

# Principal Component Analysis (PCA)

## Pra que usamos/ para que precisamos do PCA?

- para o modelo treinar mais rápido (ganho de performance)
- para explicar os dados
- diminuir uma grande quantidade de features não significativas, aumentando a relevância das features significativas

## Como funciona a interpretação das novas features do PCA sendo que nós não temos um “nome” para elas? \*\*segunda aula

O PCA tem como objetivo encontrar um hiperplano que minimiza as distâncias entre as instâncias do dataset ou então maximiza as distâncias entre as projeções dos pontos e a origem do hiperplano.

## Exemplo de explicação dos dados

PC1 79%

PC2 15%

Uma representação 2d nos daria 94% da variação dos dados.

DÚVIDA:

Como podemos saber quais features estão no PC1 e quais estão no PC2? Há algum método que nos diz isso?

## Como encontrar o PC1 e o PC2 em casos de bidimensionalidade

A primeira etapa consiste em encontrar o ponto médio dos dados bidimensionais e trazer esse ponto para a origem (0,0) sem mudar a distância relativa dos dados. Depois uma linha que passa pela origem é traçada aleatoriamente e roda até achar sua versão otimizada. Para achar essa linha, ou o modelo que minimiza a distância dos dados em relação a linha ou ele encontra uma linha que maximiza a distância

dos pontos projetados para a origem. O método utilizado se chama Singular Vector Decomposition (SVD) é um método de otimização de um vetor e é utilizado no PCA a fim de determinar os principais components (PC). O número de “principal components”, normalmente, equivale à quantidade de features do dataset.

## “Eigenvalue”

É a soma do quadrado das distâncias entre as instâncias do dataset e os principais components (PC). E o “eigenvalue” nos mostra a proporção que cada PC representa sobre a variação dos dados. Que nos mostra quais PCs usar. No nosso exemplo com duas features o PC1 representa 79% da variação dos dados e o PC2 representa 15% da variação, logo, uma representação 2d desses dois PCs nos daria 94% da variação dos dados.

## Integrantes

Isaac Higuchi RA00306191

Lucas Lopes RA00303799

Carlos Eduardo RA00297767

Thiago Carvalho RA00297792