escola britânica de artes criativas & tecnologia

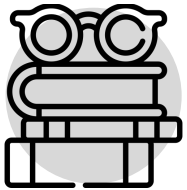
Profissão Cientista de Dados

escola
britânica de
artes criativas
& tecnologia

GLOSSÁRIO

PCA

- **Explore a redução de dimensionalidade**
- **Aprenda como fazer**
- **Analise a matéria-prima**
- **Entenda a Geomatria do PCA**
- **Capture a Informação**
- **Examine a explicação da variância**



Dica: para encontrar rapidamente a palavra que procura aperte o comando CTRL+F e digite o termo que deseja achar.

Explore a redução de dimensionalidade

Explore a redução de dimensionalidade

Análise de Componentes Principais (PCA)

Técnica popular em ciência de dados, especialmente em big data, para a redução de dimensionalidade. Cria novas colunas e trabalha com um número menor delas, resultando em perda de informação, mas a ideia é perder o mínimo possível.

Dimensionalidade

Número de colunas em uma base de dados. Afeta a complexidade do problema e o recurso de máquina necessário, tanto em termos de tempo quanto de memória.

Explore a redução de dimensionalidade

Maldição da Dimensionalidade

Fenômeno que ocorre quando o número de variáveis aumenta, tornando o espaço vetorial mais esparso e a extrapolação mais exagerada. Pode levar a um aumento da chance de overfitting e a uma queda na precisão do modelo após um certo ponto, mesmo com esforços para controlar o overfitting.

Aprenda como fazer

Aprenda como fazer

- **Componentes principais**

  São as novas variáveis criadas pelo PCA, que são combinações lineares das variáveis originais e são ordenadas de acordo com a quantidade de informação que agregam ao conjunto de dados.

- **Pacote sklearn**

  É uma biblioteca de aprendizado de máquina em Python que fornece uma série de ferramentas para modelagem de dados, incluindo o PCA.

- **Grind Search**

  É uma técnica de otimização de hiperparâmetros que busca a melhor combinação de hiperparâmetros para um modelo de aprendizado de máquina.

Analise a matéria-prima

Analise a matéria-prima

- **Redundância de Informação**
  Ocorre quando duas variáveis têm uma alta correlação, indicando que conhecendo o valor de uma, pode-se determinar o valor da outra.

- **Variância**
  É uma medida de dispersão que mostra o quão longe os números estão do valor médio (média).

- **Variabiliade**
  É a medida da dispersão de um conjunto de valores. Na estatística, a variabilidade é uma medida de quão longe um conjunto de números está espalhado.

Entenda a Geometria do PCA

Entenda a Geometria do PCA

- **Auto valores**
  São usados em PCA para determinar a importância de cada componente principal. Eles representam a quantidade de variância que é explicada por cada componente principal.

- **Auto vetores**
  São usados em PCA para determinar a direção de cada componente principal no espaço de dados. Eles são os eixos de maior variabilidade nos dados.

Capture a Informação

Capture a Informação

- **Auto valores**
  São os vetores que, quando multiplicados por uma matriz, resultam no mesmo vetor, apenas escalado.

- **Variância Explicada**
  É a proporção da variância total dos dados que é explicada por cada componente principal na análise PCA.

Examine a explicação da variância

Examine a explicação da variância

- **Autovetores**
  Vetores que apontam na direção de maior variabilidade dos dados.

- **Componentes principais**
  Projeções dos pontos dos dados originais nos autovetores.

- **Autores**
  Representam a variância das componentes principais.

- **Critério de variância explicada**
  Define um percentual mínimo de variabilidade que se deseja explicar para decidir quantas componentes principais utilizar.

Examine a explicação da variância

- **Critério do cotovelo**
  Critério subjetivo para decidir quantas componentes principais utilizar, baseado na identificação de um "cotovelo" no gráfico de variância explicada.

- **Algoritmo de Floresta Aleatória**
  É um método de aprendizado de máquina que opera construindo múltiplas árvores de decisão durante o treinamento e produzindo a classe que é a moda das classes (classificação) ou a média das previsões (regressão) das árvores individuais.

escola britânica de artes criativas & tecnologia

Bons estudos!