

Tera

AULA 30

Recommender Systems

Instrutor: [Raphael Ballet](#)

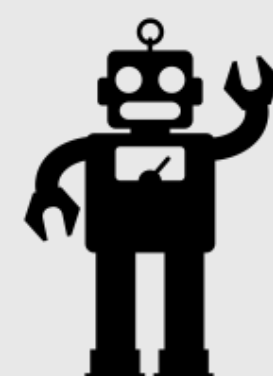
Background:

- Engenheiro de Controle e Automação (IMT)
- Mestre em Sistemas Aeroespaciais e Mecatrônica (ITA)
- Lead Data Scientist – Elo7

Interesses:



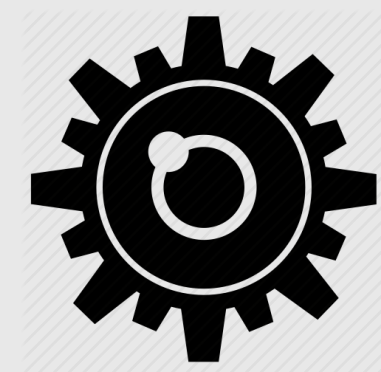
[Drones](#)



[Robótica](#)



[Aprendizado de
Máquina](#)



[Visão
Computacional](#)



[Processamento de
Linguagem Natural](#)

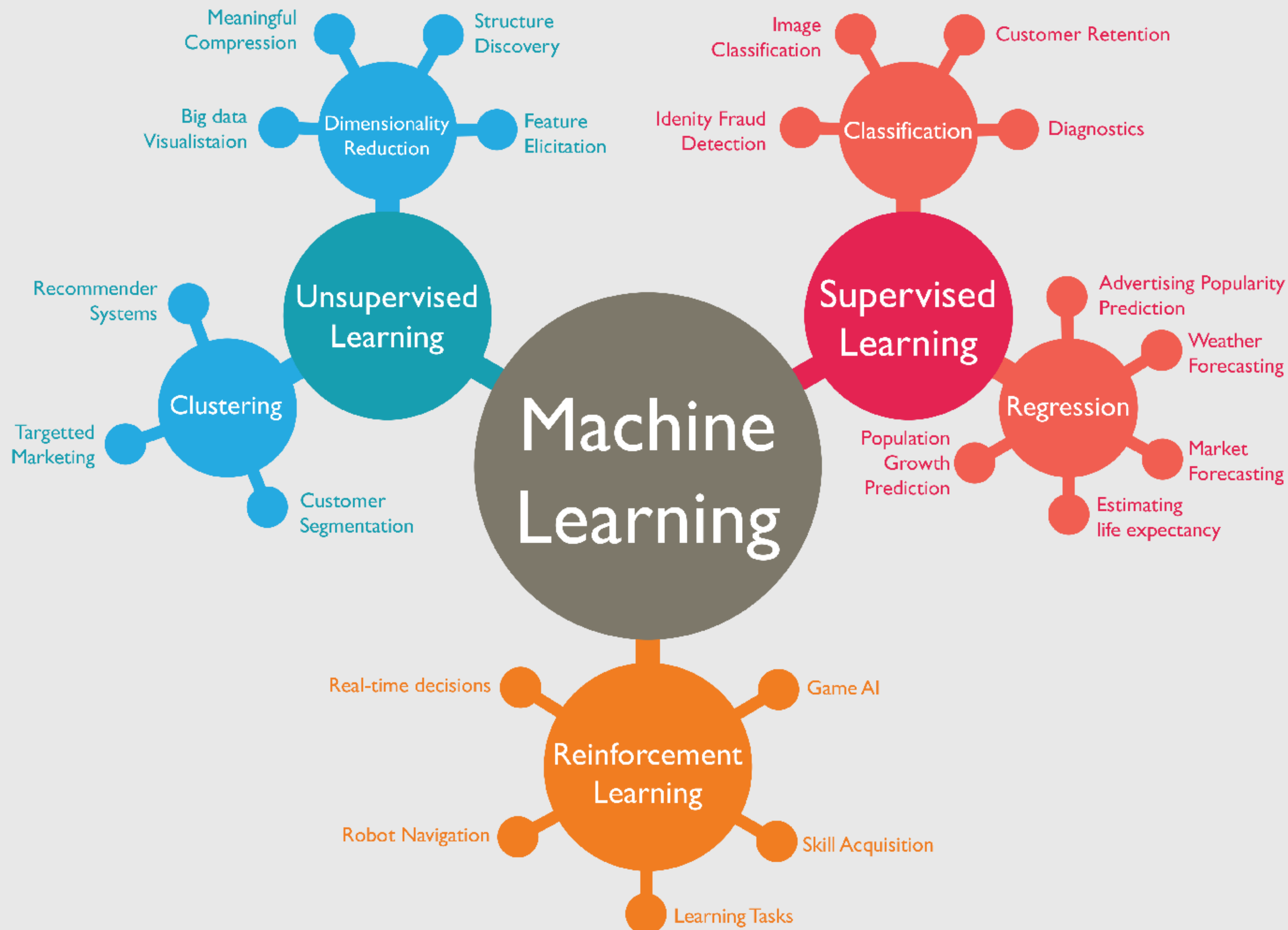


[Sistemas de
recomendação](#)

Planejamento:

1. Introdução
2. Sistemas de Recomendação: Conteúdo
3. Redução de Dimensionalidade
4. Topic Analysis
5. Filtro Colaborativo

1. INTRODUÇÃO









1. INTRODUÇÃO


- Por que recomendação?

1. INTRODUÇÃO



Produtos ▾ lembrancinha

¹  

 Raphael ▾

Categorias

[Acessórios](#)

[Aniversário e Festas](#)

[Bebê](#)

[Bijuterias](#)

[Bolsas e Carteiras](#)

[Casa](#)

[Casamento](#)

[Convites](#)

[Decoração](#)

[Doces](#)

[Eco](#)

[Infantil](#)

[Jogos e Brinquedos](#)

[Jóias](#)

[Lembrancinhas](#)

[Papel e Cia](#)

[Pets](#)

[Religiosos](#)

[Roupas](#)

[Saúde e Beleza](#)

[Técnicas de Artesanato](#)


Materiais para artesanato

Lembrancinha


1206310 PRODUTOS ENCONTRADOS

Buscas relacionadas: [lembrança](#) [lembrancinhas](#) [lembrancinha diferente](#) [lembrancinha personalizadas](#) [lembrancinha personalizada](#)


Preço até Cidade Filtrar por Ordenar por




LEMBRANCINHA URSINHO ARTICUL...
Atelier Aninha Ioannou
R\$ 15,00




LEMBRANCINHA CASAMENTO SABO...
Saboaria Natural Empório do Banho, desde 2...
R\$ 8,46




Lembrancinha batizado
Amô Atelier
R\$ 6,99




Tag para lembrancinha
My Party
R\$ 8,00 FRETE GRÁTIS




Lembrancinha Minions
Vila em Festa
R\$ 3,00



Bolo de frutas cristalizadas
Patricia Bertola Confeitaria
R\$ 30,00



Dedoche Turma da Galinha Pintadinha
Lilia B Covolo - Encontro dos Pontos
R\$ 38,00



Vaso com Rosas e Flores - Decoração
Origami Star
R\$ 60,00



1. INTRODUÇÃO

- Existe muita demanda e muita oferta, mas como unir os dois?



1. INTRODUÇÃO

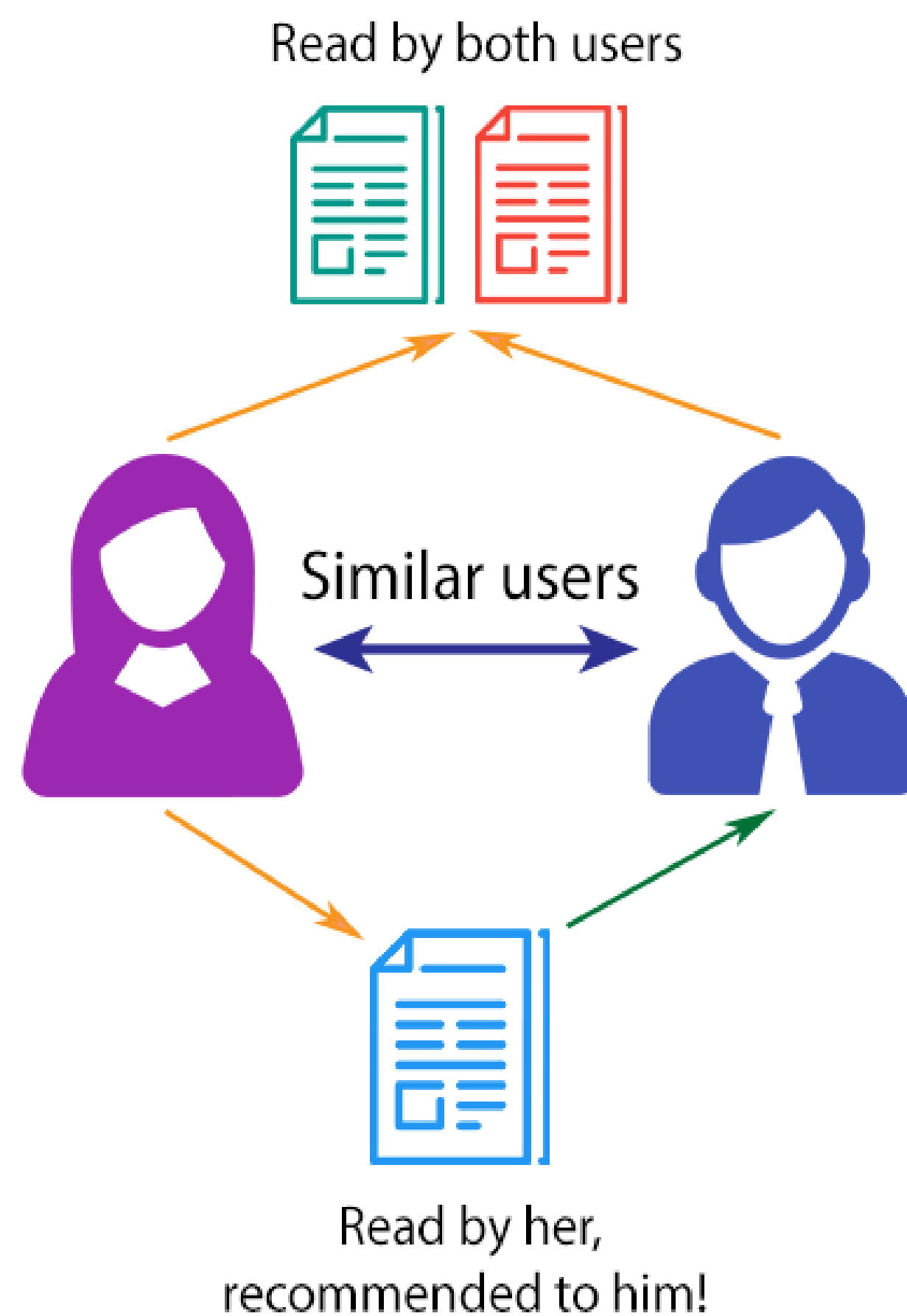
- Existe muita demanda e muita oferta, mas como unir os dois?
- Como recomendar produtos que nem mesmo o usuários sabia que queria?

2. Sistemas de Recomendação

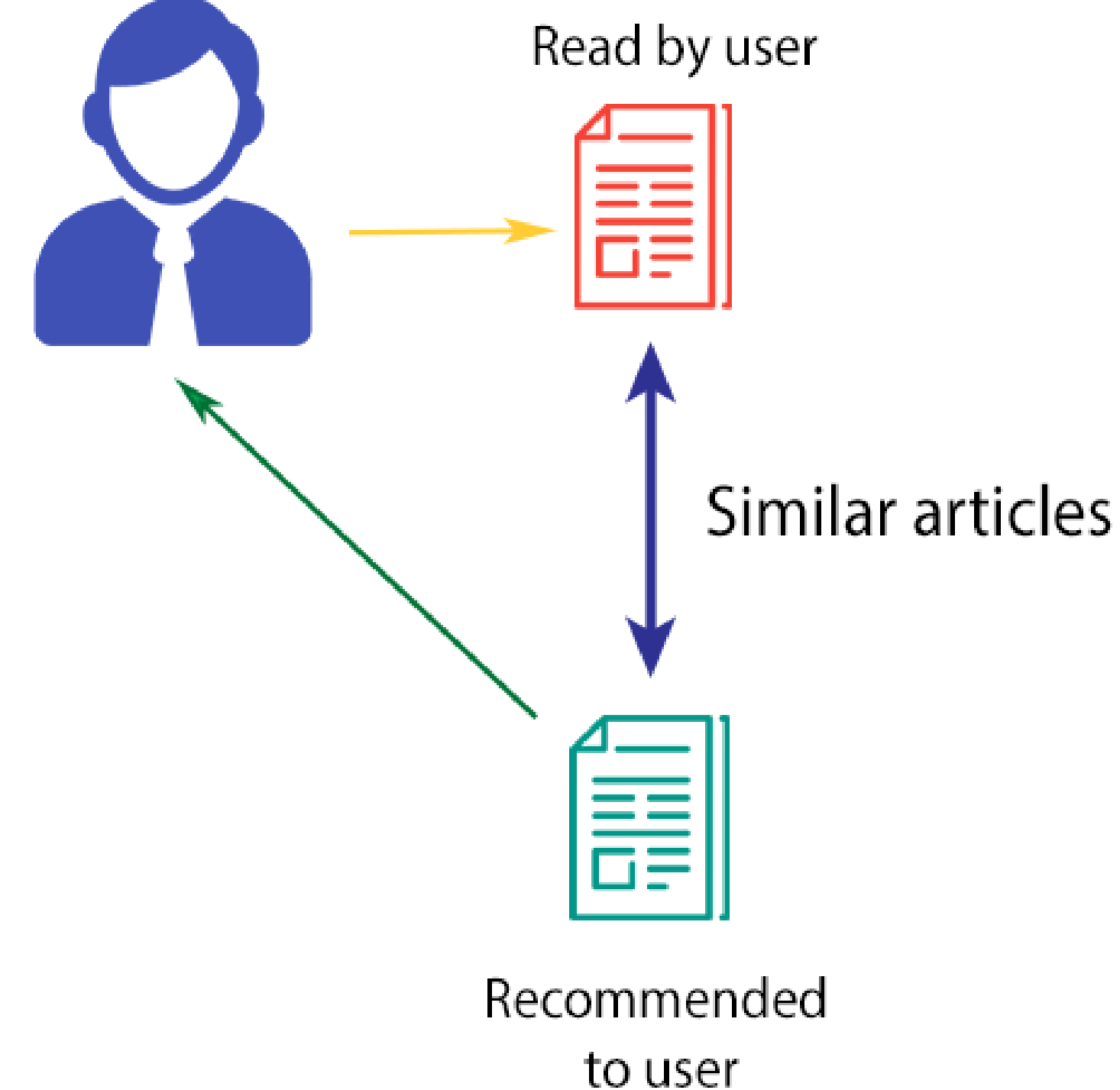
- Existem 3 grandes grupos:
 - Proximidade de documentos (produtos, músicas, filmes etc)
 - Proximidade entre usuários (filtro colaborativo)
 - Híbrido → Mistura dos outros 2

2. Sistemas de Recomendação

COLLABORATIVE FILTERING



CONTENT-BASED FILTERING



2. Sistemas de Recomendação

- Proximidade de documentos:
 - Distância entre documentos
 - Similaridade de temas (tópicos)

2. Sistemas de Recomendação

- Proximidade de documentos:
 - Distância entre documentos
 - Similaridade de temas (tópicos)



Clustering / Topic Analysis

2. Sistemas de Recomendação

- Exemplo – Recomendação artigos NY Times



leu / gostou

Recomendação




- Esporte
- Baseball
- Campeonato

- Esporte
- Baseball


2. Sistemas de Recomendação

- Case: Recomendação de produtos Elo7


Veja também estes produtos [>](#)




Ombreira de pérolas para vestidos
Bordado Studio
R\$ 49,00





Sabonete Lavandas de Provence
Solar difusores
Sob consulta



Prato para mini bolo "Coelhinho"
Amora criação de coisas
R\$ 75,00







2. Sistemas de Recomendação

- Problema: Vetores muito esparsos:
 - Distâncias semelhantes
 - Tempo de cálculo muito alto
 - Excessivo espaço em memória utilizado

2. Sistemas de Recomendação

- Solução:
 - Clustering
 - Redução de dimensionalidade
 - Análise de Tópicos

3. Redução de Dimensionalidade

- Dois objetivos principais:
 - Facilitar visualização e intuição
 - Amenizar o problema de similaridade entre observações

3. Redução de Dimensionalidade

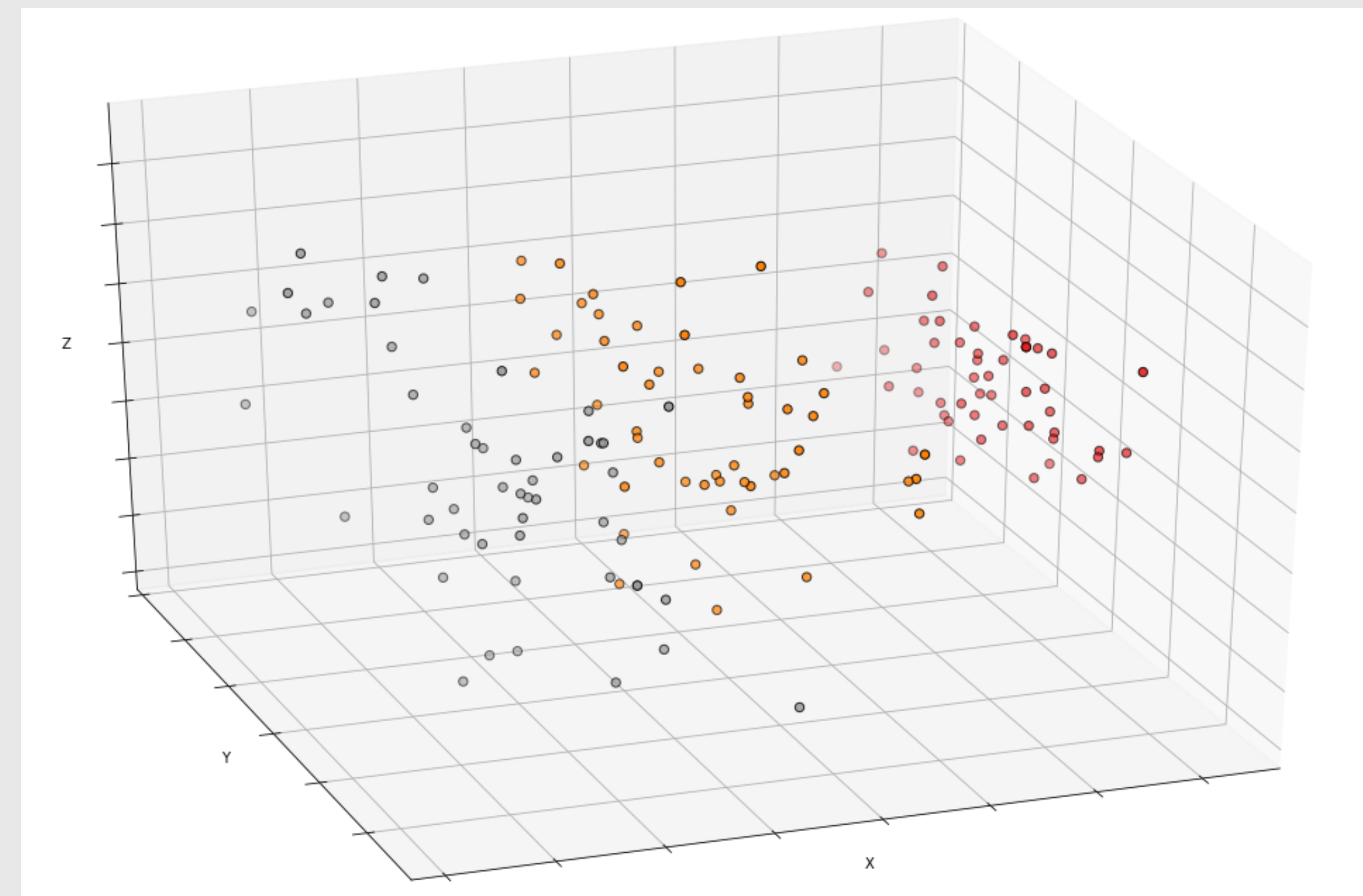
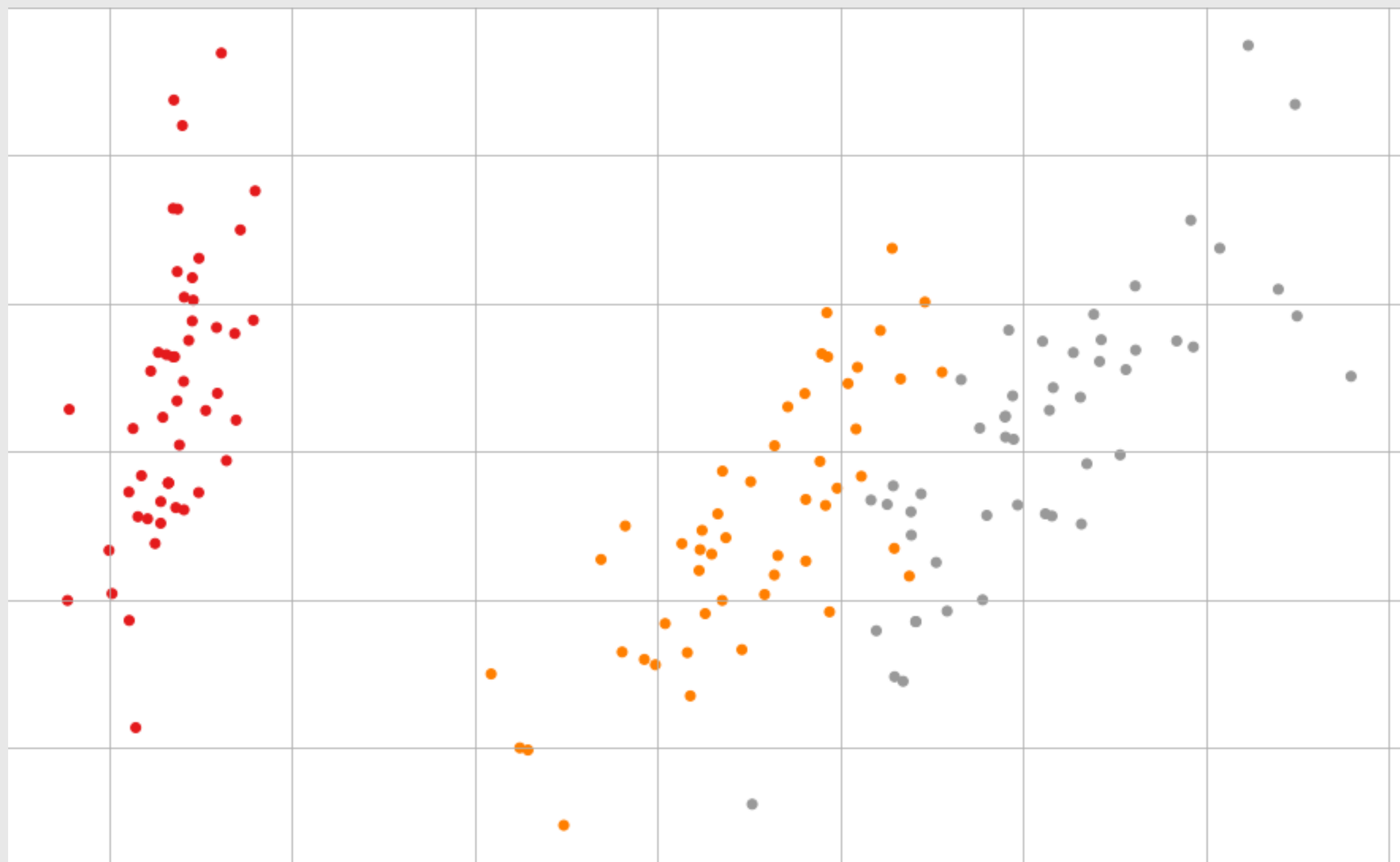
- Dois objetivos principais:
 - Facilitar visualização e intuição
 - Amenizar o problema de similaridade entre observações

Maldição da dimensionalidade



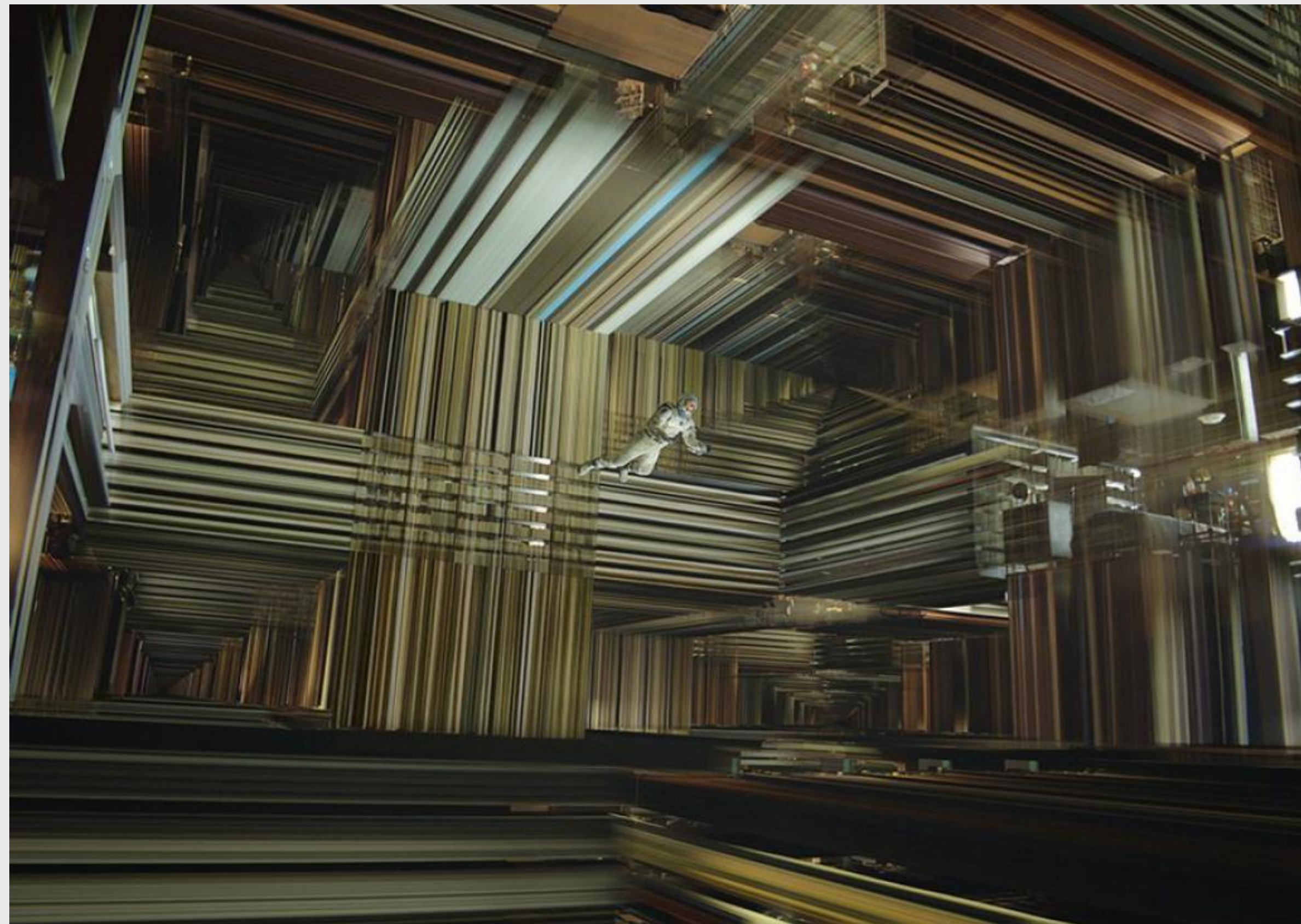
T 3. Redução de Dimensionalidade

- Visualização dos dados – 2D e 3D



T 3. Redução de Dimensionalidade

- Visualização mais do que 3D?



T 3. Redução de Dimensionalidade

- Visualização NLP (>1k dimensões)

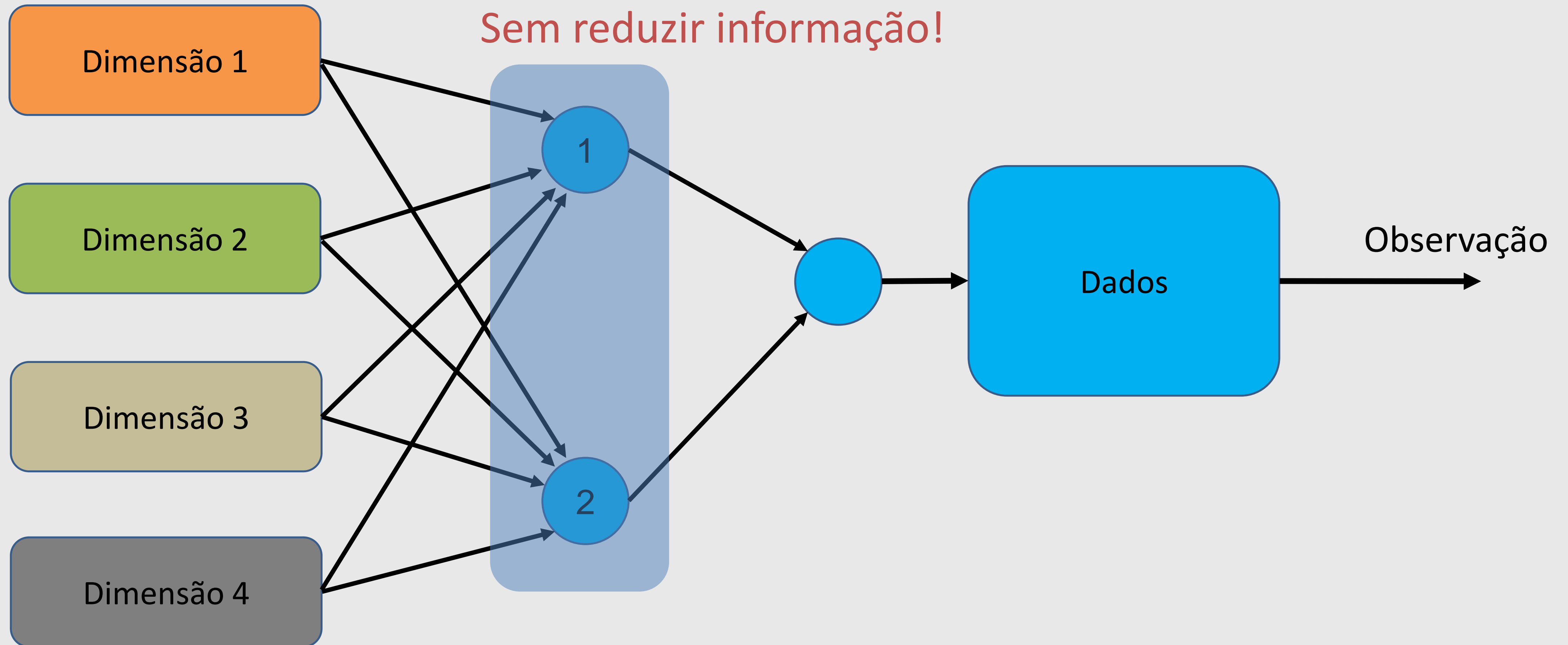


T 3. Redução de Dimensionalidade

- Objetivo

Sumário – Menor dimensão

Sem reduzir informação!



3. Redução de Dimensionalidade

- Técnicas principais:
 - **PCA** e SVD
 - Topic Analysis (**NMF** e LDA)

3. PCA

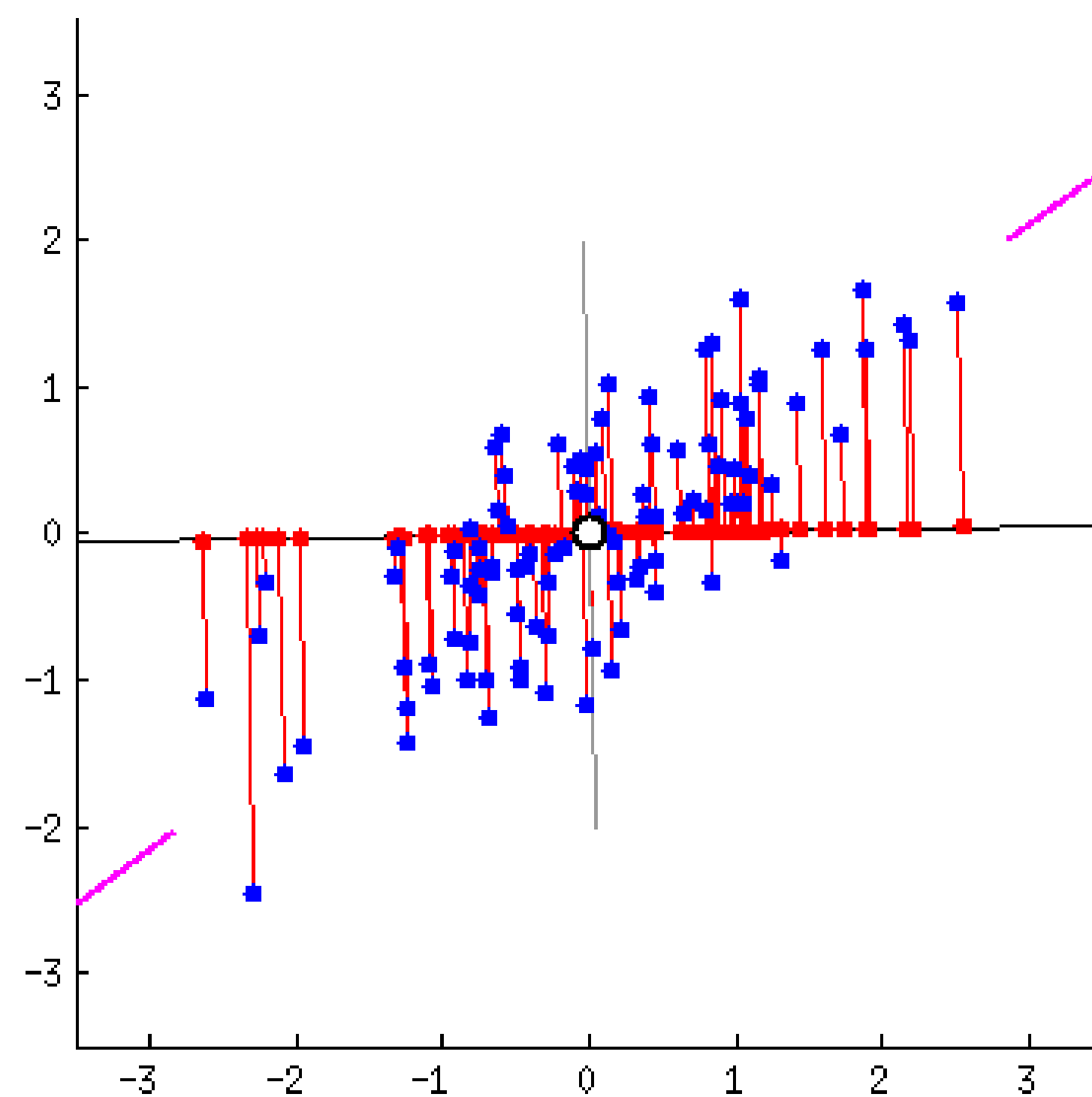
- **Principal Component Analysis:**
 - Encontra atributos de maior variação → “mais importantes”
 - Elimina atributos de menor variação → “menos explicativos”

3. PCA

- PCA: Componentes principais:
 - **Primeiro componente:**
 - Direção de maior variação nos dados
 - **Segundo componente:**
 - Direção da segunda maior variação e ortogonal ao primeiro
(descorrelacionado do primeiro)
 - ...

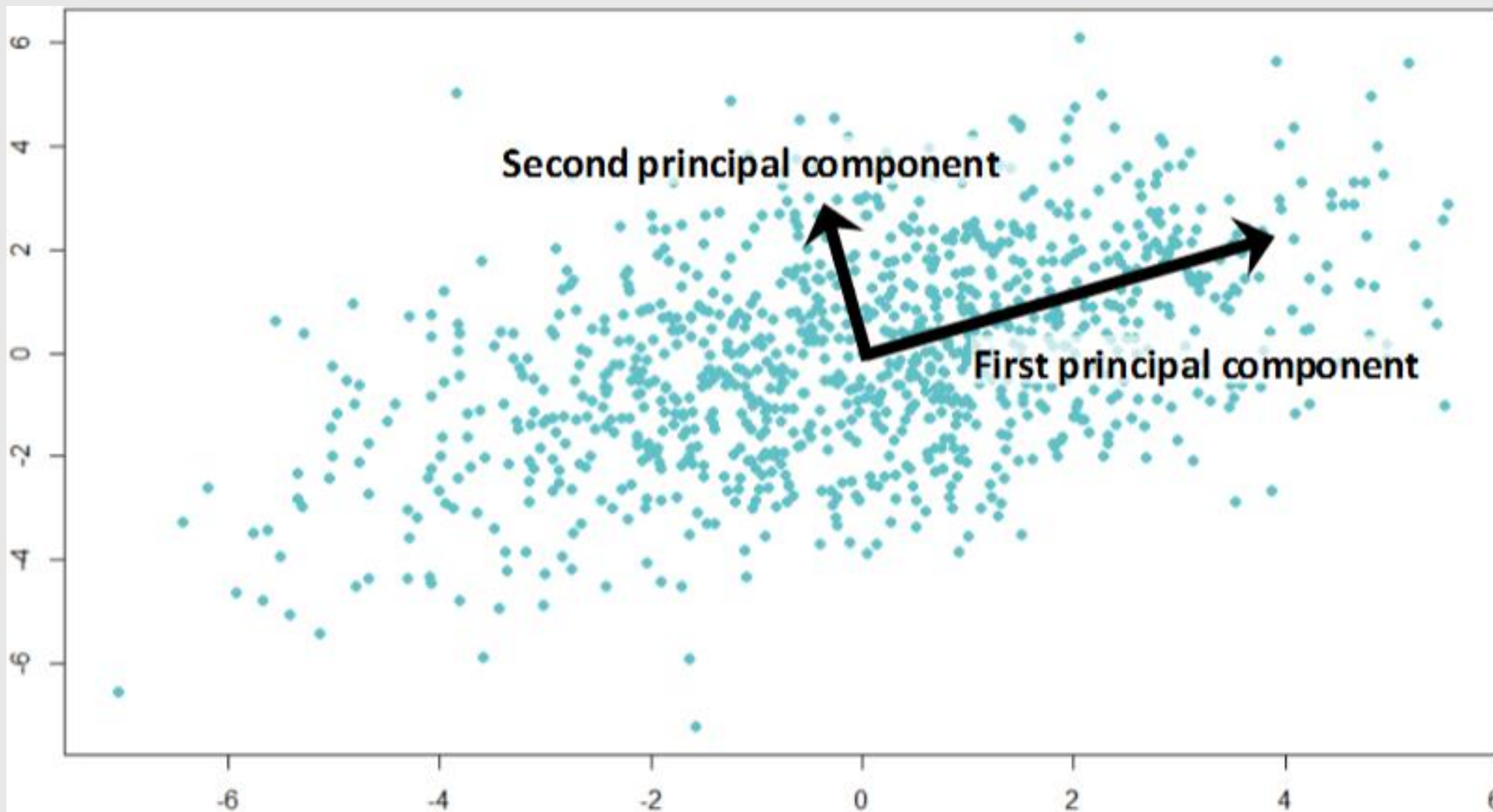
3. PCA

- Componentes principais:



3. PCA

- Componentes principais:

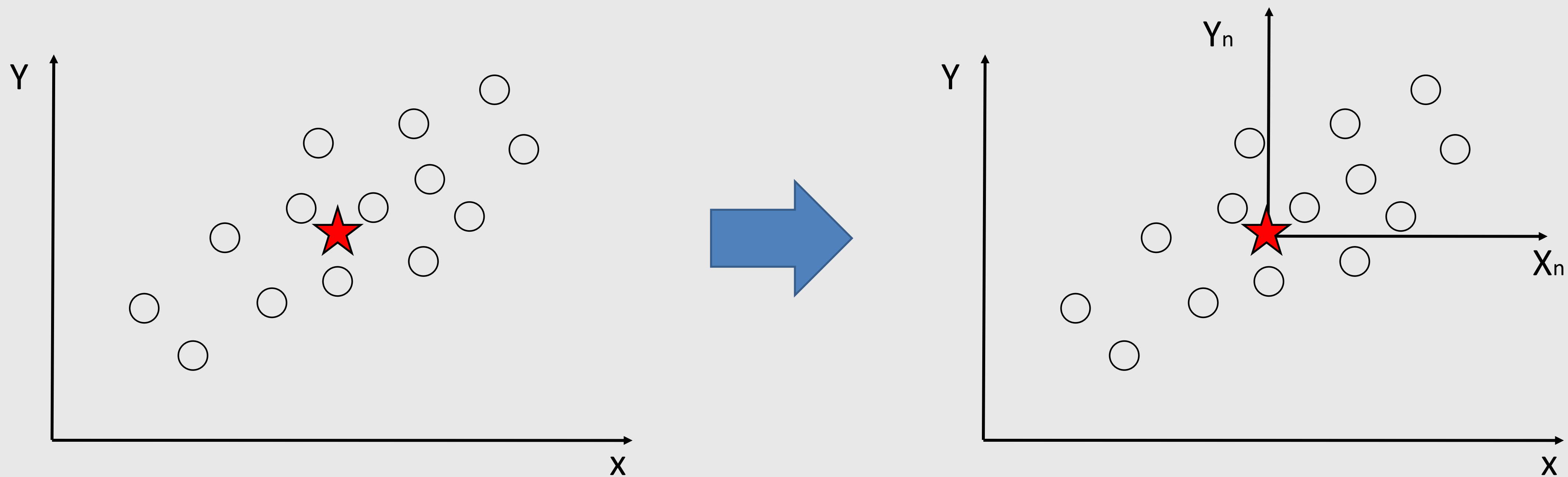


3. PCA

- Passos do algoritmo PCA:
 - 1) Remove média dos dados
 - 2) Rotaciona os eixos para descorrelacionar os atributos
 - 3) Ordena os componentes principais em nível de variância
 - 4) Remove os componentes menos variantes (Opcional)

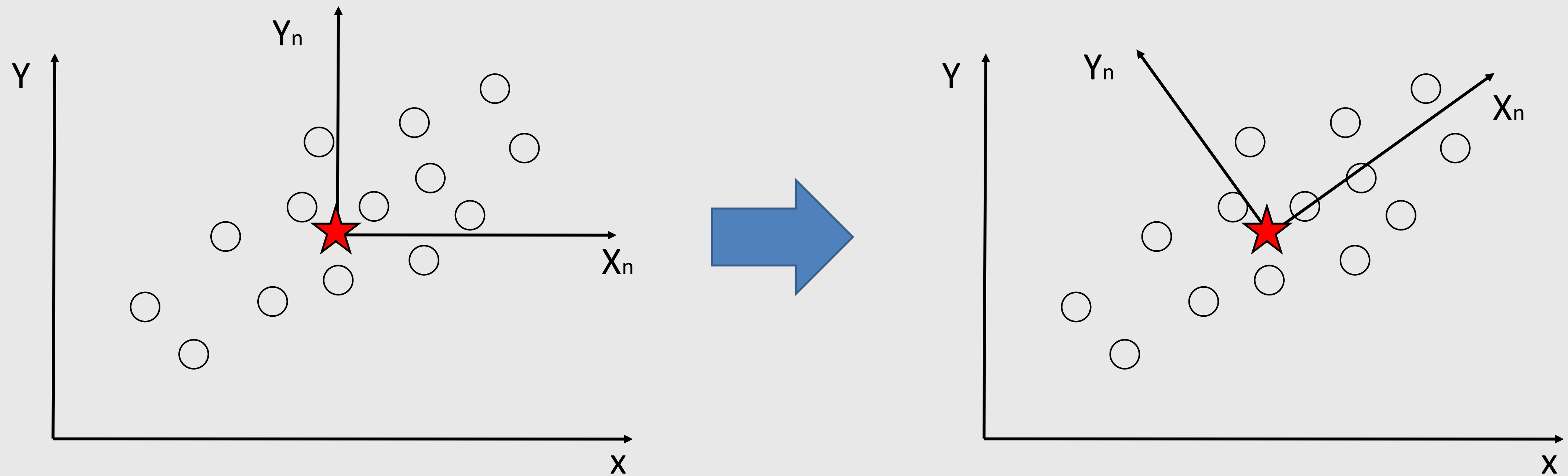
3. PCA

1) Remove média amostral dos dados



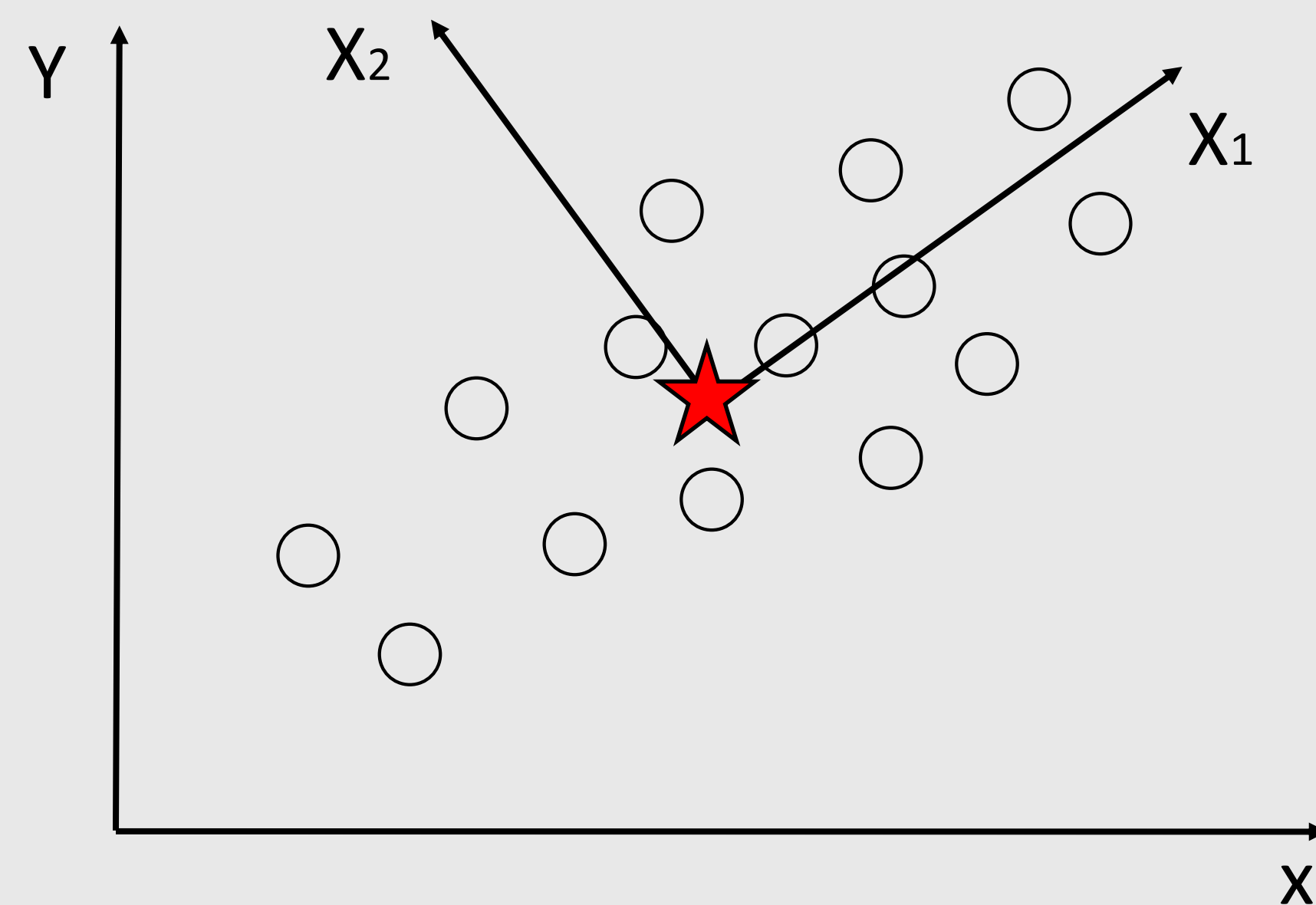
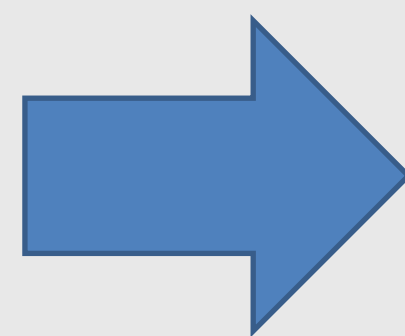
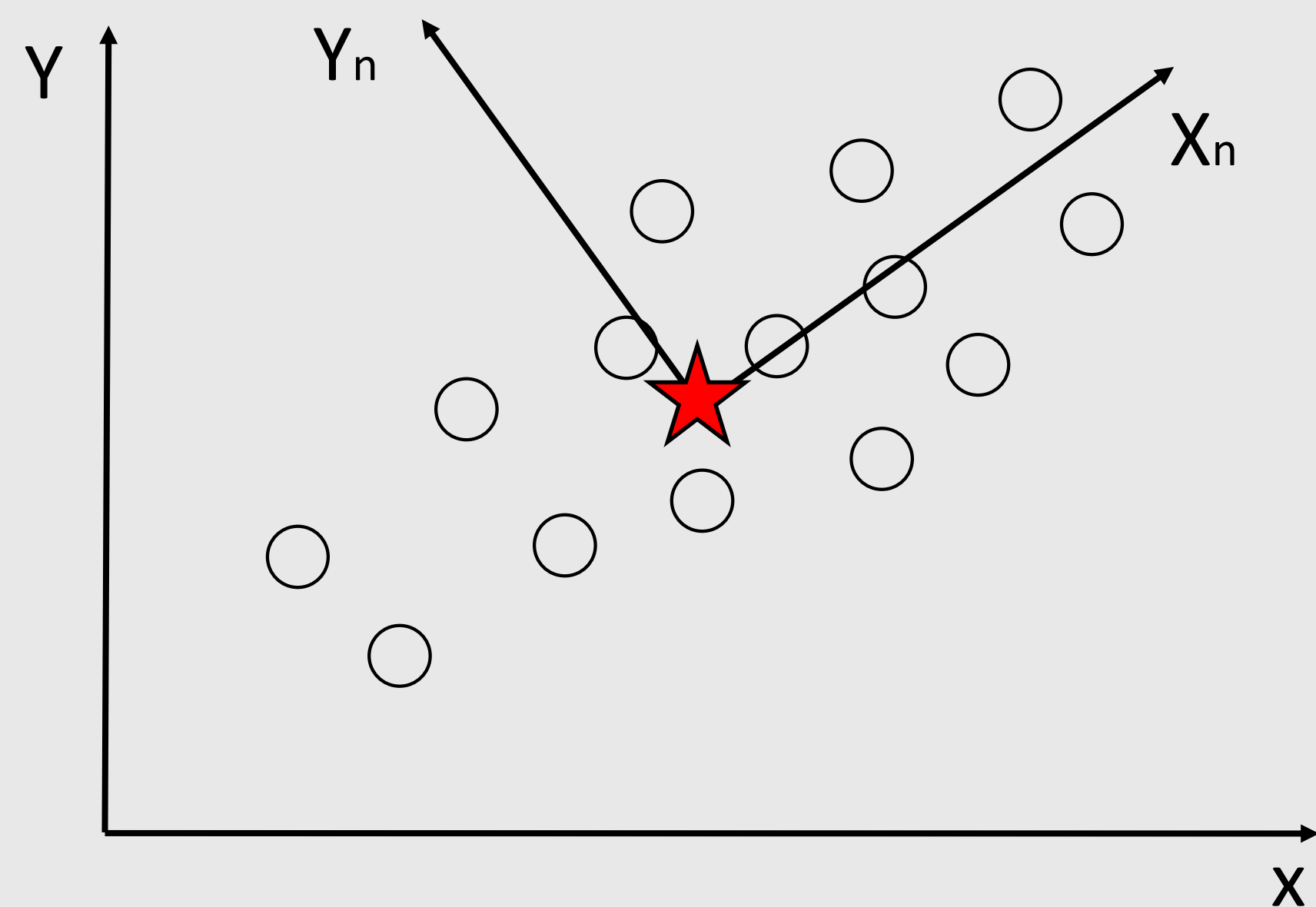
3. PCA

2) Rotaciona os eixos para descorrelacionar os atributos



3. PCA

3) Ordena os componentes principais em nível de variância

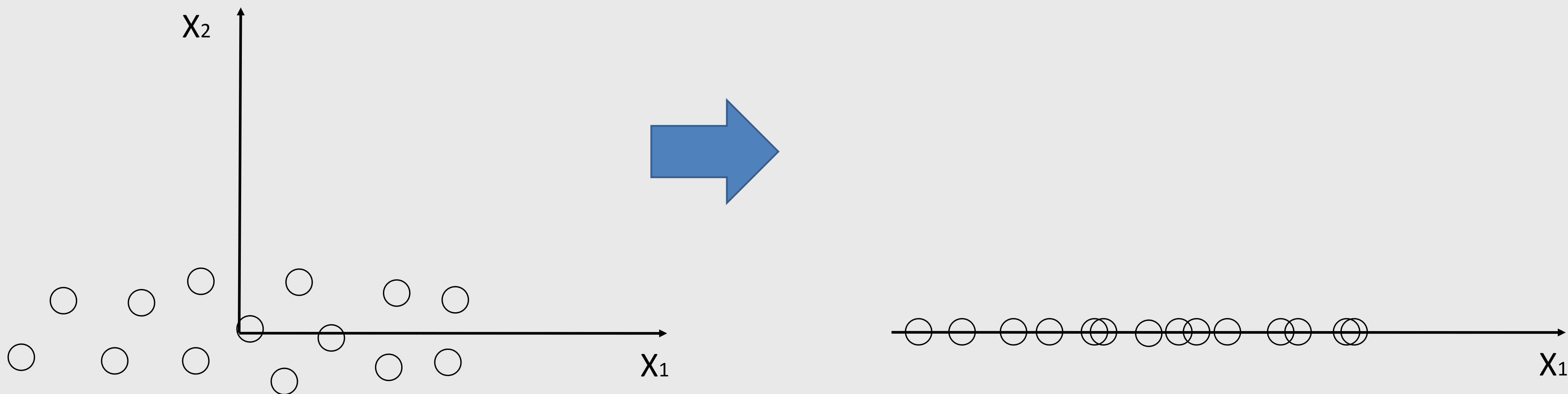


X_1 : Primeiro componente principal

X_2 : Segundo componente principal

3. PCA

4) Remove os componentes menos variantes (Opcional)

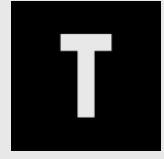


X_1 : Primeiro componente principal

X_2 : Segundo componente principal

3. PCA

- Quando remover componentes principais:
 - Atributos muito correlacionados
 - Componentes secundários pouco variantes
 - → baixa **variância explicada**
 - Balanço entre precisão e simplificação
 - **Encontrar dimensão intrínseca**



3. PCA

- Exemplo: notebook

3. PCA

- **Vantagens:**

- Permite reduzir dimensionalidade do problema sem perder informação
- Menor dimensionalidade
- → Maior velocidade e menos memória
- Resultados determinísticos

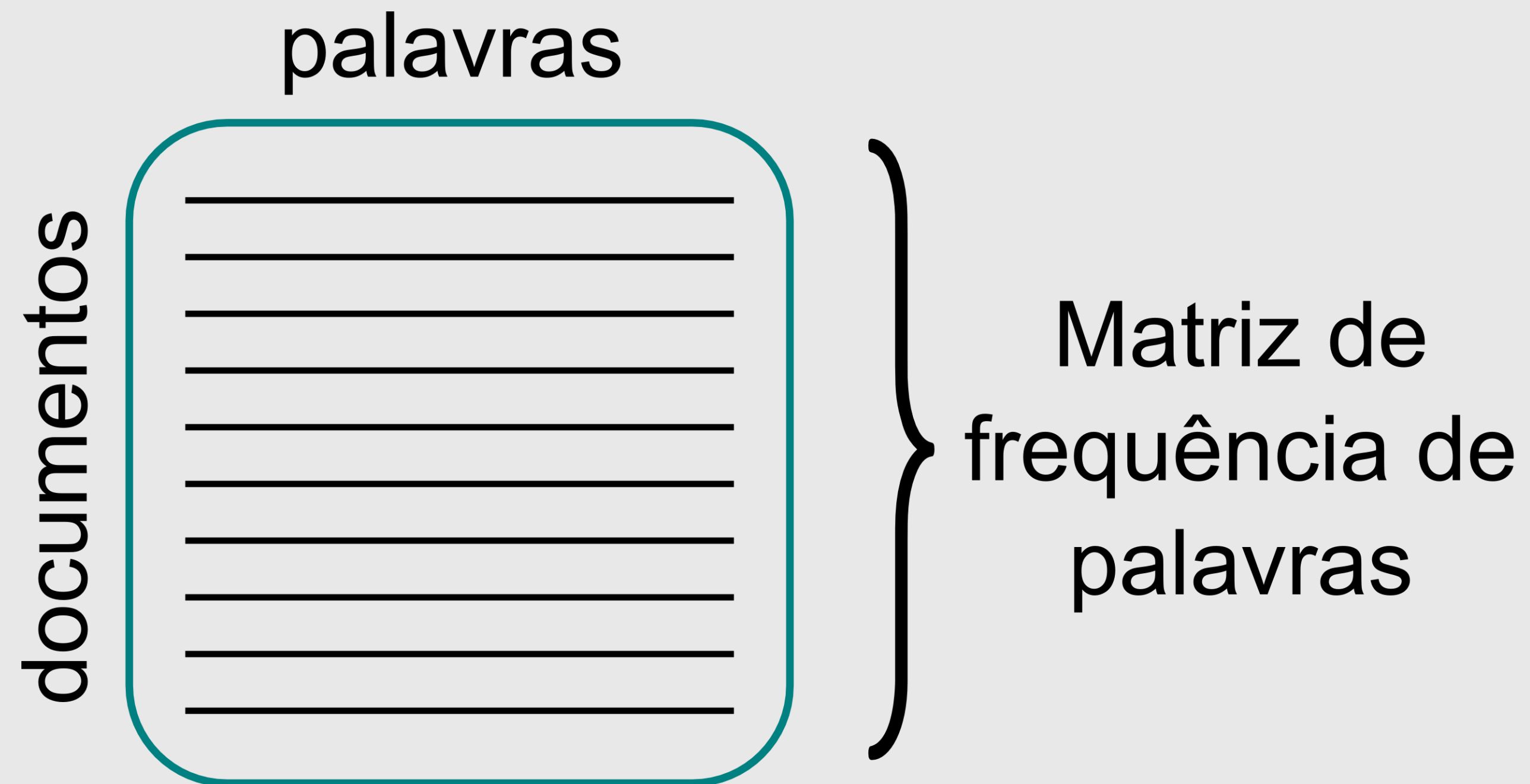
3. PCA

- **Desvantagens:**

- Dimensões resultantes (componentes principais) não representam os atributos
- Perde a “explicabilidade” do algoritmo
- Má escolha de número de componentes pode prejudicar análise

4. Topic Analysis

- Problema já conhecido:
 - Documentos + Palavras = Muitas dimensões



4. Topic Analysis

- **Solução:**
 - Redução de dimensionalidade
 - Clustering
 - ...




4. Topic Analysis

- **Solução:**
 - Redução de dimensionalidade
 - Clustering
- **Problema:**
 - Perda de interpretabilidade dos dados

4. Topic Analysis

- Objetivo:
 - Encontrar estrutura implícita nos documentos – Tópicos / Temas



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia

Toolbox
What links here
Related changes
Upload file
Special pages
Permanent link
Cite this page
Expand citations
DYK check

Print/export
Create a book
Download as PDF
Printable version

Languages
العربية
Български
Català

Daniel MietchenMy talkMy sandboxMy preferencesMy watchlistMy contributionsLog out00:41:29

ArticleTalkReadEditView historySearch

Dengue fever

A featured article from Wikipedia, the free encyclopedia

For other uses, see *Dengue fever (disambiguation)*.

Dengue fever (UK /dɛnɡɛt/ or US /dɛnɡiː/), also known as **breakbone fever**, is an infectious tropical disease caused by the dengue virus. Symptoms include fever, headache, muscle and joint pains, and a characteristic skin rash that is similar to measles. In a small proportion of cases the disease develops into the life-threatening **dengue hemorrhagic fever**, resulting in bleeding, low levels of blood platelets and blood plasma leakage, or into **dengue shock syndrome**, where dangerously low blood pressure occurs.

Dengue is transmitted by several species of mosquito within the genus *Aedes*, principally *A. aegypti*. The virus has four different types; infection with one type usually gives lifelong immunity to that type, but only short-term immunity to the others. Subsequent infection with a different type increases the risk of severe complications. As there is no vaccine, prevention is sought by reducing the habitat and the number of mosquitoes and limiting exposure to bites.

Treatment of acute dengue is supportive, using either oral or intravenous rehydration for mild or moderate disease, and intravenous fluids and blood transfusion for more severe cases. The incidence of dengue fever has increased dramatically since the 1960s, with around 50–100 million people infected yearly. Early descriptions of the condition date from 1779, and its viral cause and the transmission were elucidated in the early 20th century. Dengue has become a global problem since the Second World War and is endemic in more than 110 countries. Apart from eliminating the mosquitoes, work is ongoing on a vaccine, as well as medication targeted directly at the virus.

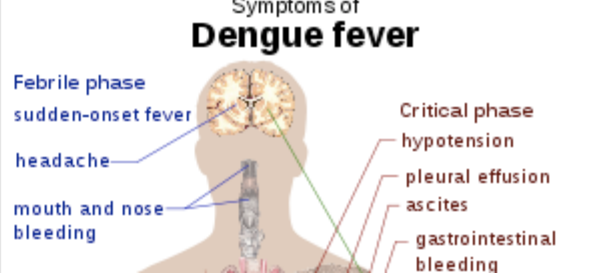
Contents [show]

Signs and symptoms

[edit]


Typically, people infected with dengue virus are asymptomatic (80%) or only have mild symptoms such as an uncomplicated fever.^{[1][2][3]} Others have more severe illness (5%), and in a small proportion it is life-threatening.^{[1][3]} The incubation period (time between exposure and onset of symptoms) ranges from 3–14 days, but most often it is 4–7 days.^[4] Therefore, travelers returning from endemic areas are unlikely to have dengue if fever or other symptoms start more than 14 days after arriving home.^[5] Children often experience symptoms similar to those of the common cold and gastroenteritis (vomiting and diarrhea),^[6] and generally have less severe symptoms than adults,^[7] but are more susceptible to the severe complications.^[5]

Symptoms of Dengue fever



Dengue fever

Classification and external resources



The typical rash seen in dengue fever

ICD-10	A90 ↗
ICD-9	061 ↗
DiseasesDB	3564 ↗
MedlinePlus	001374 ↗
eMedicine	med/528 ↗
MeSH	C02.782.417.214 ↗

75% Health

20% Medicine

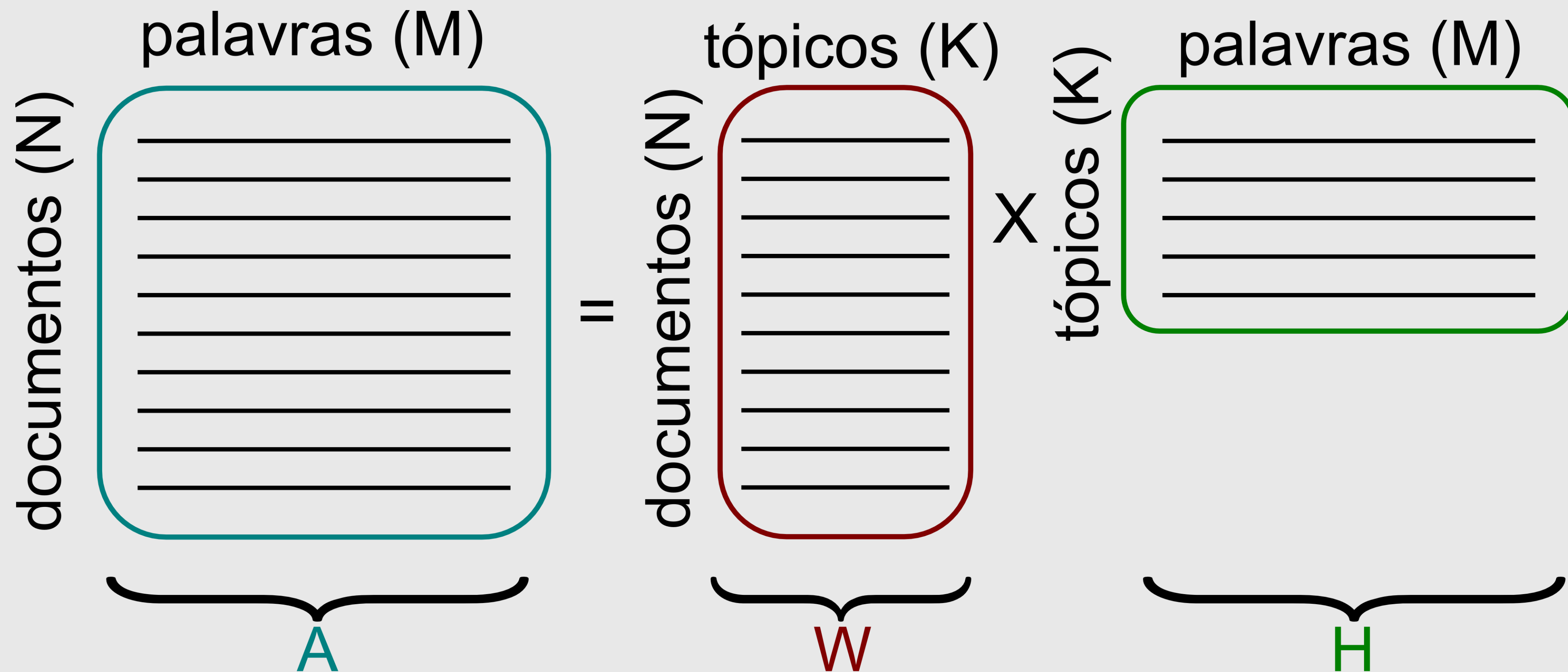
5% Biology

4. Topic Analysis – NMF

- **Non-Negative Matrix Factorization (NMF)**
- Principal objetivo:
 - Decompor a matriz de frequência de palavras em representações de tópicos
 - Documentos são compostos de combinações de tópicos
 - Tópicos são compostos de combinações de palavras

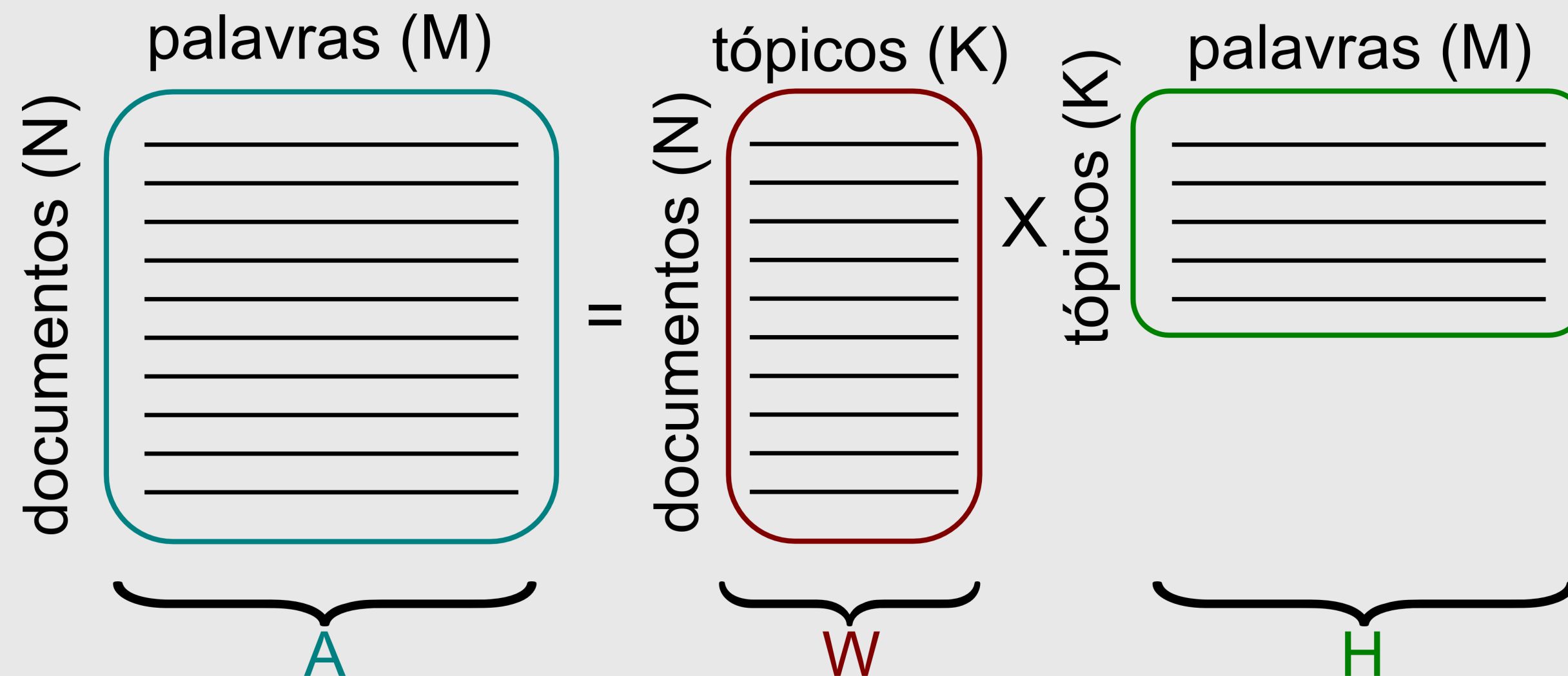
4. Topic Analysis – NMF

- NMF: Fatoração $\rightarrow A = WH$



4. Topic Analysis – NMF

- Matrizes:
 - A: Matriz de frequência de termos (M) em documentos (N)
 - W: Matriz de pesos → distribuição de tópicos (K) nos documentos
 - H: Matriz de atributos → distribuição de palavras nos tópicos



4. Topic Analysis – NMF

- Principais características:
 - Precisa definir o número de tópicos
 - Matrizes A , W e H não podem ter valores negativos
 - Matrizes W e H podem reconstruir matriz A (aprox.)

4. Topic Analysis – NMF

- NMF pode ser utilizado em vários outros cenários:
 - **Segmentação de fontes sonoras do áudio:**
 - Documentos: áudio
 - Features: espectograma do áudio
 - **Segmentação de imagens:**
 - Documentos: imagem
 - Features: pixels

T

4. Topic Analysis – NMF

- Exemplo: notebook

4. Topic Analysis – NMF

- **Vantagens:**

- Tópicos são interpretáveis
- Naturalmente agregador (clustering)
- Pode ser utilizado em outros contextos (ex: imagens, áudio etc)

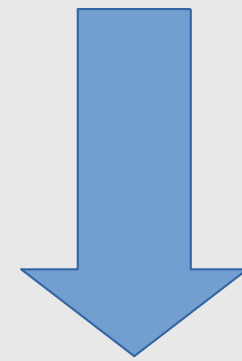
4. Topic Analysis – NMF

- **Desvantagens:**

- Solução aproximada
- Pode causar overfitting
- Limitação de utilizar apenas features positivas

5. Filtro Colaborativo

- Proximidade entre usuários (filtro colaborativo):
 - Usuários semelhantes consomem documentos semelhantes



**Clustering Topic
Analysis**

5. Filtro Colaborativo

- Proximidade entre usuários (filtro colaborativo):
 - **Documentos:**
 - Histórico de consumo do usuário (compra, avaliação etc)
 - **Atributos / Features:**
 - Lista de itens de consumo (produtos, livros, filmes etc)


5. Filtro Colaborativo

- Exemplo: Recomendação de filmes

Close

Other Movies You Might Enjoy

[Amelie](#)

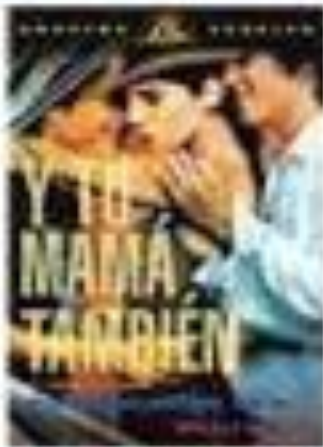


Add

★★★★☆

☐ Not Interested

[Y Tu Mama Tambien](#)




Add

★★★★☆

☐ Not Interested

[Guys and Balls](#)




Add

★★★★☆

☐ Not Interested

[Mostly Martha](#)

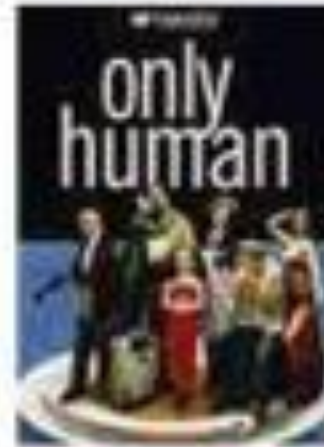


Add

★★★★☆

☐ Not Interested

[Only Human](#)




Add

★★★★☆

☐ Not Interested


[Russian Dolls](#)



Add

★★★★☆

☐ Not Interested



Eiken has been added to your Queue at position 2.

This movie is available now.

Move To Top Of My Queue

[< Continue Browsing](#) [Visit your Queue >](#)

Close

5. Filtro Colaborativo

- Exemplo: Recomendação de filmes



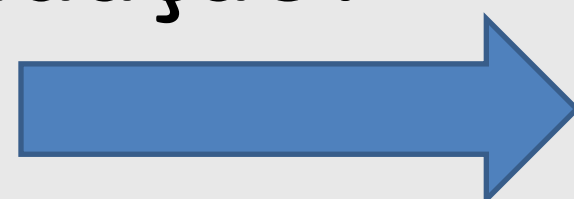
A diagram illustrating collaborative filtering for movie recommendations. It features a grid of user ratings for six movies. The movies are represented by posters at the top: Sherlock, House of Cards, Avengers, Armco Development, Breaking Bad, and Walking Dead. The users are represented by avatars on the left. The grid shows ratings from 1 to 5, with empty cells indicating no rating.

	SHERLOCK	HOUSE OF CARDS	AVENGERS	ARMCO DEVELOPMENT	Breaking Bad	WALKING DEAD
User 1	2			4	5	
User 2	5		4			1
User 3			5		2	
User 4		1		5		4
User 5			4			2
User 6	4	5		1		

5. Filtro Colaborativo

- Exemplo: Recomendação de filmes

Recomendação!



	2			4	5	
	5		4			1
			5		2	
		1		5		4
			4			2
	4	5		1		

5. Filtro Colaborativo

- **Recomendação:**
 - Encontrar items / usuários semelhantes
→ Menor distância entre vetores
- **Problema:**
 - Vetores muito esparsos: muitas dimensões sem valores

6. Filtro Colaborativo

- Exemplo: Recomendação produtos Elo7

Histórico de compras

	prod0	prod1	prod2	prod3	...
user0	0	1	0	0	...
user1	0	0	1	0	...
user2	1	0	0	0	...

➡ ~8 milhões de produtos

↓
30 milhões de acessos por mês!

6. Filtro Colaborativo

- Diferentes métodos:

- **Memória:**

a) Item-Item: “Quem comprou isso também comprou ...”



b) User-Item: “Usuários semelhantes a você compraram ...”

6. Filtro Colaborativo

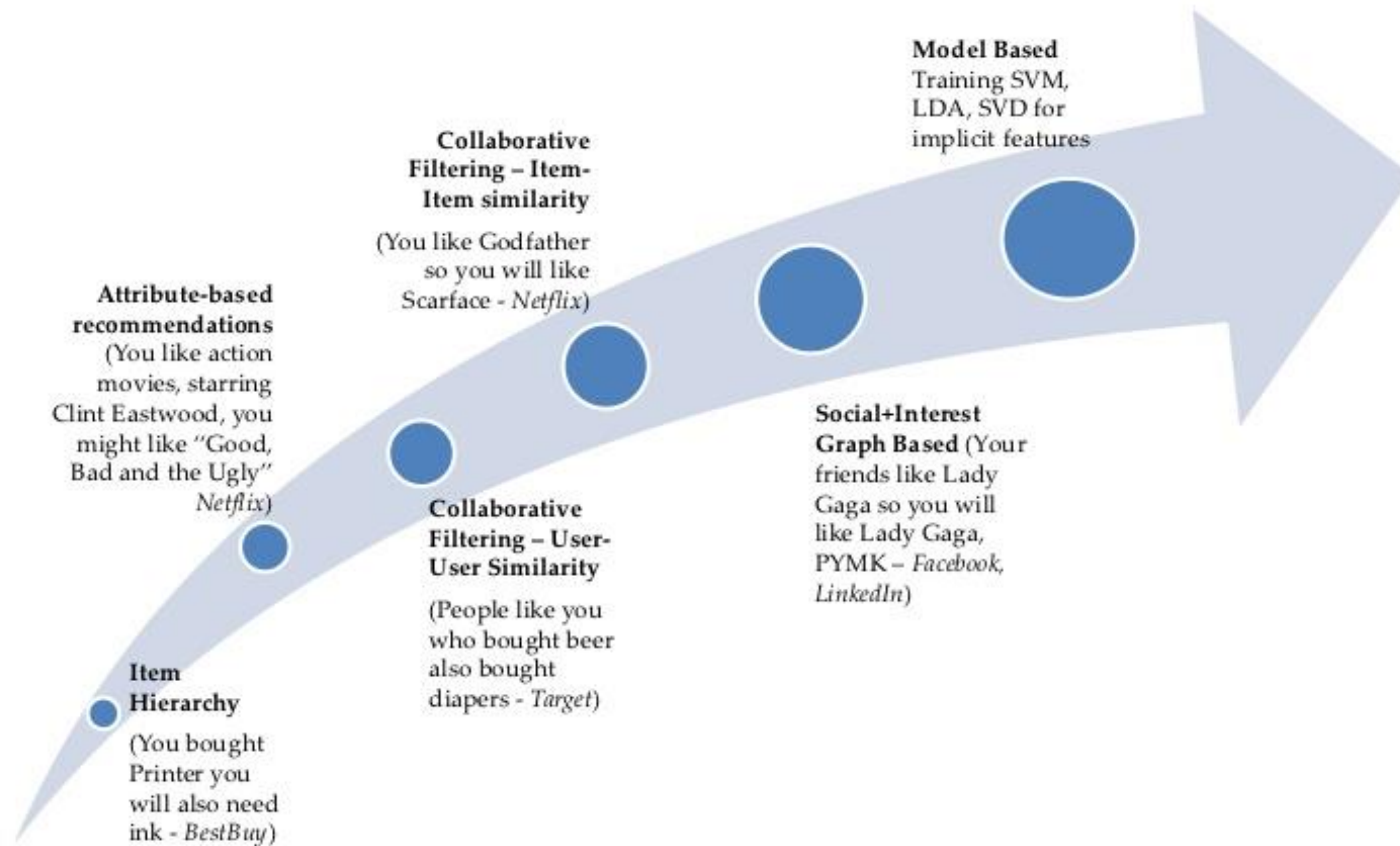
- Diferentes métodos:

- **Modelo:**

- a) Fatorização de Matrizes: SVD, PCA, NMF etc

6. Filtro Colaborativo

Recommender Approaches








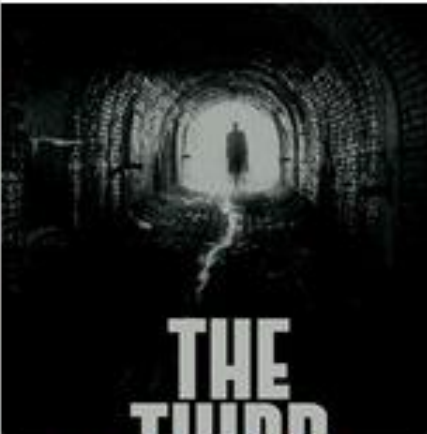

6. Filtro Colaborativo

- Case MovieLens:

top picks

[see more](#)








based on your ratings, MovieLens recommends these movies

Band of Brothers 2001 [R] 705 min 𐀀	Casablanca 1942 [PG] 102 min 𐀀	One Flew Over the Cuckoo's Nest 1975 [R] 133 min 𐀀	The Lives of Others 2006 [R] 137 min 𐀀	Sunset Boulevard 1950 [NR] 110 min 𐀀	The Third Man 1949 [NR] 104 min 𐀀	Pat 1957
						
★★★★★	★★★★★	★★★★★	★★★★★	★★★★★	★★★★★	★★★★★

recent releases

[see more](#)

movies released in last 90 days that you haven't rated

Cantinflas 2014 [PG] 106 min 𐀀	Felony 2014 𐀀	What If 2014 [PG-13] 102 min 𐀀	Frank 2014 [R] 96 min 𐀀	Sin City: A Dame to 2014 [R] 102 min 𐀀	If I Stay 2014 [PG-13] 106 min 𐀀	Are 2014
						

T

OBRIGADO!