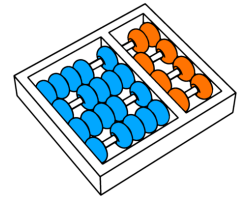


Introdução ao Processamento Digital de Imagem
MC920 / MO443

Prof. Hélio Pedrini

Instituto de Computação

UNICAMP

http://www.ic.unicamp.br/~helio

1º Semestre de 2023

Roteiro

Dimensionalidade dos Dados

- Conforme visto anteriormente, uma representação de dados comum é:

$$\boldsymbol{D} = \left[\begin{array}{c|cccc} & \boldsymbol{X}_1 & \boldsymbol{X}_2 & \ldots & \boldsymbol{X}_d \\ \hline \boldsymbol{x}_1 & x_{11} & x_{12} & \ldots & x_{1d} \\ \boldsymbol{x}_2 & x_{21} & x_{22} & \ldots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_n & x_{n1} & x_{n2} & \ldots & x_{nd} \end{array}\right]$$

em que $\boldsymbol{x}_i$ é a i-ésima linha e uma d-tupla dada por:

$$\boldsymbol{x}_i = \left(x_{i1}, x_{i2}, \ldots, x_{id}\right)$$

e $\boldsymbol{X}_j$ denota a j-ésima coluna e uma n-tupla dada por:

$$\boldsymbol{X}_j = \left(x_{1j}, x_{2j}, \ldots, x_{nj}\right)$$

Dimensionalidade dos Dados

- O número de instâncias ou amostras n é chamado de tamanho do conjunto dos dados.
- O número de atributos ou características d é chamado de dimensionalidade dos dados.

Dimensionalidade dos Dados

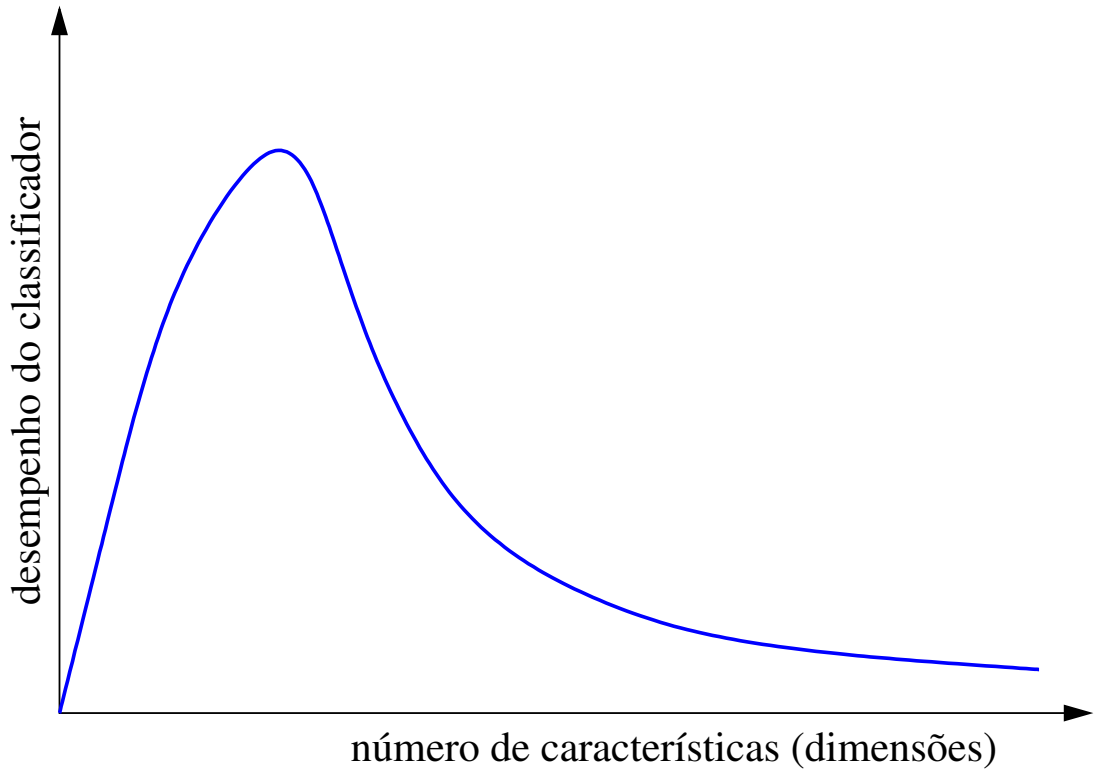
- Conjuntos de dados podem ter tipicamente um grande número de atributos.
- A redução de dimensionalidade visa diminuir o número de atributos de um conjunto de dados.
- Benefícios da redução de dimensionalidade:
  - Redução da complexidade de tempo computacional.
  - Redução da complexidade de espaço de armazenamento.
  - Eliminação de atributos redundantes ou irrelevantes.
  - Geração de modelo mais simples e mais compreensível.
  - Visualização mais intuitiva.

Maldição da Dimensionalidade

- O termo maldição da dimensionalidade foi introduzido pelo matemático americano Richard Bellman.
- A maldição da dimensionalidade refere-se ao fenômeno que surge ao se analisar dados em espaços de alta dimensionalidade (tipicamente, centenas ou milhares de dimensões).
- Muitas abordagens de análise de dados tornam-se significativamente mais complexas com o aumento da dimensionalidade dos dados.

Maldição da Dimensionalidade

- Na prática, a maldição da dimensionalidade implica que o desempenho de um classificador tende a se degradar a partir de um determinado número de características (dimensões).



Maldição da Dimensionalidade

- Quando a dimensionalidade aumenta, os dados se tornam cada vez mais esparsos no espaço que eles ocupam.
- Para um problema de classificação, isto significa que não há objetos de dados suficientes para permitir a criação de um modelo que atribua, de forma confiável, uma classe a todos os objetos possíveis.
- Como consequência, muitos algoritmos de classificação e agrupamento apresentam problemas em termos de eficácia e eficiência.

Maldição da Dimensionalidade

- Exemplo: classificador capaz de distinguir dois tipos de objetos (representados por quadrados e triângulos).
- Apenas 10 amostras estão disponíveis para treinar o classificador neste cenário simples.
- Em uma primeira tentativa com um classificador linear, observa-se que o uso de uma única característica não permite uma separação perfeita das amostras de dados.
- Uma possibilidade para melhorar o resultado seria aumentar o número de características.
- Assume-se um espaço de características em cada dimensão com intervalo de 5 unidades.

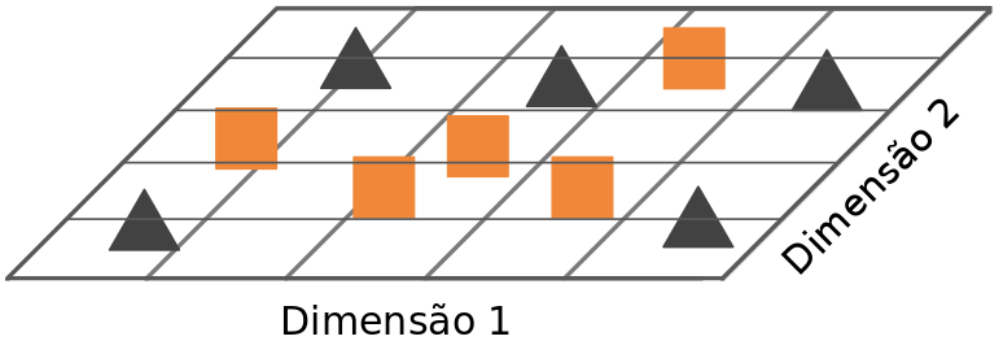
Maldição da Dimensionalidade

À medida que a dimensionalidade aumenta, os dados tornam-se progressivamente esparsos no espaço que ocupam.

10 amostras de dados
1 dimensão: 5 regiões

Maldição da Dimensionalidade

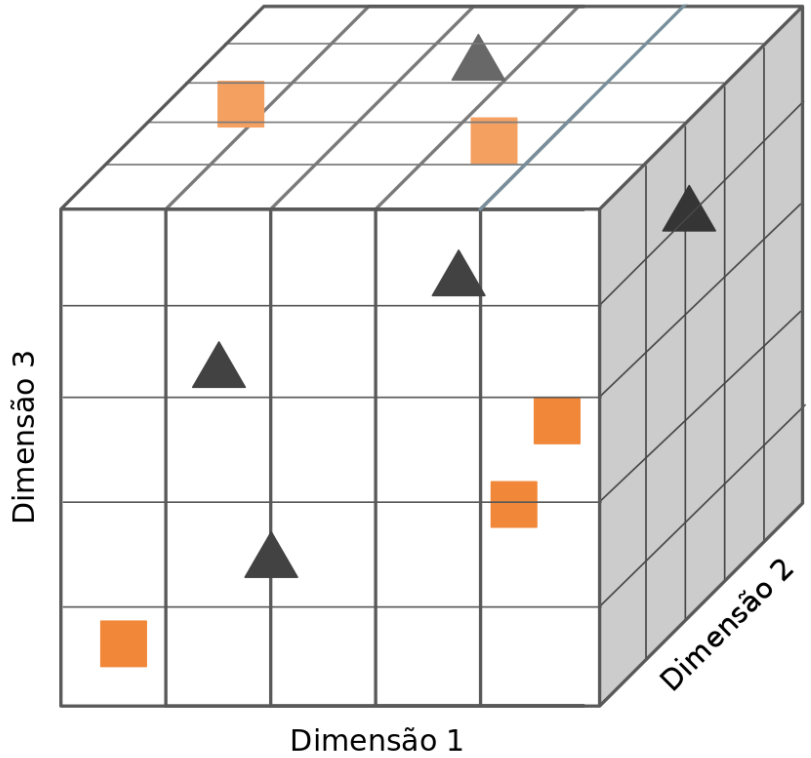
À medida que a dimensionalidade aumenta, os dados tornam-se progressivamente esparsos no espaço que ocupam.



10 amostras de dados
2 dimensões: 25 regiões

Maldição da Dimensionalidade

À medida que a dimensionalidade aumenta, os dados tornam-se progressivamente esparsos no espaço que ocupam.



10 amostras de dados
3 dimensões: 125 regiões

Maldição da Dimensionalidade

- A partir do exemplo, pode-se pensar que o aumento do número de características, até que resulte em uma separação perfeita das amostras, é a melhor estratégia para treinar o classificador.
- Entretanto, a densidade das amostras de treinamento decresceu exponencialmente com o aumento da dimensionalidade do problema.
- 1 dimensão:
  - densidade: $\frac{10 \text{ amostras de dados}}{5 \text{ intervalos}} = 2$ amostras por intervalo.
- 2 dimensões:
  - densidade: $\frac{10 \text{ amostras de dados}}{25 \text{ intervalos}} = 0{,}4$ amostras por intervalo.
- 3 dimensões:
  - densidade: $\frac{10 \text{ amostras de dados}}{125 \text{ intervalos}} = 0{,}08$ amostras por intervalo.

Maldição da Dimensionalidade

- À medida que a dimensionalidade aumenta, a esparsidade dos dados também aumenta (considerando que o conjunto de treinamento seja fixo), tornando-se mais fácil encontrar um hiperplano que separe as amostras.
- Entretanto, o uso demasiado de características pode causar o problema de *overfitting*, já que o classificador passa a aprender detalhes específicos dos dados de treinamento, mas não generaliza quando novas amostras são apresentadas.

Maldição da Dimensionalidade

- O exemplo anterior pode ser considerado de uma maneira diferente:
  - com 1 característica (dimensão): para se obter uma cobertura de 20% do intervalo do espaço com o conjunto de treinamento, a quantidade de amostras deveria ser de 20% $(0,20^1 = 0,2)$ da população disponível.
  - com 2 características (dimensões): para se obter uma cobertura de 20% do intervalo do espaço de características com o conjunto de treinamento, a quantidade de amostras deveria ser de 45% $(0,45^2 \approx 0,2)$ da população em cada dimensão.
  - com 3 características (dimensões): para se obter uma cobertura de 20% do intervalo do espaço de características com o conjunto de treinamento, a quantidade de amostras deveria ser de 58% $(0,58^3 \approx 0,2)$ da população em cada dimensão.

Maldição da Dimensionalidade

Observações:

- Se a quantidade de dados de treinamento for fixa, o problema de *overfitting* pode ocorrer se adicionarmos mais características.
- Por outro lado, se adicionarmos mais dimensões, a quantidade de dados de treinamento cresce exponencialmente para manter a mesma cobertura e evitar *overfitting*.

Redução da Dimensionalidade

Algumas técnicas para reduzir a dimensionalidade dos dados são:

- Seleção de Atributos ou Características:
  - Processo que escolhe um subconjunto ótimo de atributos de acordo com uma função objetivo.

$$[\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_d] \xrightarrow[k \ll d]{} [\boldsymbol{X}_{i_1}, \boldsymbol{X}_{i_2}, \ldots \boldsymbol{X}_{i_k}]$$

- Extração de Atributos ou Características:
  - Ao invés de escolher um subconjunto de atributos, define novas dimensões em função de todos os atributos do conjunto original.

$$[\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_d] \xrightarrow[k \ll d]{} [\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots \boldsymbol{Z}_k] = f([\boldsymbol{X}_{i_1}, \boldsymbol{X}_{i_2}, \ldots \boldsymbol{X}_{i_k}])$$

Redução da Dimensionalidade

Seleção de Características:

- Muitas características são redundantes ou irrelevantes ao problema.
- Tais características podem reduzir o desempenho do algoritmo em questão.
- Muitas características podem ser eliminadas por meio de senso comum ou conhecimento do domínio.
- Entretanto, selecionar o melhor subconjunto de características normalmente requer uma abordagem sistemática.

Redução da Dimensionalidade

Seleção de Características:

- Atributos irrelevantes individualmente podem ser úteis em conjunto.
- Nem sempre os melhores k atributos, segundo algum critério de ordenação, constituem o melhor subconjunto:
  - Atributos devem ser não correlacionados.
  - O melhor subconjunto é o mais complementar.

Redução da Dimensionalidade

Seleção de Características:

- A abordagem ideal é experimentar todos os subconjuntos possíveis de características como entrada para o algoritmo de aprendizado de máquina e então selecionar o subconjunto que produza os melhores resultados.
- Infelizmente, este processo é computacionalmente proibitivo, já que o número de subconjuntos envolvendo d atributos é 2^d (crescimento exponencial).

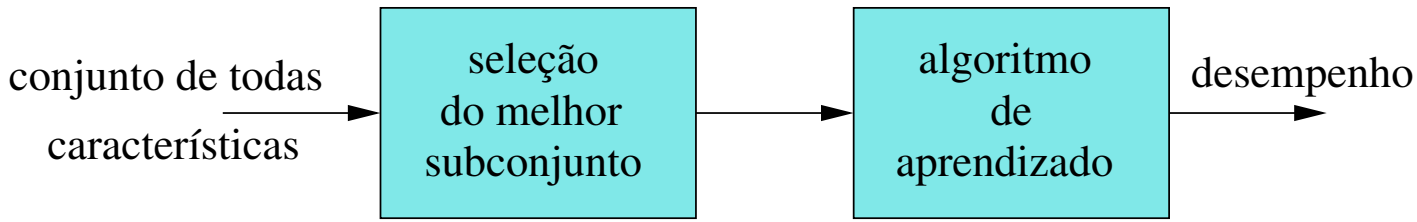
Redução da Dimensionalidade

Duas técnicas automáticas comuns para a seleção de características são:

- Abordagens de Filtros.
- Abordagens de Envoltório.

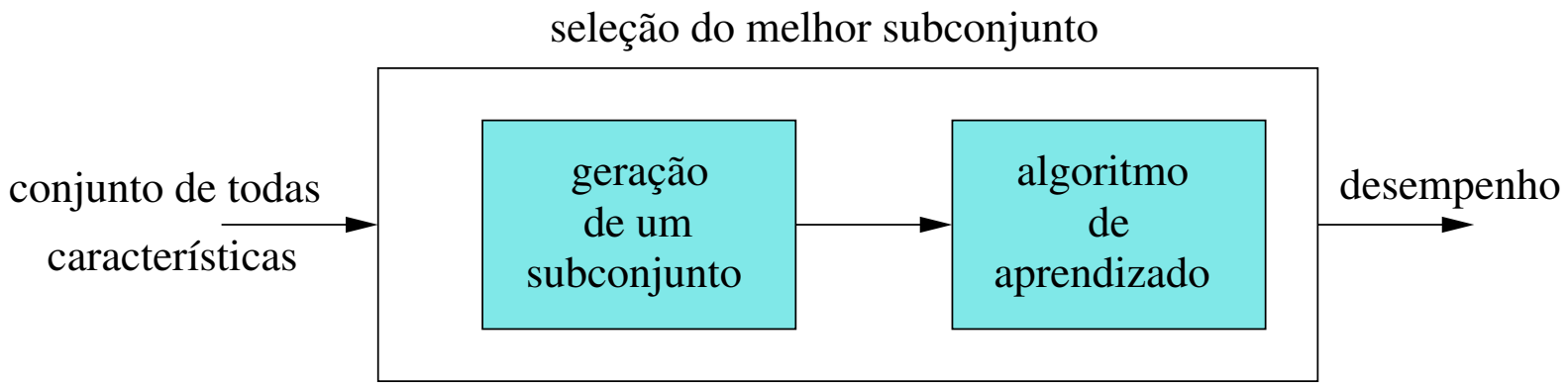
Redução da Dimensionalidade

- Abordagens de Filtros:
  - As características são selecionadas antes que o algoritmo de aprendizado de máquina seja executado, usando alguma abordagem que seja independente da tarefa de agrupamento ou classificação.
  - Por exemplo, pode-se selecionar conjuntos de atributos cuja correlação de pares seja tão baixa quanto possível.
  - A saída da abordagem é o conjunto de atributos por ela selecionados.
  - Entretanto, atributos considerados relevantes por um filtro não necessariamente são úteis para diferentes famílias de algoritmos de aprendizado.



Redução da Dimensionalidade

- Abordagens de Envoltório:
  - Elas geram um subconjunto candidato de atributos, executa o algoritmo de aprendizado com apenas esses atributos no conjunto de treinamento e usa a precisão do classificador extraído para avaliar o subconjunto de atributos em questão.
  - Este processo é repetido para cada subconjunto candidato, até que o critério de parada seja satisfeito.
  - O algoritmo de aprendizado é responsável por conduzir a busca por um subconjunto adequado de atributos.
  - A qualidade de um subconjunto candidato é avaliada utilizando o próprio algoritmo de aprendizado como uma caixa-preta.



Redução da Dimensionalidade

Estratégias:

- Busca para Frente:
  - A busca é iniciada sem atributos e os mesmos são adicionados um a um.
  - Cada atributo é adicionado isoladamente e o conjunto resultante é avaliado segundo um critério.
  - O atributo que produz o melhor critério é incorporado.
- Busca para Trás
  - Inicia-se com todo o conjunto de atributos, eliminando um atributo a cada passo.

Redução da Dimensionalidade

Processo iterativo:

- Pode-se requerer, por exemplo, que a medida de avaliação não apenas cresça a cada passo, mas que ela cresça mais do que uma determinada constante.
- Alguns critérios de parada são:
  - Parar de remover ou adicionar atributos quando nenhuma das alternativas melhorar o desempenho do classificador.
  - Continuar gerando subconjuntos de atributos até que um extremo do espaço de busca seja alcançado e escolher o melhor desses subconjuntos.
  - Ordenar os atributos segundo algum critério e utilizar um parâmetro para determinar o ponto de parada, por exemplo, o número de atributos desejado no subconjunto.

Redução da Dimensionalidade

Outros métodos de busca:

- Busca bidirecional.
- Busca aleatória.
- Busca melhor-primeiro.
- Busca tabu (metaheurística).
- Algoritmos evolutivos.

Redução da Dimensionalidade

Extração de Características:

- Todos os atributos dos dados originais são usados.
- Os atributos são transformados ou combinados em um conjunto reduzido de características melhor representativo, segundo algum critério.
- Esse mapeamento normalmente é uma função dependente do problema.

$$\underbrace{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{bmatrix} \longrightarrow \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_k \end{bmatrix}}_{Z=f(X)} = \underbrace{\begin{bmatrix} W_{11} & W_{12} & \dots & W_{1d} \\ W_{21} & W_{12} & \dots & W_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ W_{k1} & W_{k2} & \dots & W_{kd} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{bmatrix}}_{W \cdot X}$$

Redução da Dimensionalidade

Extração de Características:

- Os dados no espaço original d-dimensional são projetados em um espaço de menor dimensão.
- Espera-se que o conjunto resultante da transformação preserve informação relevante para a tarefa desejada.
- Exemplos de técnicas:
  - Análise de Componentes Principais.
  - Análise de Componentes Independentes.
  - Redução Dimensional Não-Linear.
  - Escala Multidimensional (IsoMap e FastMap).

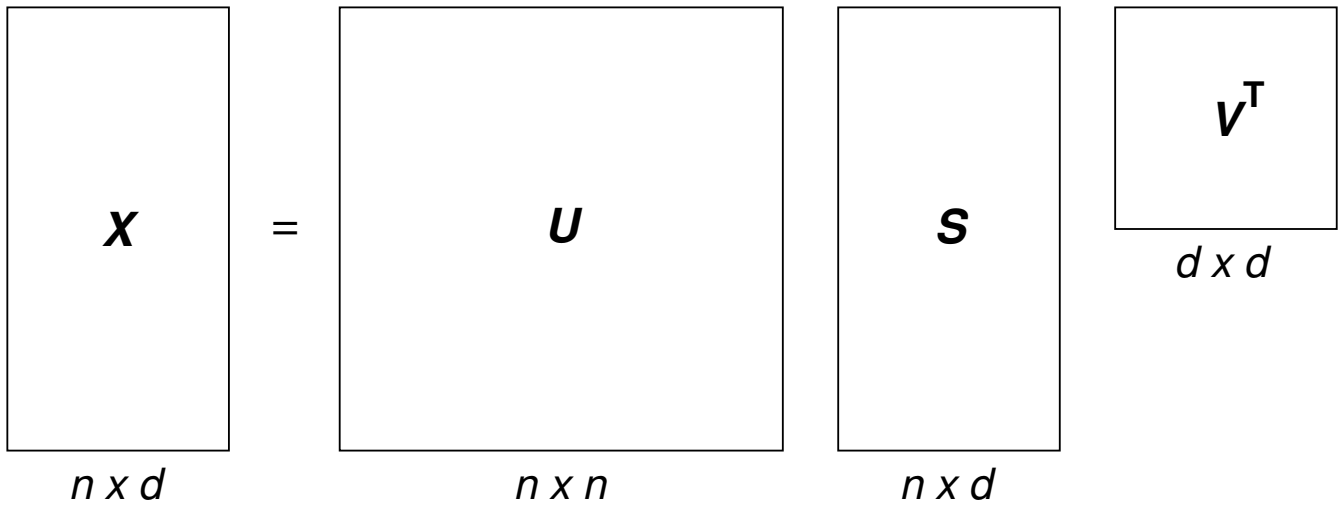
Fatoração de Matrizes

Decomposição em Valores Singulares

- Qualquer matrix $\boldsymbol{X}_{n,d}$ pode ser fatorada utilizando a decomposição em valores singulares:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\boldsymbol{T}}$$

em que $\boldsymbol{U}_{n,n}$ é uma matriz ortogonal, $\boldsymbol{S}_{n,d}$ é uma matriz diagonal de valores singulares (ordenados do maior para o menor) e $\boldsymbol{V}_{d,d}$ é uma matriz ortogonal.



Fatoração de Matrizes

Decomposição em Valores Singulares

- Exemplo:

$$\begin{bmatrix} 3 & 1 & 7 & 5 \\ 2 & 9 & 8 & 4 \\ 7 & 6 & 8 & 6 \\ 2 & 1 & 4 & 6 \\ 9 & 7 & 6 & 3 \end{bmatrix} = \underbrace{\begin{bmatrix} -0.3392 & -0.5749 & 0.0700 & 0.7140 & 0.1994 \\ -0.4909 & 0.2204 & -0.8344 & -0.0072 & 0.1193 \\ -0.5581 & -0.0507 & 0.2019 & -0.1032 & -0.7966 \\ -0.2618 & -0.5870 & 0.0512 & -0.6925 & 0.3235 \\ -0.5138 & 0.5232 & 0.5055 & 0.0006 & 0.4547 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 24.2823 & 0 & 0 & 0 \\ 0 & 6.5302 & 0 & 0 \\ 0 & 0 & 5.4732 & 0 \\ 0 & 0 & 0 & 1.9411 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} -0.4552 & -0.4927 & -0.6134 & -0.4168 \\ 0.2903 & 0.6400 & -0.2873 & -0.6508 \\ 0.8416 & -0.4820 & -0.2433 & 0.0087 \\ 0.0129 & -0.3395 & 0.6942 & -0.6346 \end{bmatrix}}_{V^T}$$

Fatoração de Matrizes

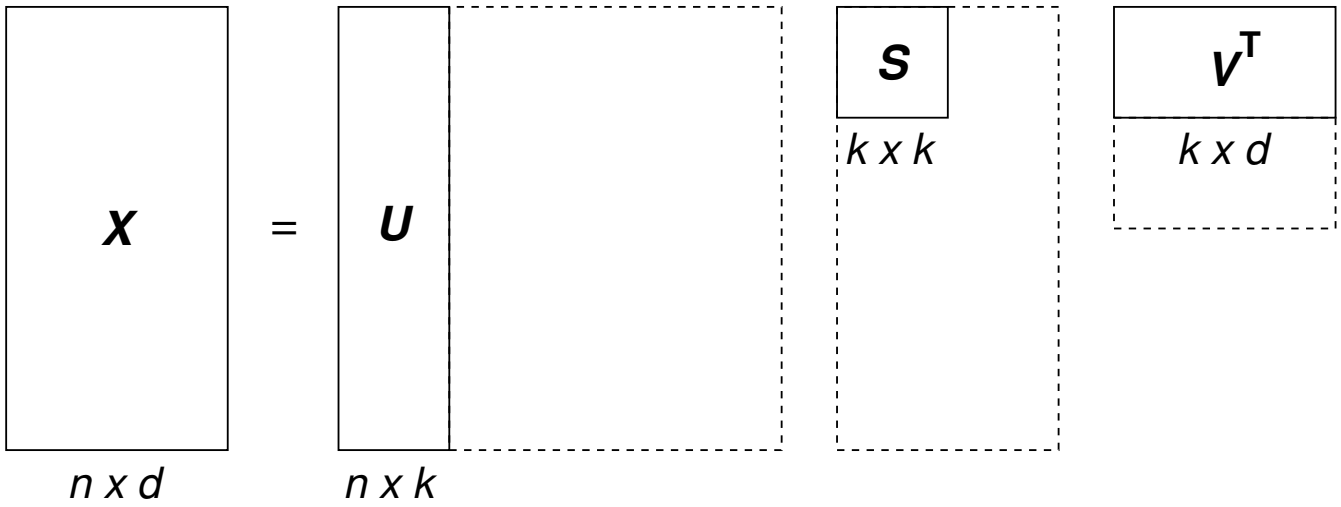
Decomposição em Valores Singulares

- Como as matrizes $\boldsymbol{U}$ e $\boldsymbol{V}$ são ortogonais:
  - $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_{n,n}$ (matriz identidade)
  - $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_{d,d}$ (matriz identidade)

Fatoração de Matrizes

Decomposição em Valores Singulares

- Os valores singulares da matriz $\boldsymbol{S}$ podem ser pensados como valores de importância para diferentes características na matriz.
- A decomposição em valores singulares pode ser utilizada na redução de dimensionalidade:



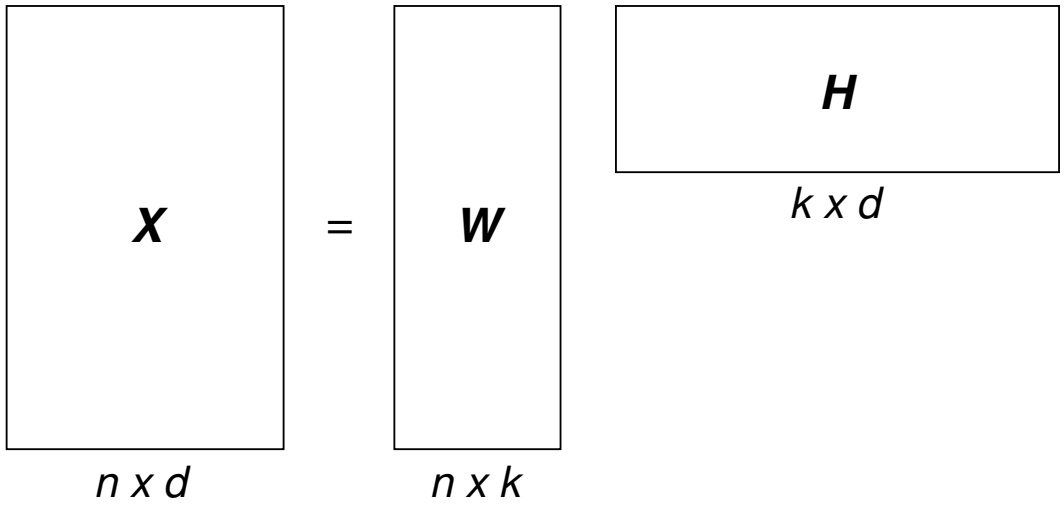
- Apenas k vetores colunas de $\boldsymbol{U}$ e k vetores linhas de $\boldsymbol{V}^T$ correspondentes aos k maiores valores singulares de $\boldsymbol{S}$ são calculados.

Fatoração de Matrizes

Fatoração de Matrizes Não-Negativas

- Uma matriz $\boldsymbol{X}_{n,d}$ é decomposta em duas matrizes $\boldsymbol{W}_{n,k}$ e $\boldsymbol{H}_{k,d}$:

$$\boldsymbol{X} \approx \boldsymbol{W}\,\boldsymbol{H}$$



- As três matrizes contêm apenas números não negativos.
- A matriz $\boldsymbol{W}$ pode ser interpretada como um conjunto de pesos e a matriz $\boldsymbol{H}$ como um conjunto de componentes (bases).

Fatoração de Matrizes

Fatoração de Matrizes Não-Negativas

- Uma amostra $\boldsymbol{x}_i$ pode ser pensada como uma soma ponderada de componentes, ou seja:

$$\boldsymbol{x}_i \approx [w_{i,1}, w_{i,2}, \ldots, w_{i,k}][h_1, h_2, \ldots, h_k]^T = \sum_{j=1}^{k} w_{i,j}\, h_i$$

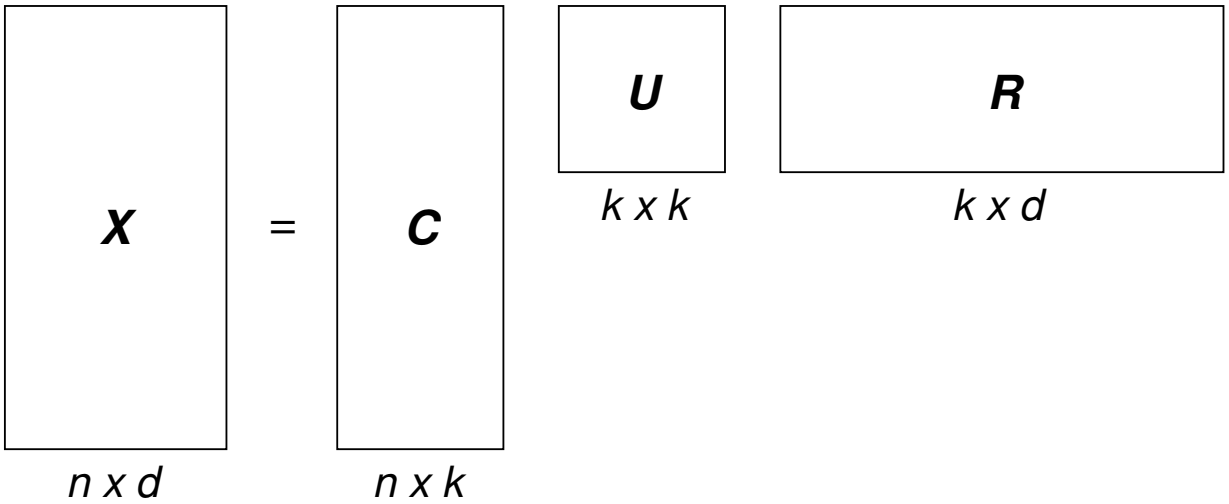
- Dessa forma, a fatoração de matrizes não negativas decompõe cada amostra de dados em uma combinação de componentes.

Fatoração de Matrizes

Decomposição CUR de Matrizes

- Uma matriz $\boldsymbol{X}_{n,d}$ é decomposta em três matrizes $\boldsymbol{C}_{n,k}$, $\boldsymbol{U}_{k,k}$ e $\boldsymbol{R}_{k,d}$:

$$\boldsymbol{X} \approx \boldsymbol{C}\,\boldsymbol{U}\,\boldsymbol{R}$$



- A matriz $\boldsymbol{C}$ é formada por colunas de $\boldsymbol{X}$, a matriz $\boldsymbol{R}$ é formada por linhas de $\boldsymbol{X}$, enquanto a matriz $\boldsymbol{U}$ é construída de forma que o produto das três matrizes aproxime a matriz $\boldsymbol{X}$.

Fatoração de Matrizes

Decomposição CUR de Matrizes

- Exemplo:

$$\begin{bmatrix} 3 & 1 & 7 & 5 \\ 2 & 9 & 8 & 4 \\ 7 & 6 & 8 & 6 \\ 2 & 1 & 4 & 6 \\ 9 & 7 & 6 & 3 \end{bmatrix} \approx \underbrace{\begin{bmatrix} 7 & 3 & 1 \\ 8 & 2 & 9 \\ 8 & 7 & 6 \\ 4 & 2 & 1 \\ 6 & 9 & 7 \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} -0.0669 & 0.0285 & 0.1819 \\ 0.1312 & -0.1037 & 0.0090 \\ 0.0299 & 0.1088 & -0.1625 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 9 & 7 & 6 & 3 \\ 2 & 9 & 8 & 4 \\ 3 & 1 & 7 & 5 \end{bmatrix}}_{R}$$

Análise de Componentes Principais

- Método[1] criado por Karl Pearson em 1901.
- Utiliza uma transformação ortogonal para converter um conjunto de observações de variáveis, possivelmente correlacionadas, em um conjunto de valores de variáveis linearmente não correlacionadas.
- As variáveis linearmente não correlacionadas são chamadas de componentes principais.
- A transformação é definida de forma que a primeira componente principal tenha a maior variância possível (ou seja, é responsável pelo máximo de variabilidade nos dados) e, cada componente seguinte, por sua vez, tenha a máxima variância sob a restrição de ser ortogonal (ou seja, não correlacionada) às componentes anteriores.

[1] Em inglês, a técnica é conhecida como *Principal Component Analysis* (PCA).

Análise de Componentes Principais

- Se apenas as primeiras componentes principais forem mantidas, a dimensionalidade dos dados transformados é reduzida.
- Tornou-se popular para a redução de dimensionalidade de dados.
- Possibilita encontrar uma aproximação dos dados originais utilizando um conjunto menor de atributos.
- Sua operação auxilia a identificação das dimensões que exibem as maiores variações em um conjunto de dados.

Análise de Componentes Principais

- Dado um conjunto $\boldsymbol{D}$ com n instâncias e d atributos, uma transformação linear do conjunto de atributos $(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_d)$ para um novo conjunto de atributos $(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_d)$ pode ser calculada como:

$$\begin{aligned}
\boldsymbol{Z}_1 &= a_{11}\boldsymbol{X}_1 + a_{21}\boldsymbol{X}_2 + \ldots + a_{d1}\boldsymbol{X}_d \\
\boldsymbol{Z}_2 &= a_{12}\boldsymbol{X}_1 + a_{22}\boldsymbol{X}_2 + \ldots + a_{d2}\boldsymbol{X}_d \\
\vdots &= \qquad\qquad \vdots \\
\boldsymbol{Z}_d &= a_{1d}\boldsymbol{X}_1 + a_{2d}\boldsymbol{X}_2 + \ldots + a_{dd}\boldsymbol{X}_d
\end{aligned}$$

- As componentes principais $\boldsymbol{Z}_i$ são tipos específicos de combinações lineares escolhidas de tal modo que sejam não correlacionadas (independentes).
- Em geral, apenas algumas das primeiras componentes principais são responsáveis pela maior parte da variabilidade do conjunto de dados.

Análise de Componentes Principais

- A análise de componentes principais pode ser reduzida ao problema de encontrar os autovalores e autovetores da matriz de covariância (ou correlação) do conjunto de dados.
- A proporção da variância do conjunto de dados originais explicada pela i-ésima componente principal é igual ao i-ésimo autovalor dividido pela soma de todos os d autovalores.
- Ou seja, as componentes principais são ordenadas decrescentemente de acordo com os autovalores.
- Quando os valores dos diferentes atributos estão em diferentes escalas, pode-se utilizar a matriz de correlação ao invés da matriz de covariância.

Análise de Componentes Principais

- Dada a matriz de dados $\boldsymbol{X} = \boldsymbol{D} \in \mathbb{R}^{n \times d}$, centralizamos os pontos para que fiquem com média zero:

$$x'_{ij} = x_{ij} - \overline{x}_j$$

em que $\overline{x}_j$ é a média dos valores do atributo j.

- A matriz de covariância dos atributos pode ser calculada como:

$$\boldsymbol{\Sigma} = \boldsymbol{X'}^T \boldsymbol{X'}$$

em que $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$. O elemento i, j dessa matriz representa a correlação entre o atributo i e o atributo j. A diagonal indica a variância do respectivo atributo.

- Dessa matriz de covariância, pode-se extrair um total de d autovalores $(\boldsymbol{\lambda})$ e autovetores $(\boldsymbol{V})$ tal que:

$$\boldsymbol{\Sigma} \cdot \boldsymbol{V} = \boldsymbol{\lambda} \cdot \boldsymbol{V} \quad \Rightarrow \quad \boldsymbol{\lambda} = \boldsymbol{V}^{-1} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{V}$$

Análise de Componentes Principais

- Se ordenarmos decrescentemente todos os autovetores conforme os autovalores, tem-se que:
  - Cada autovetor i representa a i-ésima direção de maior variação.
  - O autovalor correspondente quantifica essa variação.
- Cada autovetor representa uma combinação linear dos atributos originais de tal forma a capturar a variação descrita pelo autovalor.
- Basicamente, a matriz de autovetores é uma base de dados após rotação que captura a variação em ordem crescente.
- Se um autovalor for muito pequeno, significa que não existe variação naquele eixo e, portanto, ele pode ser descartado.
- Imagine um problema de classificação utilizando apenas uma variável x_j com variância baixa. É fácil perceber que tal variável não tem poder discriminatório pois, para toda classe, ela apresenta um valor muito similar.

Análise de Componentes Principais

- De posse da matriz $\boldsymbol{V} \in \mathbb{R}^{d \times k}$ dos k primeiros autovetores com um valor significativo de $\boldsymbol{\lambda}$, é possível transformar a matriz de dados centralizada $\boldsymbol{X}'$ com:

$$\hat{\boldsymbol{X}} = \boldsymbol{V}^T \cdot \boldsymbol{X}'$$

- Isso transforma a matriz $\boldsymbol{X}'$ em uma matriz $\hat{\boldsymbol{X}} \in \mathbb{R}^{n \times k}$ com $k < d$.
- A projeção corresponde a uma transformação de rotação. Portanto, para retornar novamente aos dados, basta fazer:

$$\boldsymbol{X}' = \boldsymbol{V} \cdot \hat{\boldsymbol{X}}$$

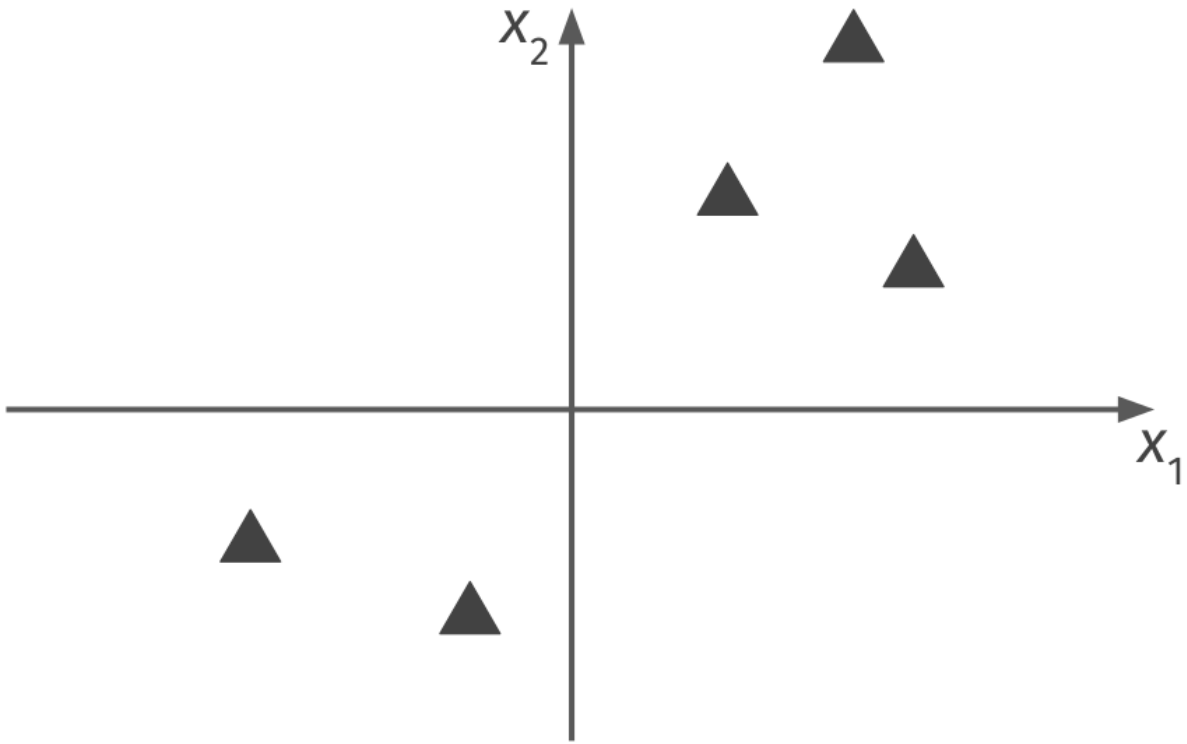
Análise de Componentes Principais

Ilustração dos passos fundamentais da análise de componentes principais:

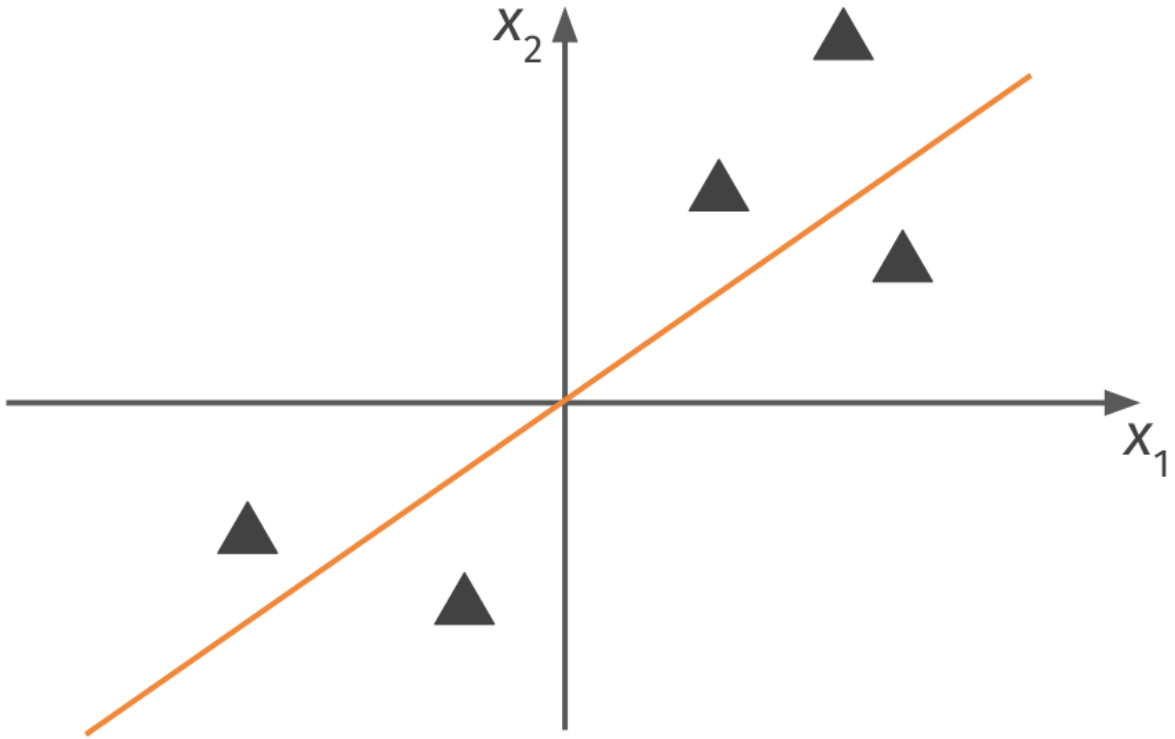
- Identificação do hiperplano que está mais próximo dos dados.
- Projeção dos dados para o hiperplano.

- As direções que contêm a maior variação dos dados são determinadas.
- Essas direções, chamadas de componentes principais, são ordenadas conforme valor de variação.
- As componentes principais são ortogonais (perpendiculares) entre si.

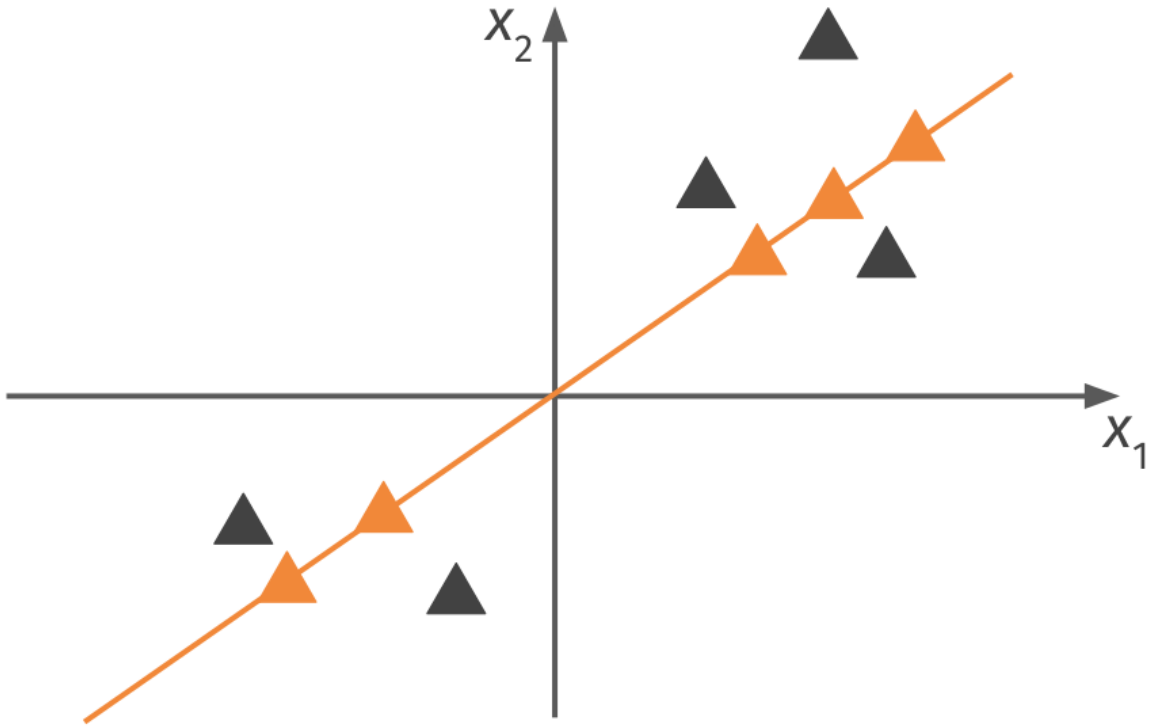
Análise de Componentes Principais



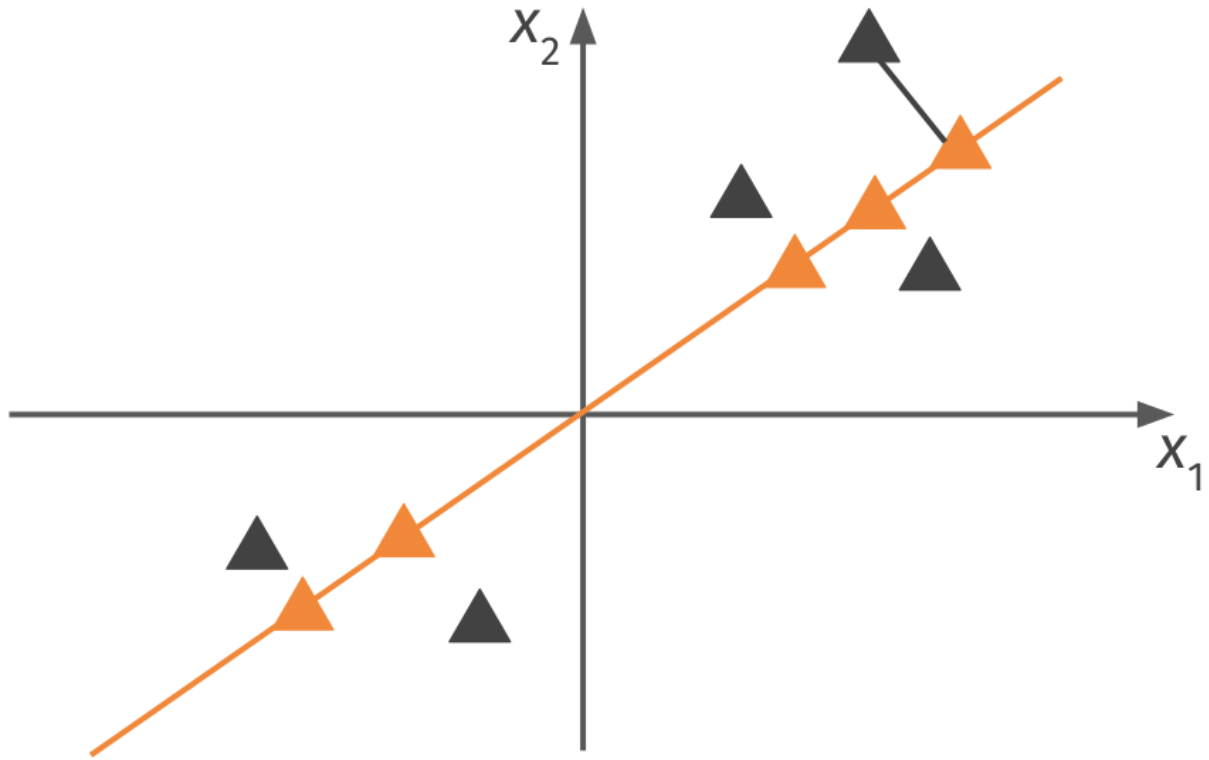
Análise de Componentes Principais



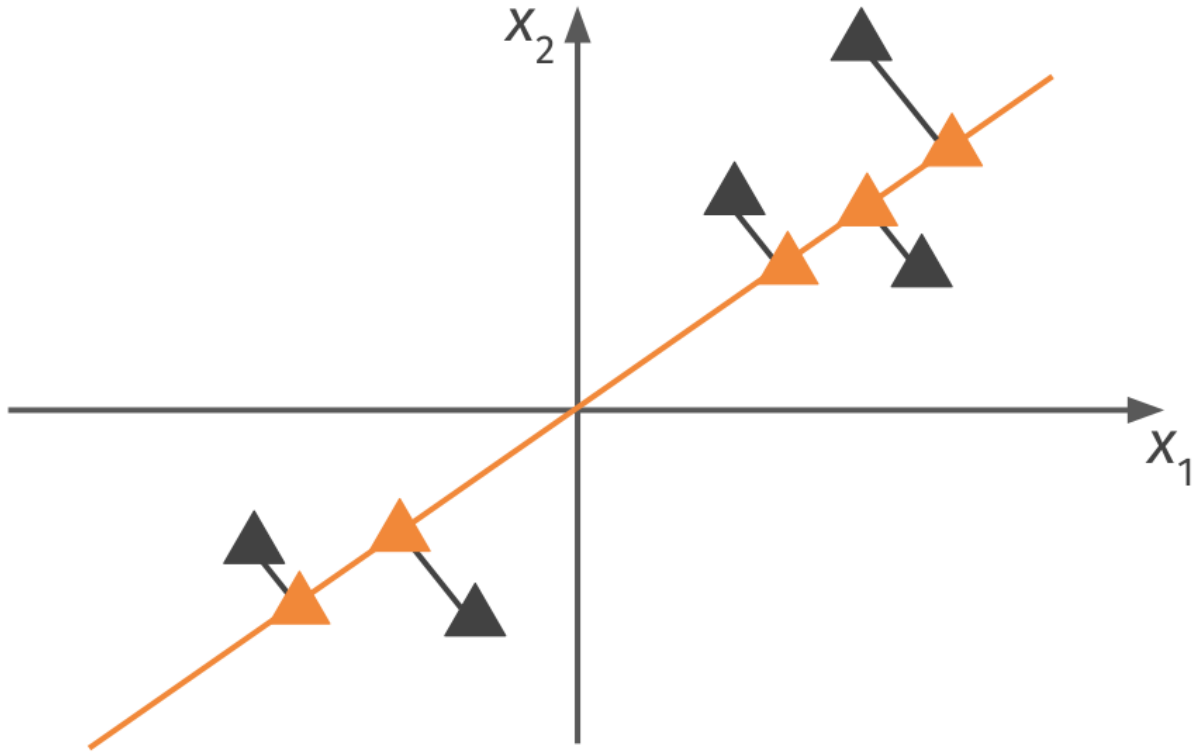
Análise de Componentes Principais



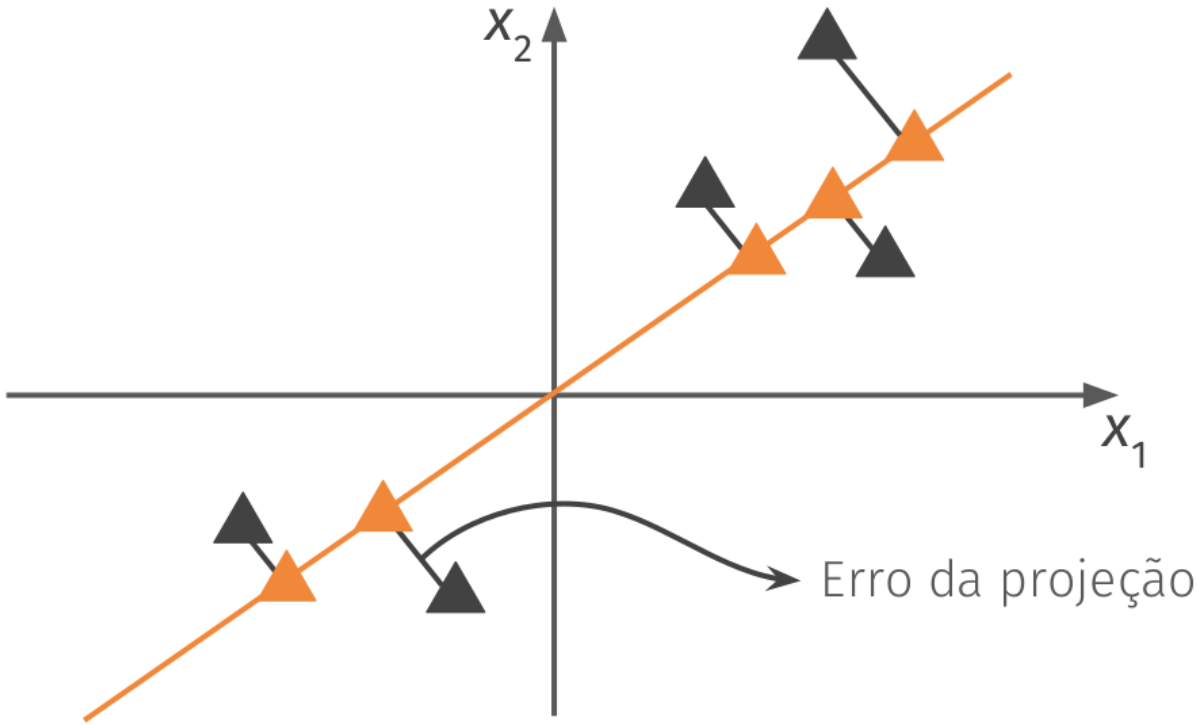
Análise de Componentes Principais



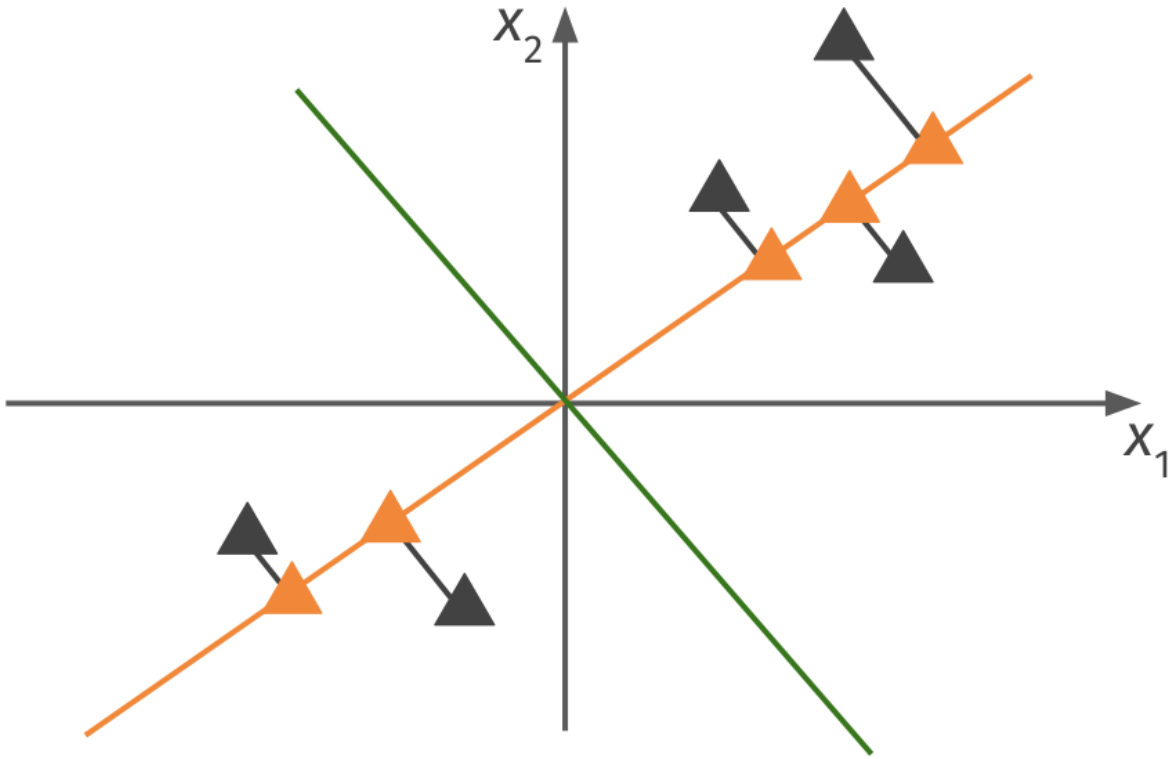
Análise de Componentes Principais



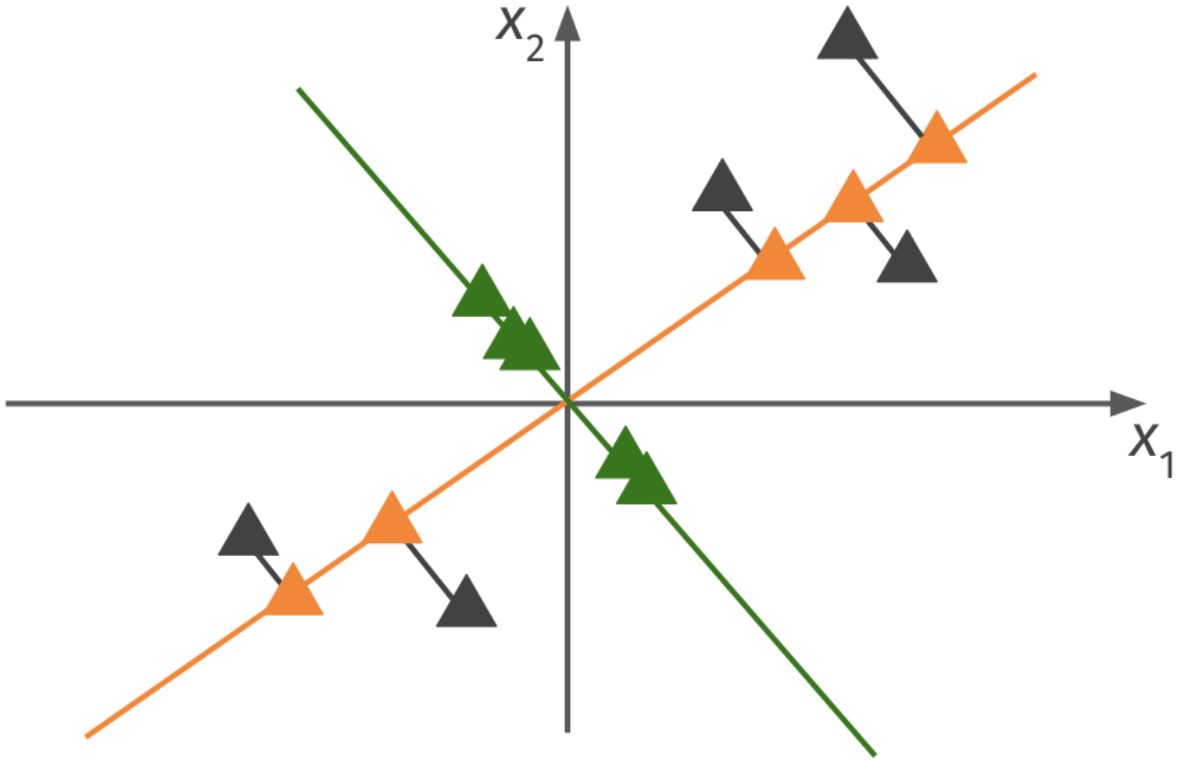
Análise de Componentes Principais



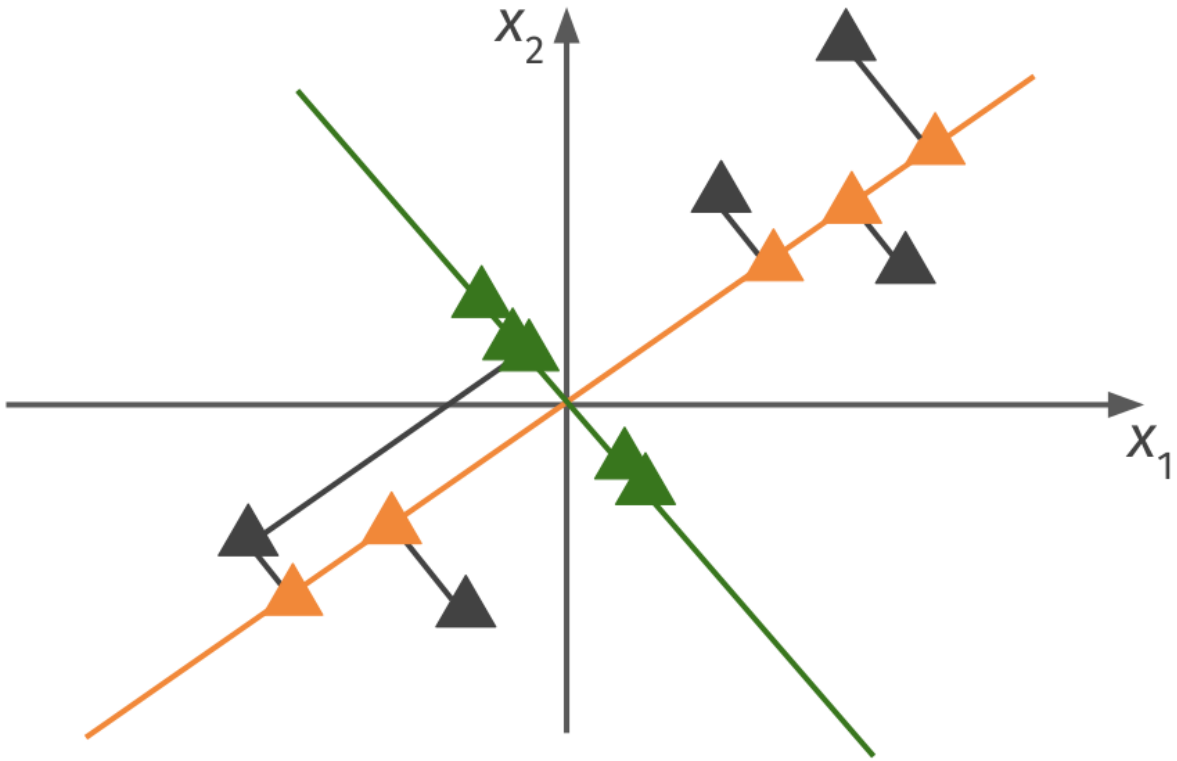
Análise de Componentes Principais



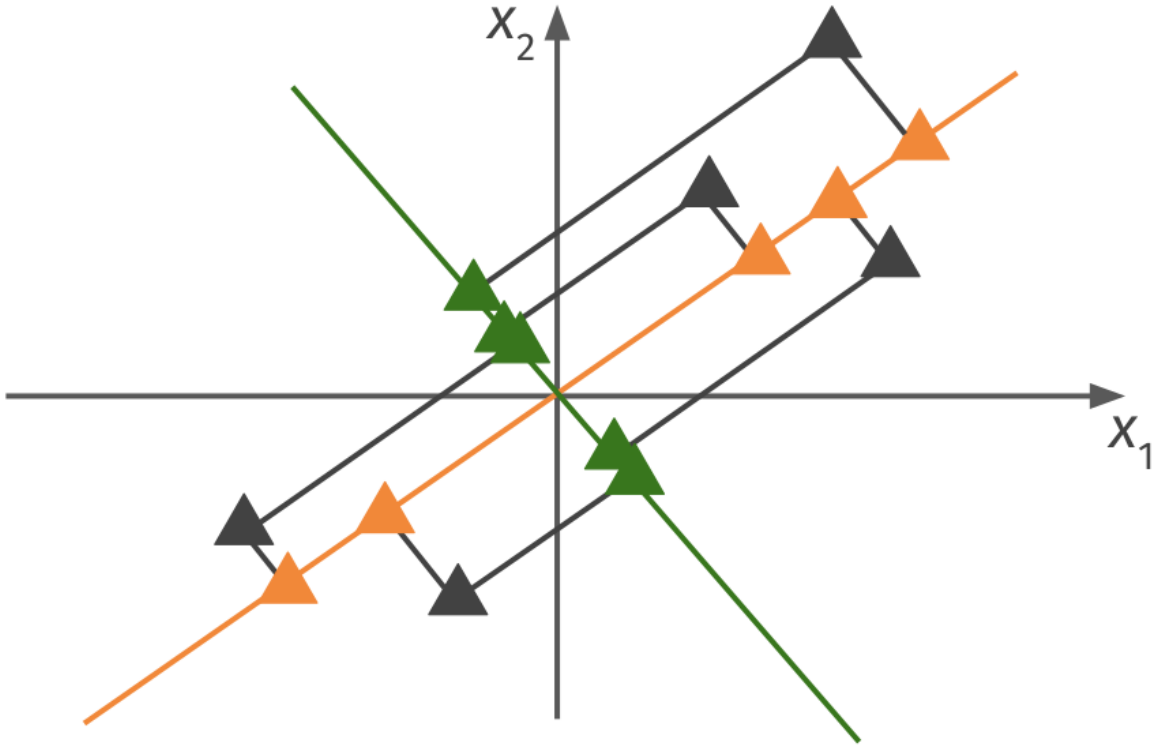
Análise de Componentes Principais



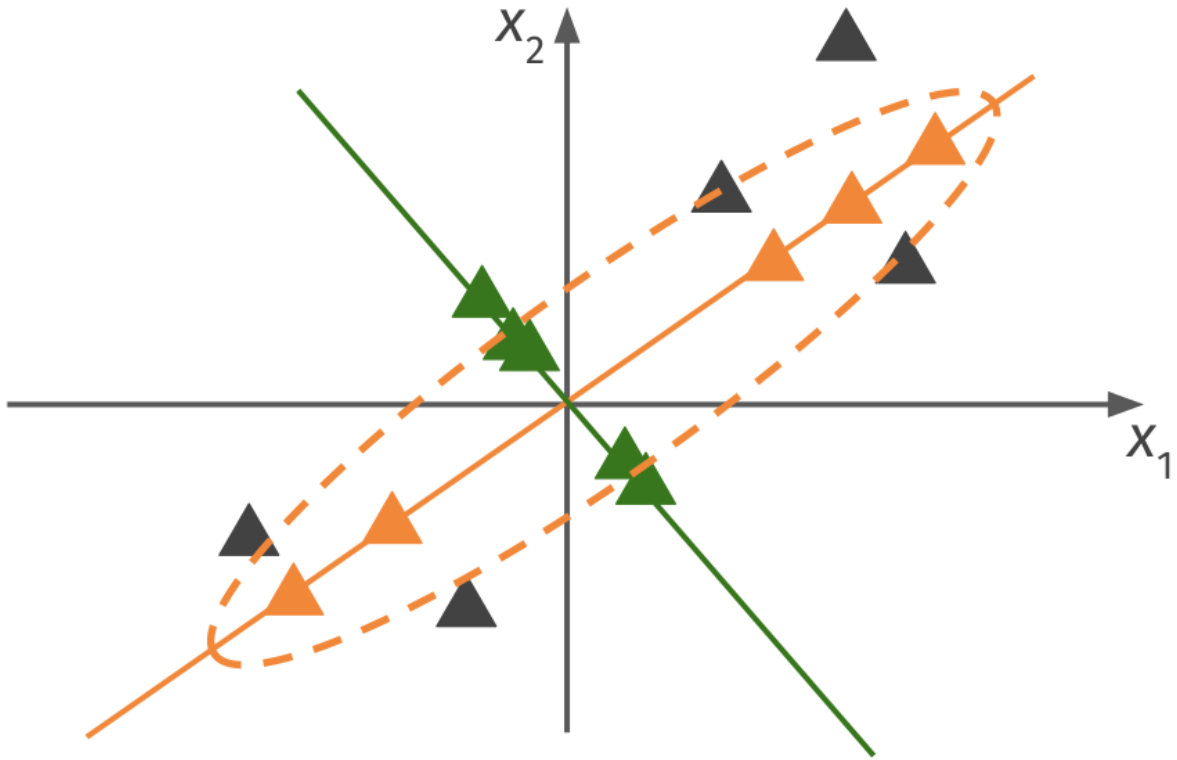
Análise de Componentes Principais



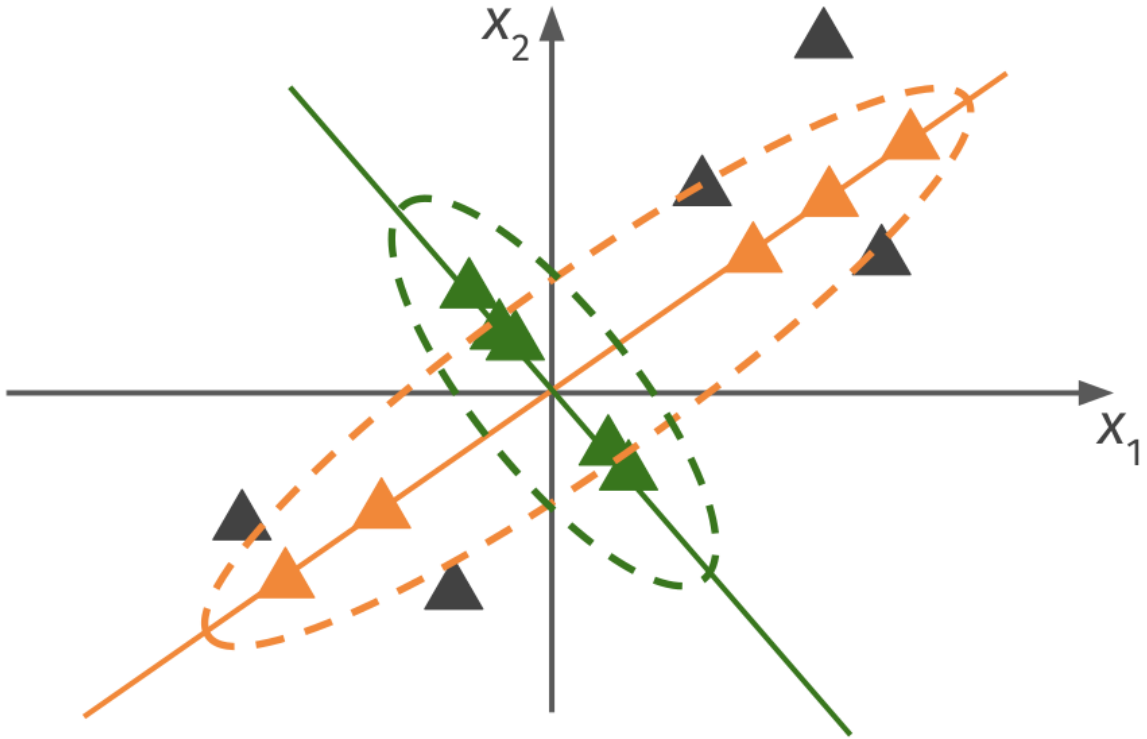
Análise de Componentes Principais



Análise de Componentes Principais

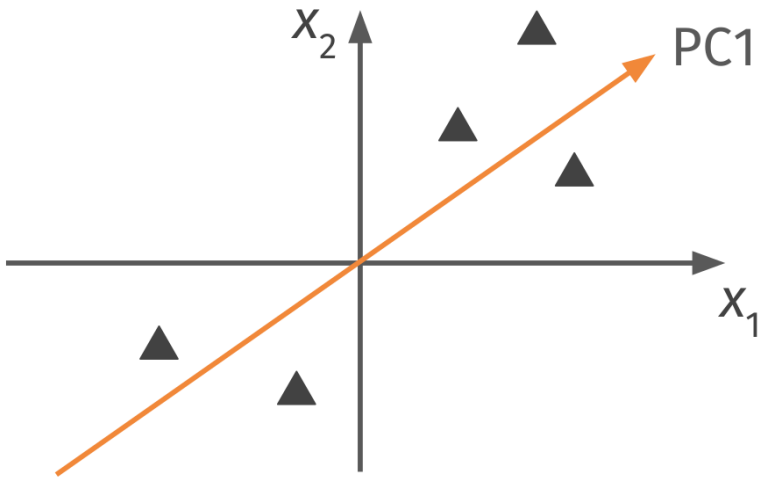


Análise de Componentes Principais



Análise de Componentes Principais

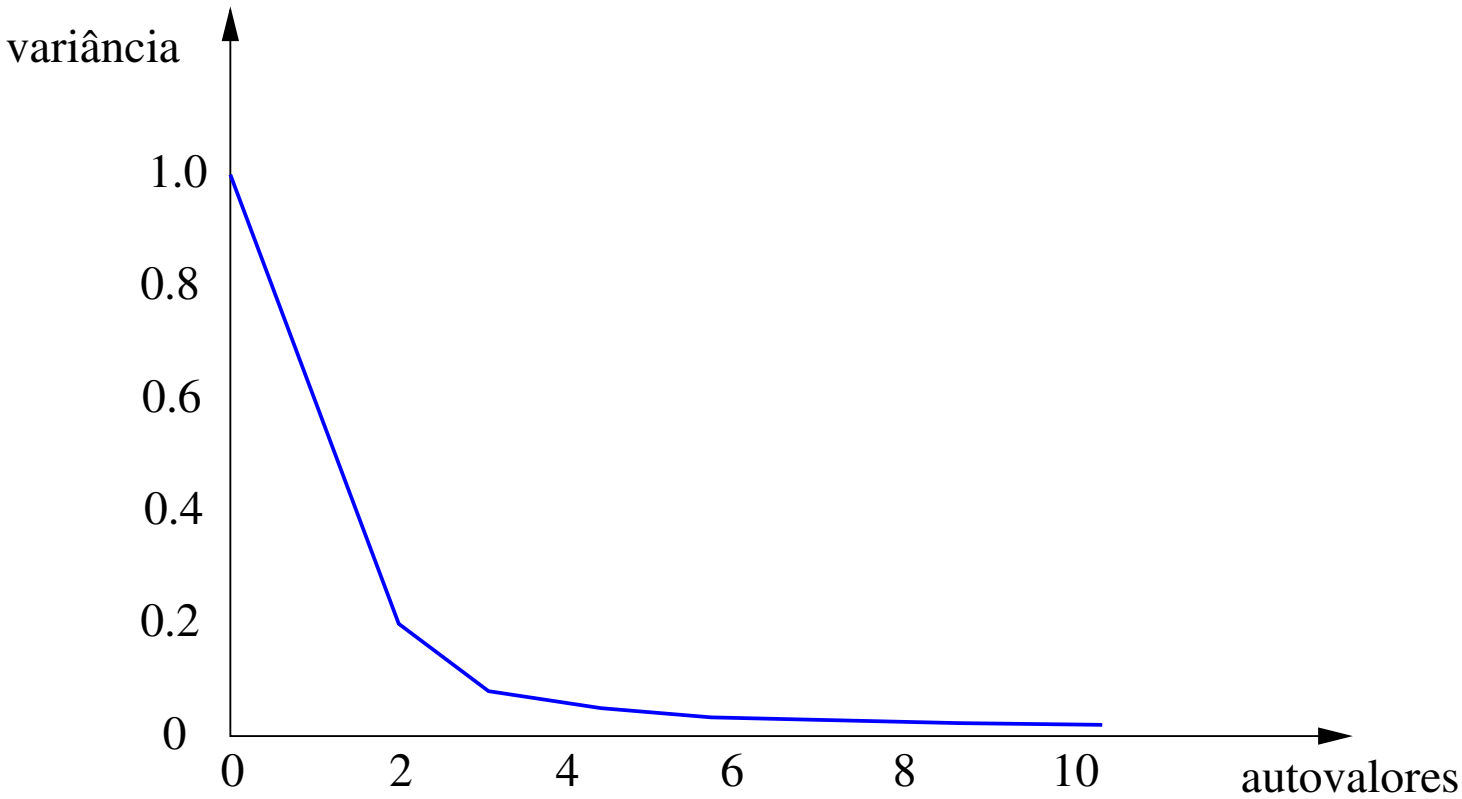
- Para reduzir de 2 dimensões para 1 dimensão:
  - Encontre uma direção (um vetor) que minimiza o erro da projeção dos dados.



- Para reduzir de n dimensões para k dimensões:
  - Encontre k vetores que minimizam o erro da projeção dos dados.

Redução da Dimensionalidade

Gráfico da fração da variância geral dos dados calculada para cada autovalor (componente principal) da matriz de covariância.



Análise de Componentes Principais

Algoritmo:

- Entrada: matriz $\boldsymbol{X}$ de dados $(n \times d)$, em que cada linha é um vetor x_i.
- Saída: matriz $\hat{\boldsymbol{X}}$ $(n \times k)$.

1. Calcular as médias das colunas (dimensões) de $\boldsymbol{X}$, formando o vetor de médias $\overline{\boldsymbol{X}}$.
2. Subtrair o vetor $\overline{\boldsymbol{X}}$ de cada linha de $\boldsymbol{X}$, formando a matriz $\boldsymbol{X}'$.
3. Calcular a matriz de covariância $\boldsymbol{\Sigma}$ de $\boldsymbol{X}'$.
4. Calcular e ordenar, decrescentemente pelos autovalores, os autovetores $\boldsymbol{v}_i$ de $\boldsymbol{\Sigma}$, formando a matriz $\boldsymbol{V}$ dos autovetores.
5. Selecionar os k primeiros autovetores em $\boldsymbol{V}$, formando $\boldsymbol{V}_k$.
6. Calcular a matriz $n \times k$ dos dados de saída $\hat{\boldsymbol{X}} = \boldsymbol{V}_k^T \boldsymbol{X}'$.

Análise de Componentes Principais

Vantagens:

- Alto poder de representação.
- Redução do custo de armazenamento.
- Fácil implementação.
- Robusta e largamente estudada.

Análise de Componentes Principais

Desvantagens:

- Assume apenas relações lineares entre os atributos.
- A interpretação dos atributos transformados torna-se mais difícil, pois seu significado original é alterado.
- Nem sempre é fácil determinar o valor de k.
- Não considera as classes das amostras (não é ótima para classificação).

Análise de Componentes Principais

Aplicações:

- Reconhecimento de faces.
- Reconstrução de imagens.
- Compressão de dados.
- Visualização de dados multidimensionais.

Análise de Componentes Principais com Núcleo

- Esta técnica[2] é uma extensão da Análise de Componentes Principais para permitir o uso de transformações não lineares (chamadas de núcleos).
- Essa transformação mapeia o espaço original dos dados para um novo espaço de atributos, de maior dimensionalidade, de forma que passem a ser linearmente separáveis.
- A análise de componentes principais é então implicitamente aplicada nesse espaço de maior dimensão, o qual não é linearmente relacionado ao espaço original.

[2] Em inglês, a técnica é conhecida como *Kernel* PCA.

Análise de Componentes Independentes

- A técnica[3] utiliza a informação mútua (conceito semelhante à entropia) entre as componentes para torná-las maximamente independentes.
- Portanto, a informação mútua entre as componentes resultantes é zero.
- Não impõe restrição de ortogonalidade.
- Assim como na análise de componentes principais, a interpretação dos atributos transformados não é intuitiva.
- Uma aplicação comum é a separação de sinais de áudio.

[3] Em inglês, a técnica é conhecida como *Independent Component Analysis* (ICA).

Mapeamento Topológico Localmente Linear

- Preserva[4] relações de vizinhança dos dados de entrada quando mapeados para um espaço de baixa dimensão, ou seja, procura manter a estrutura local dos dados de alta dimensão em uma nova representação de baixa dimensão.
- Cada ponto é expresso como uma combinação linear de pontos vizinhos.
- A reconstrução de cada ponto é realizada a partir dos seus vizinhos por meio de pesos apropriados.
- Esses pesos capturam as propriedades geométricas intrínsecas das vizinhanças locais.

[4] Em inglês, a técnica é conhecida como *Locally Linear Embedding* (LLE).

Escalonamento Multidimensional Métrico

- Método[5] de redução de dimensionalidade linear.
- Visa encontrar uma projeção dos dados em um espaço de dimensão menor que preserve as distâncias entre pares de pontos tão bem quanto possível.
- Isso é feito encontrando um conjunto de coordenadas para cada ponto no espaço de baixa dimensão que minimiza a diferença entre as distâncias originais e as distâncias na nova representação de baixa dimensão.
- No modelo geométrico procurado, quanto maior a distância observada ou dissimilaridade entre duas observações (ou menor a similaridade), mais afastados devem estar os pontos que as representam no modelo espacial.
- Assume-se normalmente que a distância entre os pontos no modelo espacial é Euclidiana.

[5] Em inglês, a técnica é conhecida como *Multidimensional Scaling* (MDS).

Mapeamento Não Linear de Sammon

- Similar ao Escalonamento Multidimensional Métrico, entretanto, esta técnica[6] utiliza uma função de custo com um fator inversamente proporcional à distância nos dados de entrada.
- Dessa forma, a preservação de distâncias longas é menos importante do que a preservação de distâncias mais curtas.
- O algoritmo inicia com uma configuração inicial de pontos no espaço de baixa dimensão e, em seguida, itera até encontrar uma configuração que minimize a função de erro. Cada iteração envolve a atualização das coordenadas dos pontos no espaço de baixa dimensão com base nas distâncias relativas entre os pontos.
- Nenhuma hipótese é feita em relação a qual tipo de função de distância utilizar, embora a distância Euclidiana seja geralmente escolhida.
- Método muito utilizado para visualização bidimensional de dados multivariados.

[6] Em inglês, a técnica é conhecida como *Sammon Mapping*.

FastMap

- Método iterativo de redução de dimensionalidade não linear que realiza um mapeamento de pontos em um espaço dimensional por um conjunto de eixos, em que cada eixo é definido por um par de pontos (pivôs) mais afastados obtidos do conjunto de dados.
- Para encontrar pontos mais afastados entre si, seria necessário computar as distâncias entre cada par de pontos, resultando em um algoritmo de complexidade quadrática pelo número de cálculos de distância.
- O método utiliza uma heurística para encontrar os pares de pontos cujas distâncias são próximas àquelas dos pontos mais distantes, levando a um algoritmo de complexidade linear.
- A aplicação da função de distância Euclidiana permite que as projeções dos pontos possam ser calculadas utilizando a lei dos cossenos.
- Intuitivamente, o método trata cada distância entre pares de pontos como uma mola, buscando rearranjar as posições dos pontos de forma a minimizar as deformações na mola.

Mapeamento Isométrico (IsoMap)

- Método de redução de dimensionalidade não linear que utiliza o conceito de distância geodésica em um espaço de alta dimensão para preservar as distâncias entre os pontos na representação de baixa dimensão.
- O algoritmo inicia com a construção de um grafo de vizinhança a partir dos dados, em que cada ponto é conectado aos seus k-vizinhos mais próximos.
- Em seguida, o algoritmo calcula as distâncias geodésicas entre todos os pares de pontos do grafo e constrói uma representação de baixa dimensão do grafo.
- Pode ser visto como uma generalização do método de Escalonamento Multidimensional Métrico: uma diferença entre eles é que o IsoMap utiliza distância por grafos, enquanto que o método de Escalonamento Multidimensional Métrico utiliza distância Euclidiana.
- Tal diferença torna o método IsoMap não linear.
- Pontos de alta dimensionalidade próximos são mapeados mais perto, enquanto pontos de alta dimensionalidade distantes são mapeados mais longe, de acordo com a distância geodésica.

Mapas Auto-Organizáveis

- Tipo de rede neural artificial baseada em aprendizado competitivo e não supervisionado[7].
- Método de redução de dimensionalidade não linear que busca a preservação de topologia do espaço original.
- O método realiza um mapeamento dos dados de entrada em um conjunto de neurônios organizados em uma grade bidimensional. Cada neurônio no mapa representa uma região do espaço de entrada.
- Durante o treinamento, os pesos dos neurônios são ajustados para que neurônios que estejam próximos na grade respondam a entradas similares. Isso significa que os neurônios no mapa organizam-se automaticamente em regiões que respondem a entradas similares.

[7] Em inglês, a técnica é conhecida como *Self-Organizing Maps* (SOM).

Mapas Auto-Organizáveis

- O processo de treinamento é realizado em duas etapas principais: inicialização e ajuste fino. Na fase de inicialização, os pesos dos neurônios são inicializados aleatoriamente e a rede é apresentada com exemplos de entrada. Na fase de ajuste fino, a rede ajusta gradualmente os pesos dos neurônios para que os neurônios que respondem a entradas similares sejam agrupados na mesma região do mapa.
- Uma vez que o mapa tenha sido treinado, os dados de entrada podem ser mapeados para o espaço de menor dimensão representado pelo mapa. Cada entrada é mapeada para o neurônio no mapa que fornece a melhor resposta para a entrada. O resultado é uma representação bidimensional dos dados de entrada em que os padrões similares são mapeados para regiões próximas umas das outras no mapa.

t-Distributed Stochastic Neighbor Embedding (t-SNE)

- Técnica não linear para redução de dimensionalidade que é particularmente adequada para a visualização de conjuntos de dados de alta dimensão.
- Baseia-se na construção de uma distribuição de probabilidade sobre pares de pontos nos dados de entrada e uma distribuição de probabilidade similar no espaço de menor dimensão em que os dados são mapeados. O objetivo é encontrar um mapeamento que minimize a diferença entre essas duas distribuições de probabilidade.
- Distâncias Euclidianas em alta dimensionalidade entre pontos são convertidas para probabilidades condicionais que representam similaridades.
- A similaridade entre dois pontos x_i e x_j é a probabilidade condicional, $P(x_j \mid x_i)$, de que o ponto x_i escolheria o ponto x_j como seu vizinho se os vizinhos fossem selecionados em proporção à sua densidade de probabilidade sob uma distribuição normal centrada em x_i.

t-Distributed Stochastic Neighbor Embedding (t-SNE)

- Se os pontos mapeados y_i e y_j corretamente modelarem a similaridade entre os pontos de alta dimensionalidade x_i e x_j, as probabilidades condicionais $P_1(x_j \mid x_i)$ e $P_2(x_j \mid x_i)$ serão iguais.
- A partir dessa observação, a técnica visa encontrar uma representação de baixa dimensionalidade que minimiza a diferença entre essas probabilidades condicionais (ou similaridades).
- A técnica minimiza a soma das divergências de Kullback-Leibler sobre todos os pontos de dados utilizando um método de gradiente descendente para ajustar o mapeamento no espaço de menor dimensão.
- A variância da distribuição normal utilizada para calcular as similaridades no espaço de maior dimensão é definida a partir de um hiperparâmetro da técnica, chamado de perplexidade (valor definido pelo usuário), que pode ser interpretado pela quantidade de vizinhos muito próximos que cada ponto tem.

Uniform Manifold Approximation and Projection (UMAP)

- Técnica de redução de dimensionalidade não linear que assume que os dados são uniformemente distribuídos em uma variedade topológica de Riemann (uma generalização do conceito métrico do espaço Euclidiano) localmente conectada.
- A técnica inicialmente converte os dados de entrada em uma matriz de similaridade, que pode ser calculada de várias maneiras diferentes, tais como a distância Euclidiana ou uma medida de similaridade baseada em correlação.
- Um grafo ponderado de k vizinhos é construído a partir da matriz de similaridade, cujas arestas conectam pontos de dados com seus vizinhos mais próximos. Uma estrutura de baixa dimensionalidade é então calculada para preservar as propriedades locais e globais do grafo.
- Quando mais vizinhos são considerados ao redor de um ponto, o mapeamento captura estrutura mais global. Quando menos vizinhos são considerados, o mapeamento preserva estruturas mais locais.

Algumas Considerações Finais

- A técnica t-SNE é computacionalmente cara e pode demandar muito mais tempo de execução do que as técnicas UMAP e Análise de Componentes Principais.
- Análise de Componentes Principais é uma técnica determinística, enquanto t-SNE e UMAP são técnicas probabilísticas.
- Redução de dimensionalidade não linear (como realizada pelas técnicas t-SNE e UMAP) pode representar relacionamentos complexos de atributos que nem sempre são possíveis com algoritmos lineares (como Análise de Componentes Principais).

Análise de Componentes Principais

Exemplo: Compressão de Imagens

- A técnica de análise de componentes principais pode ser aplicada no contexto de processamento de imagens.
- Manter apenas algumas das componentes principais pode resultar em uma imagem de menor qualidade, entretanto, que requer menor capacidade de armazenamento.
- A técnica é aplicada separadamente em cada banda de cor, cujos valores de intensidade estão no intervalo entre 0 e 255.
- Considerando apenas algumas componentes principais, a imagem colorida é gerada novamente.

Análise de Componentes Principais

- Alguns resultados da compressão de uma imagem mantendo-se diferentes números (k) de componentes principais:

imagem original

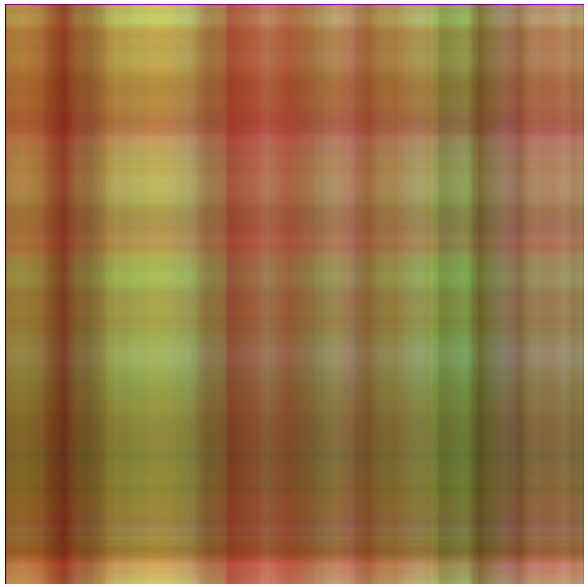


imagem comprimida (k=1)

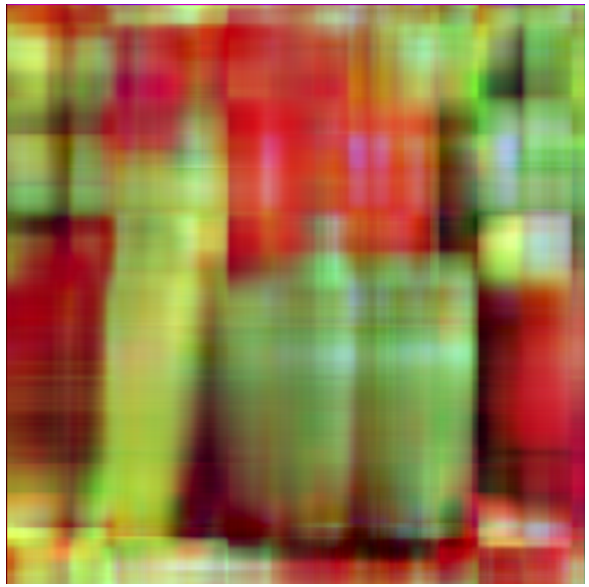


imagem comprimida (k=5)

Análise de Componentes Principais

imagem comprimida (k=10)

imagem comprimida (k=20)

imagem comprimida (k=30)

Análise de Componentes Principais

imagem comprimida (k=40)

imagem comprimida (k=50)

Distância Euclidiana

- Dadas duas amostras $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)$ e $\boldsymbol{y} = (y_1, y_2, \ldots, y_d)$ em um espaço d-dimensional, a distância Euclidiana é expressa como:

$$D(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_d - y_d)^2}$$

$$= \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}$$

Variância e Covariância

- A variância mede a variação de uma única variável aleatória (característica) X_i (por exemplo, altura de uma pessoa em uma população). Ela é expressa como:

$$\sigma(X_i)^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_i)^2$$

em que n é o número de amostras e $\overline{X}_i$ é a média da variável aleatória X_i (representada como um vetor).

- A covariância é uma medida de quanto duas variáveis aleatórias variam em conjunto (por exemplo, a altura e o peso de uma pessoa em uma população). Ela é expressa como:

$$\sigma(X_i, X_j) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X_i})(X_j - \overline{X}_j)$$

em que X_i e X_j são duas variáveis (características) com n amostras. A variância $\sigma(X_i)^2$ de uma variável X_i pode também ser expressa como a covariância com ela própria por $\sigma(X_i, X_i)$.

Matriz de Covariância

- A matriz de covariância pode ser expressa como:

$$\boldsymbol{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X} - \overline{\boldsymbol{X}})^T(\boldsymbol{X} - \overline{\boldsymbol{X}})$$

  em que o conjunto de dados é representado pela matriz $\boldsymbol{X} \in \mathbb{R}^{n \times d}$ e $\overline{\boldsymbol{X}}$ é o vetor média.
- A matriz de covariância é quadrada, em que $\Sigma_{i,j} = \sigma(X_i, X_j)$, com $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$, tal que d descreve a dimensão dos dados (número de atributos).
- A matriz de covariância é simétrica, já que $\sigma(X_i, X_j) = \sigma(X_j, X_i)$.
- As entradas da diagonal da matriz de covariância são as variâncias, enquanto as outras entradas são as covariâncias.

Matriz de Covariância

- A matriz de covariância para dois atributos X_1 e X_2 (duas dimensões) é dada por:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2}^2 \\ \sigma_{2,1}^2 & \sigma_2^2 \end{bmatrix}$$

- A matriz de covariância para d dimensões é dada por:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \dots & \sigma_{1,d} \\ \sigma_{2,1} & \sigma_2^2 & \dots & \sigma_{2,d} \\ & & \ddots & \vdots \\ \dots & \dots & & \vdots \\ \sigma_{d,1} & \sigma_{d,2} & \dots & \sigma_d^2 \end{bmatrix}$$

Matriz de Covariância

- Para a base de dados `Iris`, considerando os atributos X_1 como o comprimento da sépala e X_2 como a largura da sépala, o vetor da média (total de 150 amostras) e a matriz de covariância são:

$$\overline{\boldsymbol{X}} = \begin{bmatrix} 5.8433 & 3.0573 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.6857 & -0.0424 \\ -0.0424 & 0.1900 \end{bmatrix}$$

- Ou seja, a variância para X_1 é $\sigma_1^2 = 0.6857$ e para X_2 é $\sigma_2^2 = 0.1900$, enquanto a covariância entre os dois atributos é $\sigma_{1,2} = \sigma_{2,1} = -0.0424$.

Matriz de Covariância

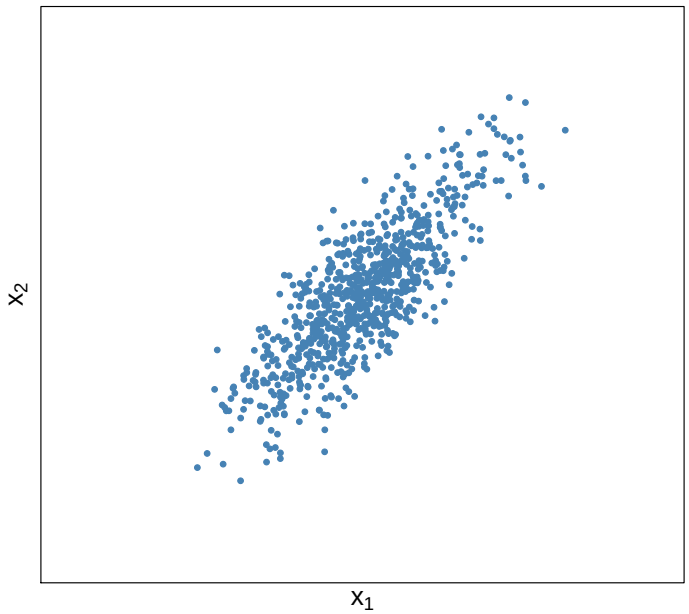
- Considerando agora os 4 atributos da base de dados `Iris`, X_1 como o comprimento da sépala, X_2 como a largura da sépala, X_3 como o comprimento da pétala e X_4 como a largura da pétala, o vetor da média (total de 150 amostras) e a matriz de covariância são:

$$\overline{\boldsymbol{X}} = \begin{bmatrix} 5.8433 & 3.0573 & 3.7580 & 1.1993 \end{bmatrix}$$

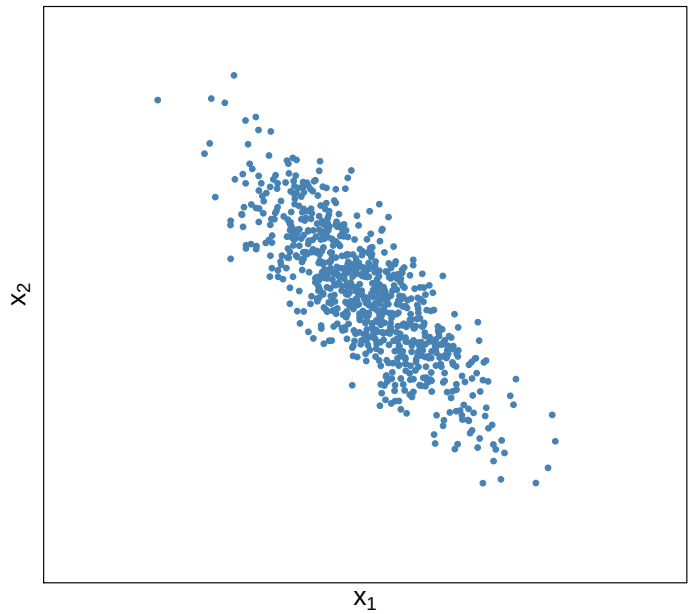
$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.6857 & -0.0424 & 1.2743 & 0.5163 \\ -0.0424 & 0.1900 & -0.3297 & -0.1216 \\ 1.2743 & -0.3297 & 3.1163 & 1.2956 \\ 0.5163 & -0.1216 & 1.2956 & 0.5810 \end{bmatrix}$$

Interpretação Geométrica da Matriz de Covariância

- A forma geral da distribuição dos dados define a matriz de covariância.



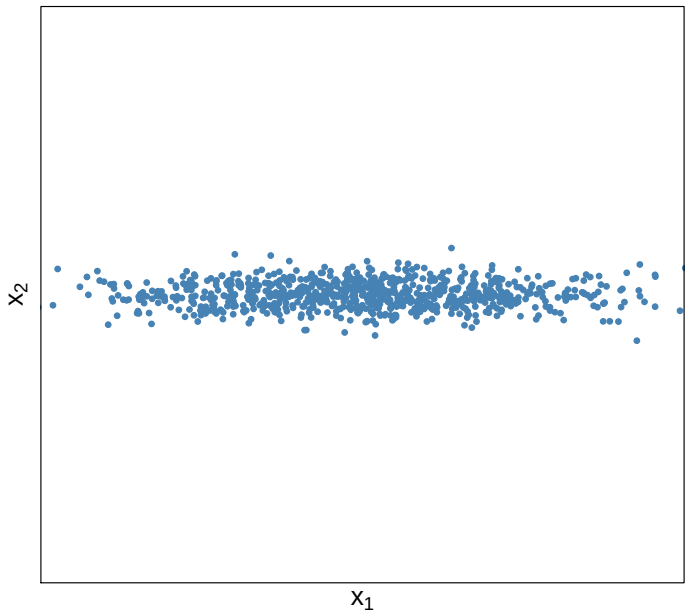
$$\boldsymbol{\Sigma} = \begin{bmatrix} 3 & 4 \\ 4 & 6 \end{bmatrix}$$

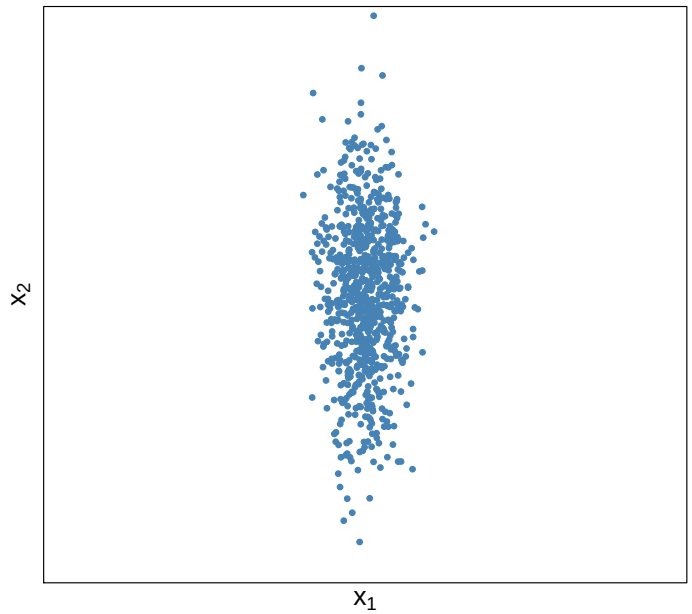


$$\boldsymbol{\Sigma} = \begin{bmatrix} 3 & -4 \\ -4 & 6 \end{bmatrix}$$

Interpretação Geométrica da Matriz de Covariância

- A forma geral da distribuição dos dados define a matriz de covariância.



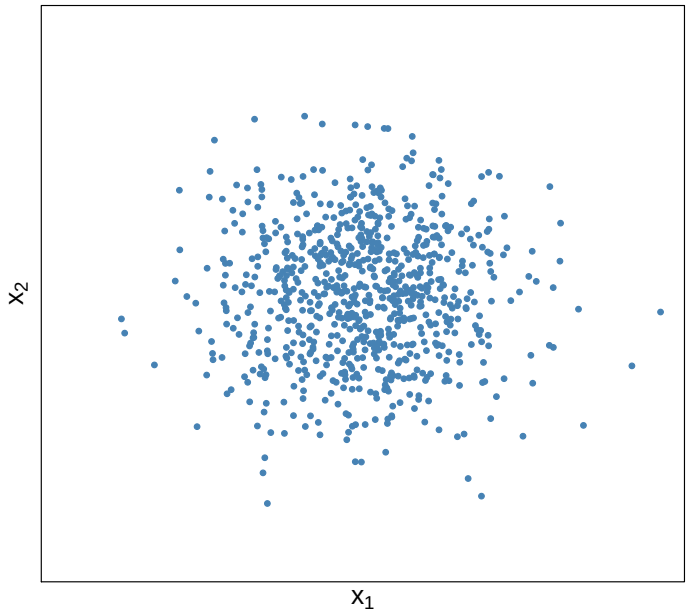


$$\boldsymbol{\Sigma} = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

Interpretação Geométrica da Matriz de Covariância

- A forma geral da distribuição dos dados define a matriz de covariância.



$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Interpretação Geométrica da Matriz de Covariância

- Dadas duas variáveis x_1 e x_2, a correlação $\rho(x_1, x_2)$ e a covariância $\sigma(x_1, x_2)$ podem ser relacionadas da seguinte forma:

$$\rho(x_1, x_2) = \frac{\sigma(x_1, x_2)}{\sigma(x_1)\sigma(x_2)}$$

  em que $\sigma(x_1)$ e $\sigma(x_2)$ são os desvios padrões de x_1 e x_2, respectivamente.
- A correlação está definida no intervalo $[-1, +1]$.
- A covariância está definida no intervalo $[-\infty, +\infty]$.

Autovalores e Autovetores

- Os autovalores e os autovetores podem ser calculados apenas a partir de uma matriz quadrada (matriz de covariância).
- Os autovetores são ortogonais entre si.
- No contexto de redução de dimensionalidade baseada na análise de componentes principais, os autovetores da matriz de covariância são calculados para encontrar as características mais representativas.
- Os autovetores da matriz de covariância representarão as novas características e serão escolhidos de acordo com seus autovalores.

Autovalores e Autovetores

- Considerando novamente os 4 atributos da base de dados `Iris`, X_1 como o comprimento da sépala, X_2 como a largura da sépala, X_3 como o comprimento da pétala e X_4 como a largura da pétala, os autovalores e autovetores da matriz de covariância são:

$$\boldsymbol{\lambda} = \begin{bmatrix} 4.2282 & 0.2427 & 0.0782 & 0.0238 \end{bmatrix}$$

$$\boldsymbol{V} = \begin{bmatrix} 0.3614 & 0.6566 & 0.5820 & 0.3155 \\ -0.0845 & 0.7302 & -0.5979 & -0.3197 \\ 0.8567 & -0.1734 & -0.0762 & -0.4798 \\ 0.3583 & -0.0755 & -0.5458 & 0.7537 \end{bmatrix}$$

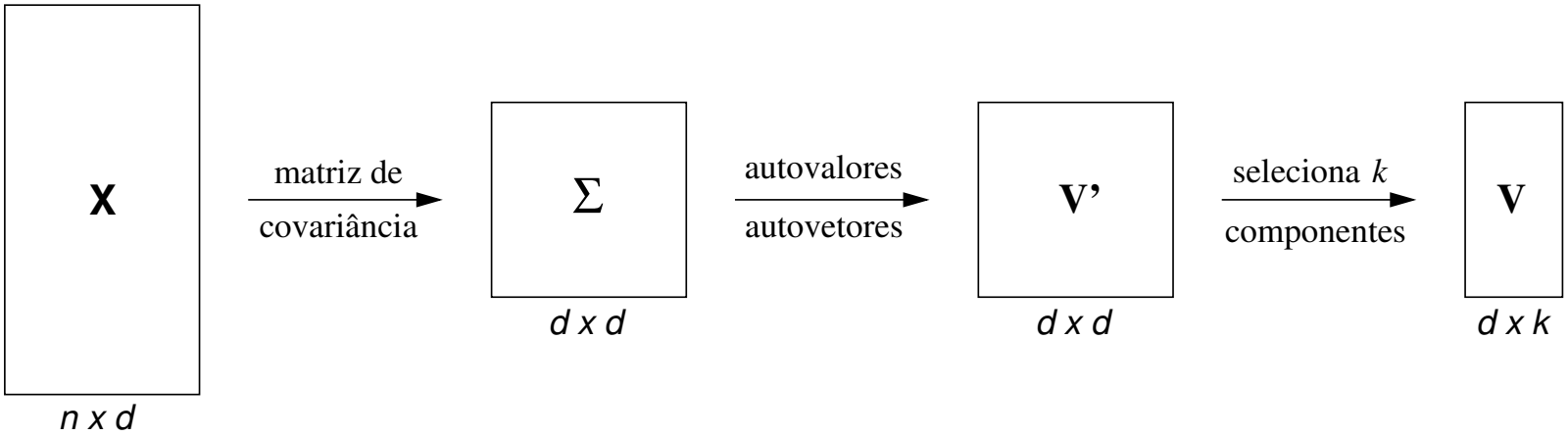
Escolha das Componentes Principais

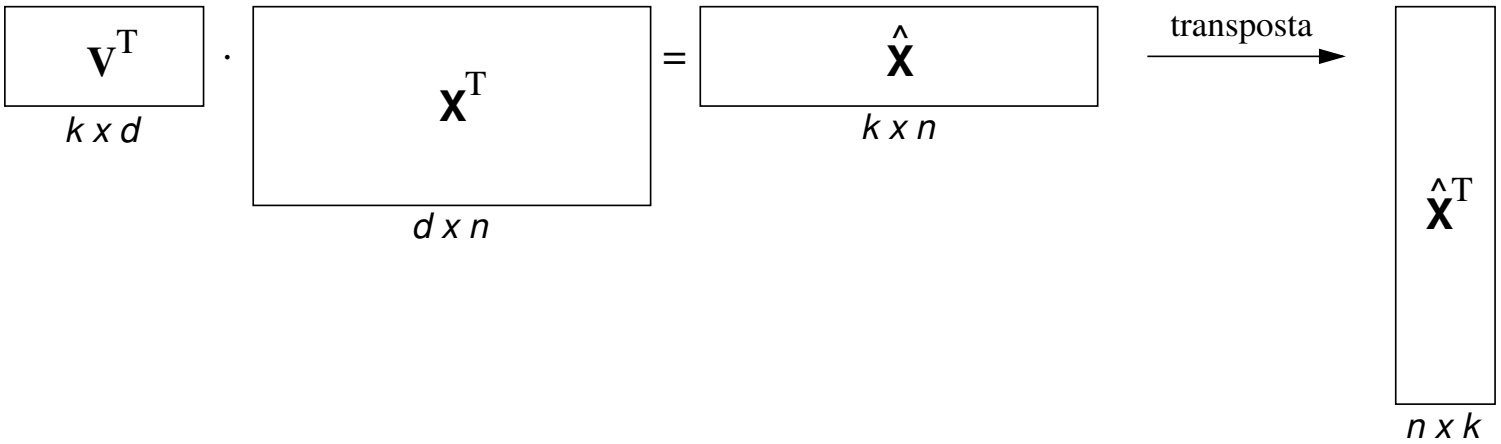
- Um autovetor estará associado a cada dimensão dos dados.
- Os autovalores são ordenados decrescentemente.
- Os autovetores com maiores autovalores são escolhidos para construir as novas características.
- Se desejarmos reduzir a dimensionalidade da base de dados `Iris` para 2, basta escolher os primeiros dois autovetores:

$$\begin{bmatrix} 0.3614 & 0.6566 \\ -0.0845 & 0.7302 \\ 0.8567 & -0.1734 \\ 0.3583 & -0.0755 \end{bmatrix}$$

Projeção dos Dados

- Após escolha das componentes principais, a nova base de dados (com menor dimensionalidade) pode ser construída pela multiplicação dos autovetores selecionados pela transposta da matriz original de dados.





Análise de Componentes Principais

- Exemplo:

Dados originais

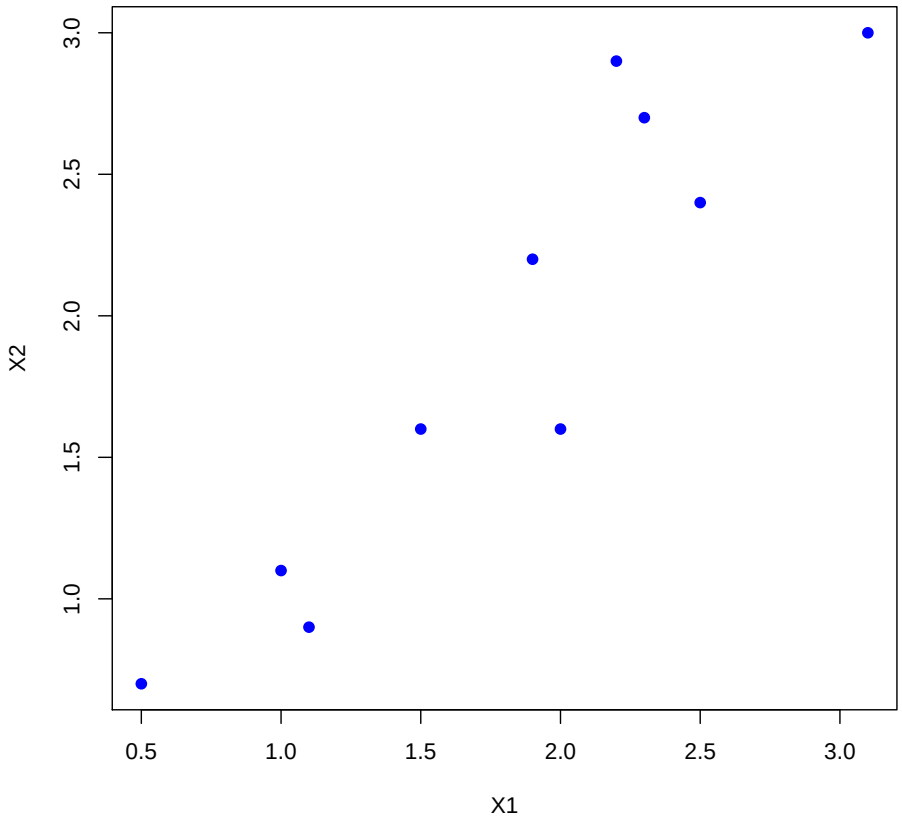
X_1	X_2
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2.0	1.6
1.0	1.1
1.5	1.6
1.1	0.9

Subtração da média

X_1'	X_2'
0.69	0.49
-1.31	-1.21
0.39	0.99
0.09	0.29
1.29	1.09
0.49	0.79
0.19	-0.31
-0.81	-0.81
-0.31	-0.31
-0.71	-1.01

Análise de Componentes Principais

- Gráfico dos dados originais:



Análise de Componentes Principais

- Matriz de covariância:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.6166 & 0.6154 \\ 0.6154 & 0.7166 \end{bmatrix}$$

- Autovalores e Autovetores:

$$\boldsymbol{U} = \begin{bmatrix} 0.0491 & 1.2840 \end{bmatrix}$$

$$\boldsymbol{V} = \begin{bmatrix} -0.7352 & -0.6779 \\ 0.6779 & -0.7352 \end{bmatrix}$$

Análise de Componentes Principais

- Projeção dos dados com duas componentes principais:

$\hat{X}_1$	$\hat{X}_2$
-0.8279	-0.1751
1.7775	0.1428
-0.9921	0.3843
-0.2742	0.1304
-1.6758	-0.2094
-0.9129	0.1752
0.0991	-0.3498
1.1445	0.0464
0.4380	0.0177
1.2238	-0.1626

Análise de Componentes Principais

- Transformação dos dados para o espaço original:

X_1	X_2
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2.0	1.6
1.0	1.1
1.5	1.6
1.1	0.9

Análise de Componentes Principais

- Projeção dos dados com uma componente principal:

$\hat{X}_1$
-0.8279
1.7775
-0.9921
-0.2742
-1.6758
-0.9129
0.0991
1.1445
0.4380
1.2238

Análise de Componentes Principais

- Transformação dos dados para o espaço original:

X_1	X_2
2.3713	2.4187
0.7050	0.6032
2.4826	2.5394
2.0959	2.1116
2.9460	3.0420
2.5289	2.5812
1.7428	1.7371
1.1341	1.0685
1.5131	1.4880
1.0804	1.0103