

Entrega 3 – Álgebra Linear, Vetores e Geometria Analítica

Nesta entrega, o objetivo é calcular e interpretar os autovalores e autovetores da matriz de dados utilizada no primeiro entregável. Esse tipo de análise é essencial em Inteligência Artificial para entender a estrutura dos dados e aplicar técnicas como PCA (Análise de Componentes Principais).

- **Obtenção dos Dados**

Utilizamos os dados da tabela 'treino_ia' do banco PostgreSQL, contendo colunas como 'Price', 'Fee' e 'TotalUsers'. Os dados foram filtrados para remover registros inválidos e calcular as colunas auxiliares 'price_com_fee' e 'price_por_pessoa'.

2. Construção da Matriz de Dados

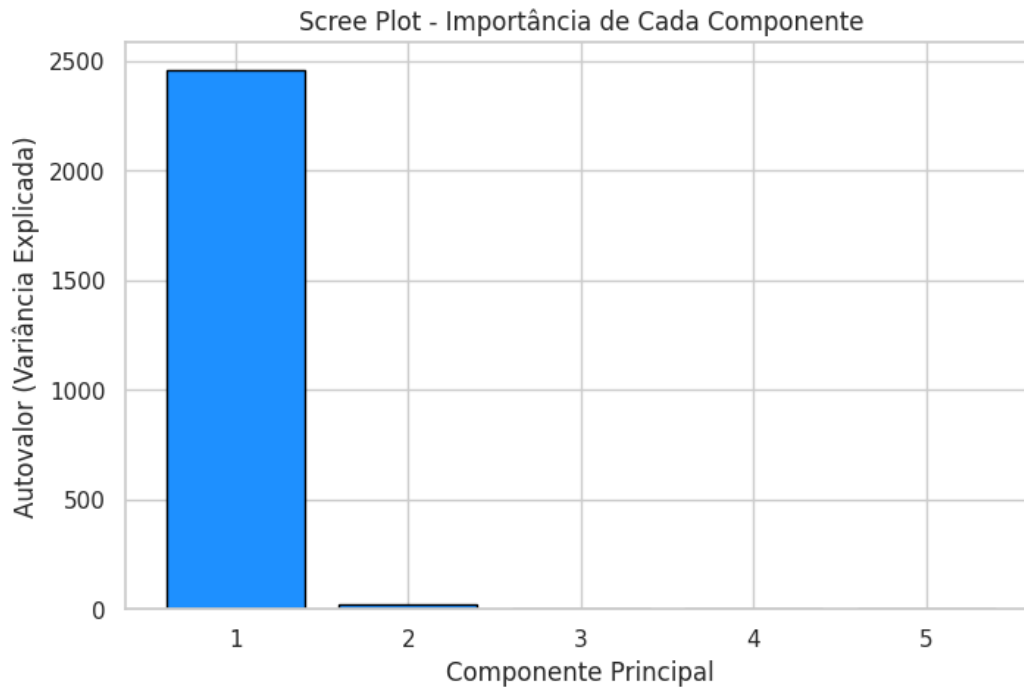
Selecionamos as colunas numéricas ['price', 'fee', 'totalusers', 'price_com_fee', 'price_por_pessoa'] e centralizamos os dados (removendo a média) antes de calcular a matriz de covariância.

3. Cálculo de Autovalores e Autovetores

A partir da matriz de covariância, utilizamos a função `numpy.linalg.eig` para calcular os autovalores e autovetores. Cada autovalor representa a quantidade de variância explicada por uma direção (componente), e o autovetor indica a direção dessa componente no espaço dos dados.

4. Visualização - Scree Plot

O gráfico de barras (Scree Plot) abaixo mostra a importância relativa de cada componente principal, ou seja, quanto da variância dos dados é explicada por cada autovalor. Esse gráfico é usado para identificar quantas dimensões são realmente relevantes na estrutura dos dados.



5. Interpretação

Os componentes principais com maiores autovalores explicam a maior parte da variabilidade nos dados. Isso permite reduzir a dimensionalidade mantendo a essência dos dados, o que é útil para visualização, compressão e modelos de IA mais eficientes.

6. Conclusão

A análise de autovalores e autovetores revelou as direções mais importantes na variação dos dados. Essa técnica é fundamental para preparar dados em projetos de Inteligência Artificial, onde muitas vezes precisamos reduzir dimensões ou entender as relações ocultas entre variáveis.

Link código-fonte:

<https://colab.research.google.com/drive/1uGSdhSnI3oJi2bh2TzpjYgzGNv5VNZWi?usp=sharing>