

# Análise de Componentes Principais

---

**Dannyel Fontenele Ribeiro – 2322130061**

# Definição



- Algoritmo não-supervisionado que examina interrelações entre variáveis
- Reduz dimensionalidade (qtd. de atributos) dos dados (Extração de atributos), mantém boa parte da informação
- Converte conjunto de variáveis correlacionadas em outro de variáveis não correlacionadas
- Dados em espaço dimensional maior são mapeados para outro menor, e a variância dos dados no menor é máxima
- Componentes principais são mais difíceis de serem interpretados
- Se a relação entre as variáveis é fraca, o PCA não reduz muito bem os dados

# Definição

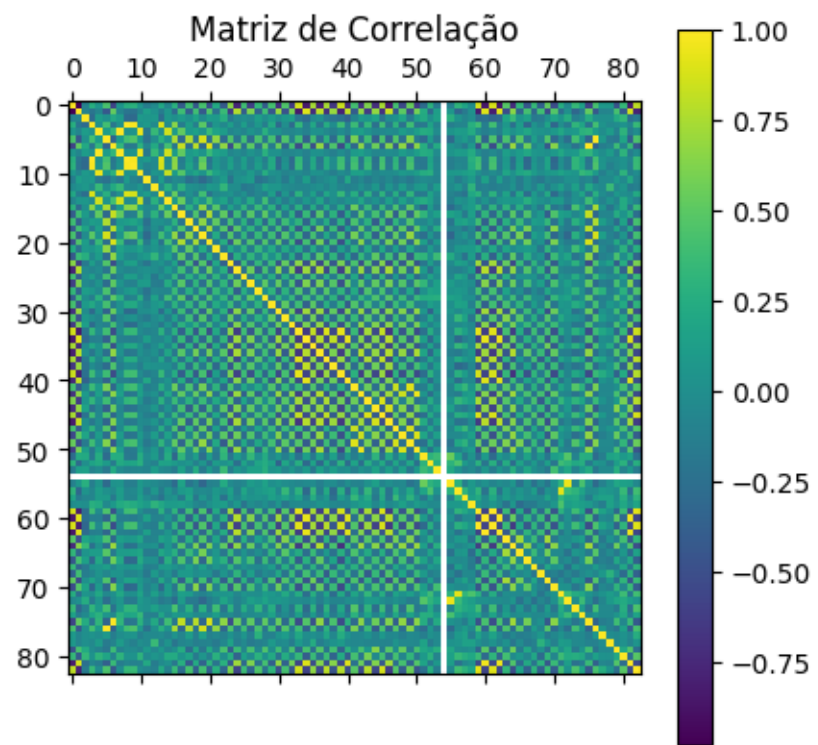
- Componentes principais: variáveis que são combinações lineares de variáveis iniciais
- Qtd. de informação decrescente por componentes principais (qtd. info.  $CP\ 1 > CP\ 2 > \dots$ )
- Descarte de componentes principais com menor informação
- Componentes principais como linhas ortogonais entre si
- Linhas com maior variância = maior dispersão de dados = maior informação



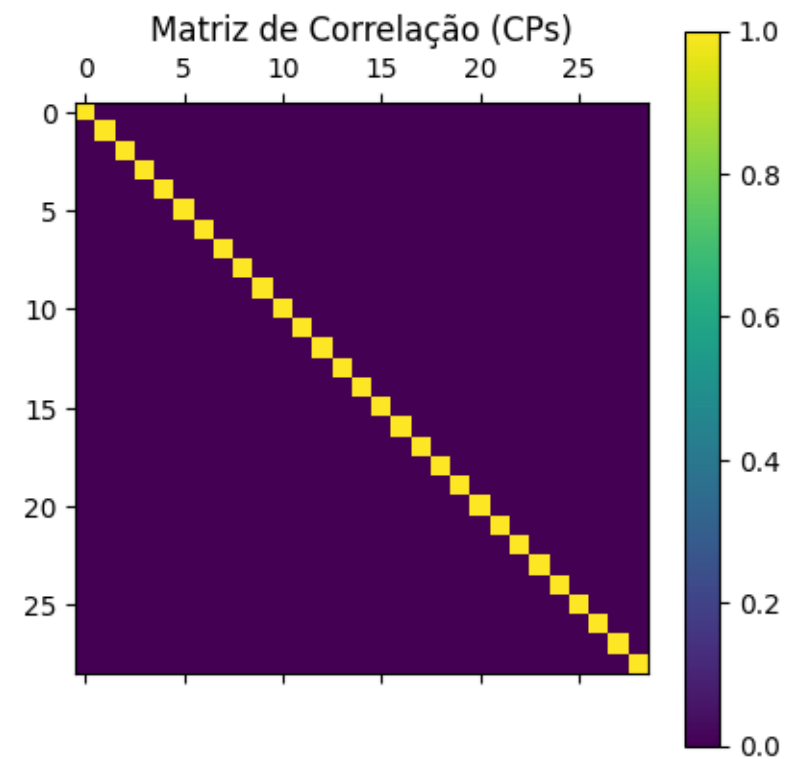
# Funcionamento

1. As variáveis são padronizadas (*z-score*)
2. Calcula-se a matriz de covariância dos dados (Matriz simétrica de ordem  $p$  atributos/dimensões)
3. Calcula-se os autovalores e autovetores da matriz ( $p$  autovetores, autovalores são qtd. de variância por CP)
4. Cria-se um vetor de atributos com os autovetores (ordenados por autovalor)
5. Mapeiam-se os dados ao longo dos novos eixos (Multiplicação de matrizes transpostas)

# Funcionamento



83 atributos



29 componentes principais

## Como é Usado



- Redução de tempo de treino de modelos e de sobreajuste
- Filtração de ruído de dados
- Compressão de Imagens
- Extração de Componentes