: C) Modelar relações entre variáveis aleatórias e suas dependências condicionais.

As Redes Bayesianas são usadas para modelar as dependências condicionais entre um conjunto de variáveis. Isso permite representar relações de causa e efeito de maneira probabilística.

O Random Forest é suscetível a overfitting?

Resposta: C) Sim, mas o Random Forest é menos suscetível a overfitting do que uma única árvore de decisão.

Enquanto o Random Forest, por sua natureza de combinar várias árvores, é menos propenso ao overfitting comparado a uma única árvore de decisão, ele ainda pode sofrer de overfitting, especialmente se as árvores individuais são muito profundas.

Qual é a ideia principal por trás da técnica de "bagging" usada no Random Forest?

Resposta: A) Combinação de modelos fracos em um modelo forte

O "bagging" (Bootstrap Aggregating) envolve treinar múltiplos modelos (neste caso, árvores de decisão) em subconjuntos de dados diferentes e depois combiná-los para criar um modelo mais robusto e preciso.

Como funciona o algoritmo Random Forest ?

Processo de treinamento:

O Random Forest cria várias árvores de decisão, cada uma treinada em uma amostra aleatória do conjunto de dados (com substituição). Cada árvore de decisão faz suas previsões independentemente.

Classificação de dados:

Para classificar um novo exemplo, o algoritmo consulta cada árvore de decisão e a classificação final é determinada pela maioria dos votos das árvores individuais.

Quais são as principais vantagens e desvantagens do Random Forest?

Vantagens:

Alta precisão, boa performance em muitos tipos de problemas, robustez a outliers e capacidade de lidar com dados não lineares.

Desvantagens:

Pode ser computacionalmente intensivo, menos interpretável do que modelos mais simples, e pode levar a overfitting em alguns casos.

Situações apropriadas:

É eficaz em situações com conjuntos de dados grandes e complexos.

Não apropriado:

Em situações onde a interpretabilidade do modelo é crucial ou em datasets muito pequenas.

Quais dos seguintes são parâmetros comuns do algoritmo K-Nearest Neighbors (KNN) ?

Resposta: B) Número de vizinhos (K) e métrica de distância

O parâmetro chave do KNN é o número de vizinhos, e a escolha da métrica de distância (como Euclidiana, Manhattan, etc.) afeta significativamente o desempenho do modelo.

Qual métrica de distância é mais adequada para calcular as distâncias entre pontos com variáveis categóricas no algoritmo K-Nearest Neighbors (KNN)?

Resposta: D) Distância de Hamming

A distância de Hamming é usada para medir a distância entre strings de texto ou sequências de símbolos, o que a torna adequada para variáveis categóricas, pois mede o número de substituições necessárias para mudar uma string em outra.