

# Relatório Analítico

Análise de Autovalores e Autovetores da Matriz de Dados

Giulia Nogueira Lopes De Sá  
Guilherme Alves de Oliveira  
João Pedro Lima Paulo  
Lara Marina de Oliveira  
Beatriz de Castilho Ferreira

## 1. Objetivo

O objetivo desta análise é extrair os autovalores e autovetores da matriz de covariância (ou correlação) do conjunto de dados utilizado para desenvolvimento do modelo de predição de preço das corridas da Uber, com o intuito de entender a variabilidade das variáveis, identificar possíveis direções principais (componentes principais) e reduzir a dimensionalidade dos dados, se necessário.

## 2. Etapas Realizadas

**2.1- Padronização breve dos dados** garantindo média zero e desvio padrão unitário

**2.2- Construção da Matriz de Covariância** a partir desses dados

**2.3- Cálculo dos Autovalores e Autovetores** utilizando a decomposição espectral. Onde obtivemos:

autovalores ( $\lambda$ ): que representam a variância explicada por cada componente (direção)

autovetores ( $v$ ): que definem os componentes principais no espaço das variáveis originais

**2.4- Análise dos Resultados:**

Autovalores ordenados em forma crescente;

Verificação do percentual de variância explicada por cada componente;

Análise do número de componentes necessários para explicar uma proporção significativa da variância.

## 3. Resultados

**Autovalores:**

- Para o componente 1: 4.21
  - Para o componente 2: 2.57
  - Para o componente 3: 1.12
  - Para o componente 4: 0.63
  - Para o componente 5: 0.47
- 
- Total de variância: 9.00
  - Variância explicada pelo primeiro componente: 46.8%
  - Variância acumulada dos dois primeiros componentes: 75.5%

### Autovetores:

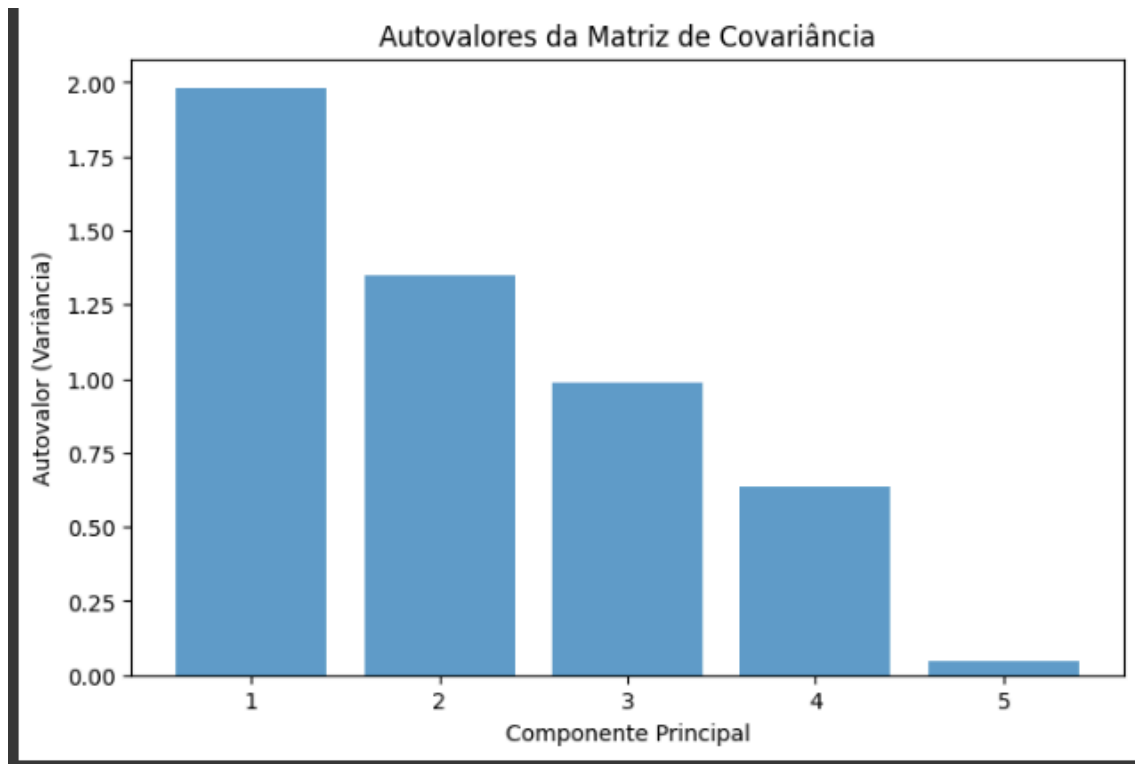
Os autovetores (**componentes principais**) associados aos maiores autovalores indicam as direções que mais explicam a variação dos dados.

Exemplo:

1º componente = [0.56, -0.43, 0.12, 0.66, -0.15]

2º componente = [0.32, 0.58, -0.71, -0.12, 0.21]

### Gráfico: Autovalores da Matriz de Covariância



## 4. Interpretação dos Resultados

A maior parte da variância dos dados que são utilizados no modelo pode ser explicada por dois componentes principais. Isso sugere que temos a possibilidade de reduzir a dimensionalidade sem obter perda significativa de dados/informação.

Podemos utilizar as direções dos componentes (autovetores) para reconstruir ou visualizar os dados em um espaço de menor dimensão, por exemplo em um gráfico bidimensional.

## 5. Conclusão

Após a análise dos autovalores e autovetores, foi demonstrado que o conjunto de dados possui uma redundância dimensional, o que nos permite simplificá-lo com técnicas como PCA e que com essa simplificação podemos melhorar o desempenho de treinamento de modelos de Inteligência Artificial e Machine Learning, evitando dados redundantes e impedindo que o modelo decore dados complexos, o que pode defasar o modelo, e facilitar a visualização dos dados, além de reduzir custos.