# Aprendizagem automática

## Sessão 6 - T

## Aprendizagem não supervisionada
## Redução da dimensionalidade

**Ciência de Dados Aplicada**

**2023/2024**

# Redução da dimensionalidade

- **Objetivo:** transformar dados de elevada dimensão numa representação de dimensão inferior, com o objetivo de captar os padrões e relações essenciais dos dados, minimizando a redundância e o ruído.

- **Porquê reduzir as dimensões?**
  - Simplifica **a análise e a visualização** de dados complexos;
  - Reduz a **complexidade computacional** e os requisitos de memória;
  - Ajuda a atenuar a **maldição da dimensionalidade**;
  - Melhora **o desempenho e a generalização do modelo**;
  - Melhora **a interpretabilidade** e a compreensão dos padrões de dados subjacentes;

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

# Redução da dimensionalidade

- Se tivermos apenas uma caraterística, podemos facilmente traçar os dados numa reta numérica.
- Mesmo com este gráfico simples, podemos ver diferenças entre as amostras 1, 2 e 3 e 4, 5 e 6.

| | Característica 1 |
|---|---|
| **Amostra 1** | 10 |
| **Amostra 2** | 11 |
| **Amostra 3** | 8 |
| **Amostra 4** | 3 |
| **Amostra 5** | 1 |
| **Amostra 6** | 2 |


**Característica 1**

# Redução da dimensionalidade

- Agora podemos traçar os dados num gráfico bidimensional.

| | Característica 1 | Característica 2 |
|---|---|---|
| Amostra 1 | 10 | 6 |
| Amostra 2 | 11 | 4 |
| Amostra 3 | 8 | 5 |
| Amostra 4 | 3 | 3 |
| Amostra 5 | 1 | 2.8 |
| Amostra 6 | 2 | 1 |

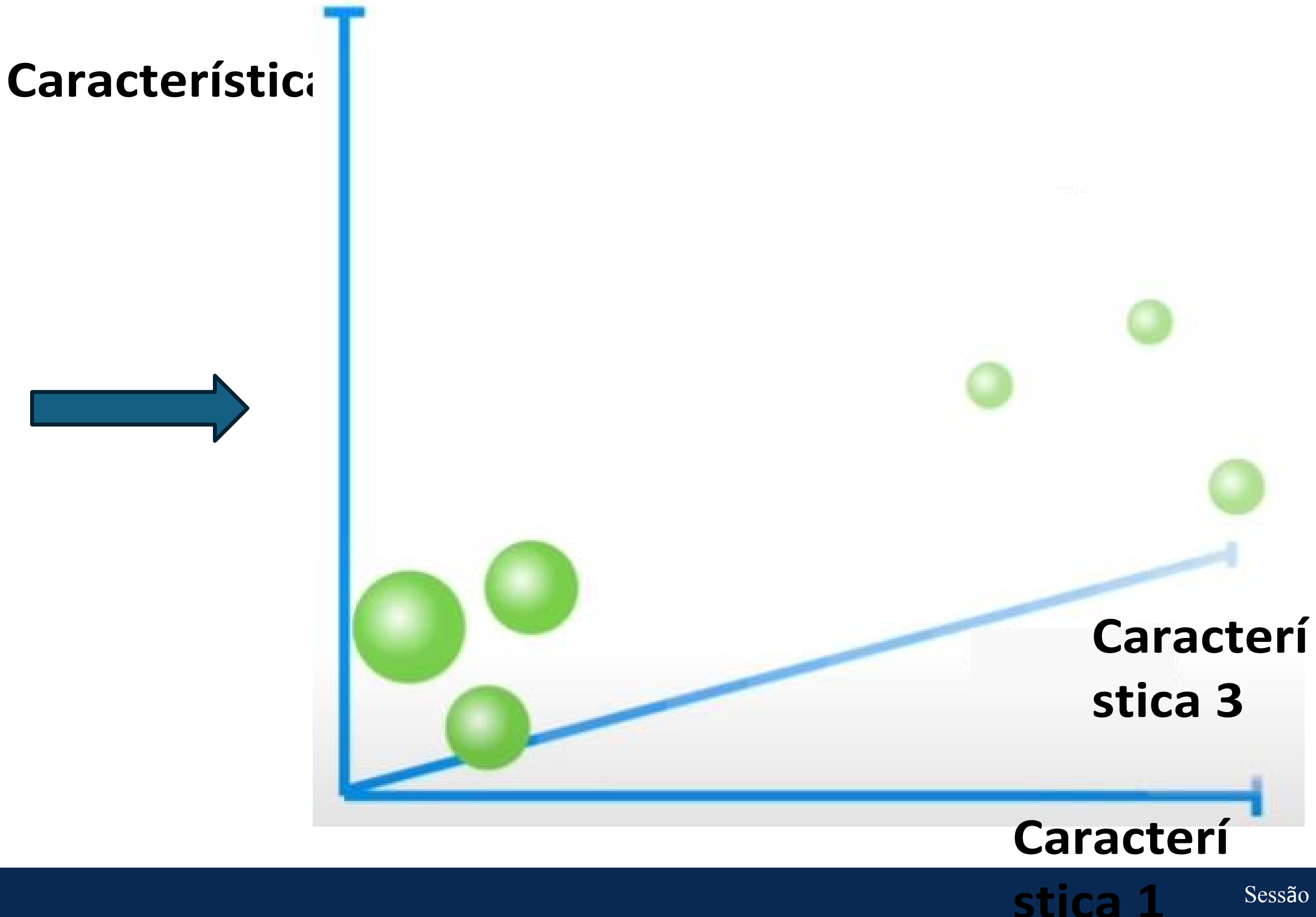
# Redução da dimensionalidade

**Característica 1**

# Redução da dimensionalidade

- Agora podemos traçar os dados num gráfico tridimensional.

| | Característica 1 | Característica 2 | Característica 3 |
|---|---|---|---|
| **Amostra 1** | 10 | 6 | 12 |
| **Amostra 2** | 11 | 4 | 9 |
| **Amostra 3** | 8 | 5 | 10 |
| **Amostra 4** | 3 | 3 | 2.5 |
| **Amostra 5** | 1 | 2.8 | 1.3 |
| **Amostra 6** | 2 | 1 | 2 |



# Redução da dimensionalidade

- E quanto a 4 ou mais dimensões?

| | Característica 1 | Característica 2 | Característica 3 | Característica 4 | ... |
|---|---|---|---|---|---|
| Amostra 1 | 10 | 6 | 12 | 5 | ... |
| Amostra 2 | 11 | 4 | 9 | 7 | ... |
| Amostra 3 | 8 | 5 | 10 | 6 | ... |
| Amostra 4 | 3 | 3 | 2.5 | 2 | ... |
| Amostra 5 | 1 | 2.8 | 1.3 | 4 | ... |
| Amostra 6 | 2 | 1 | 2 | 7 | ... |

?

# Análise de componentes principais (PCA)

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

- Identifica os **componentes principais**, que são novas variáveis que **captam a maior variação nos dados**;

- Estes **componentes são ortogonais** (não correlacionados) entre si, permitindo uma redução eficiente das dimensões;

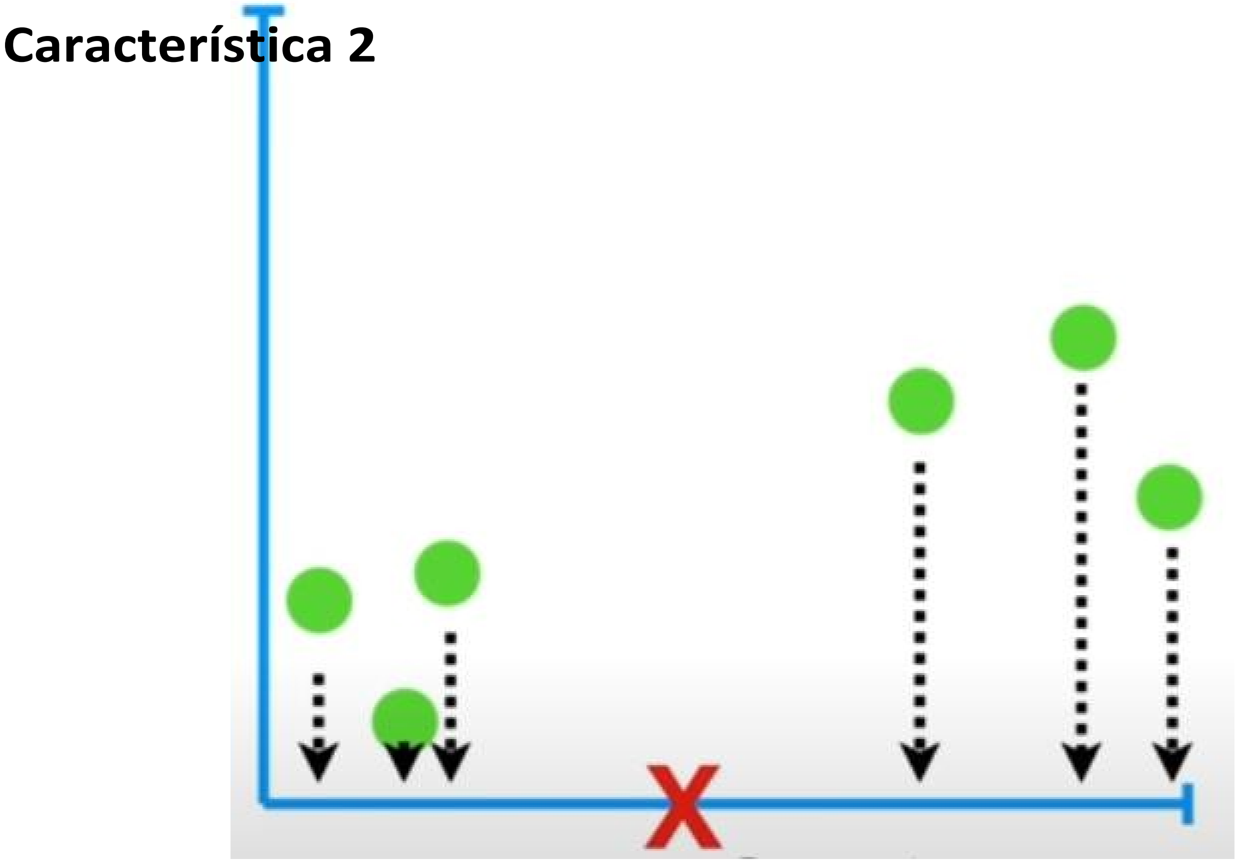
- O **primeiro componente principal explica a quantidade máxima de variância** nos dados, seguido dos componentes subsequentes por ordem decrescente de variância explicada.

# PCA passo a passo

- Vamos começar com um exemplo simples com 2 características.

1.) Centrar os dados na média;

| | Característica 1 | Característica 2 |
|---|---|---|
| **Amostra 1** | 10 | 6 |
| **Amostra 2** | 11 | 4 |
| **Amostra 3** | 8 | 5 |
| **Amostra 4** | 3 | 3 |
| **Amostra 5** | 1 | 2.8 |



| Amostra 6 | 2 | 1 |
|---|---|---|

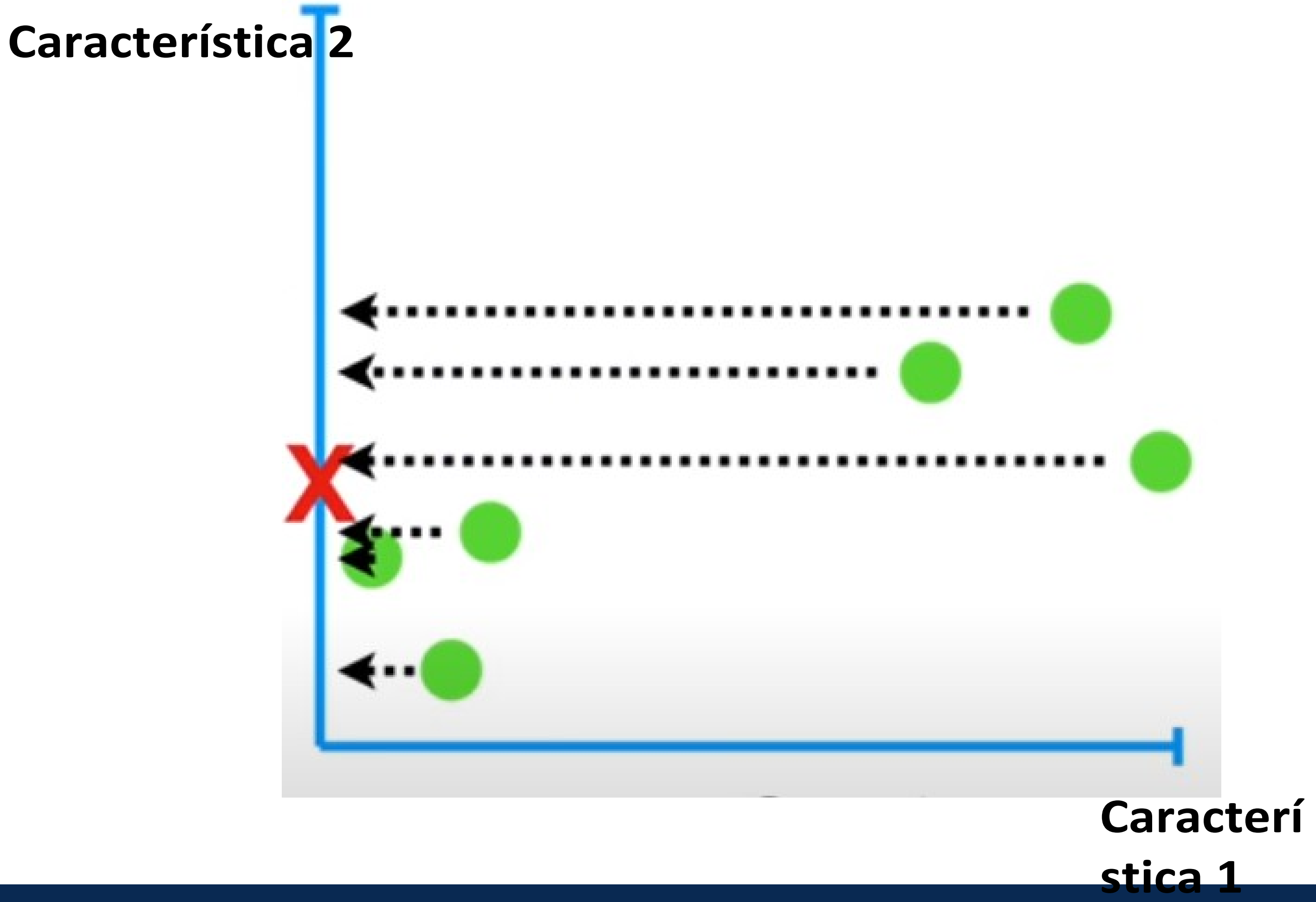
# PCA passo a passo

**Característica 1**

# PCA passo a passo

1.) Centrar a média dos dados;

| | Característica 1 | Característica 2 |
|---|---|---|
| Amostra 1 | 10 | 6 |
| Amostra 2 | 11 | 4 |
| Amostra 3 | 8 | 5 |
| Amostra 4 | 3 | 3 |
| Amostra 5 | 1 | 2.8 |
| Amostra 6 | 2 | 1 |

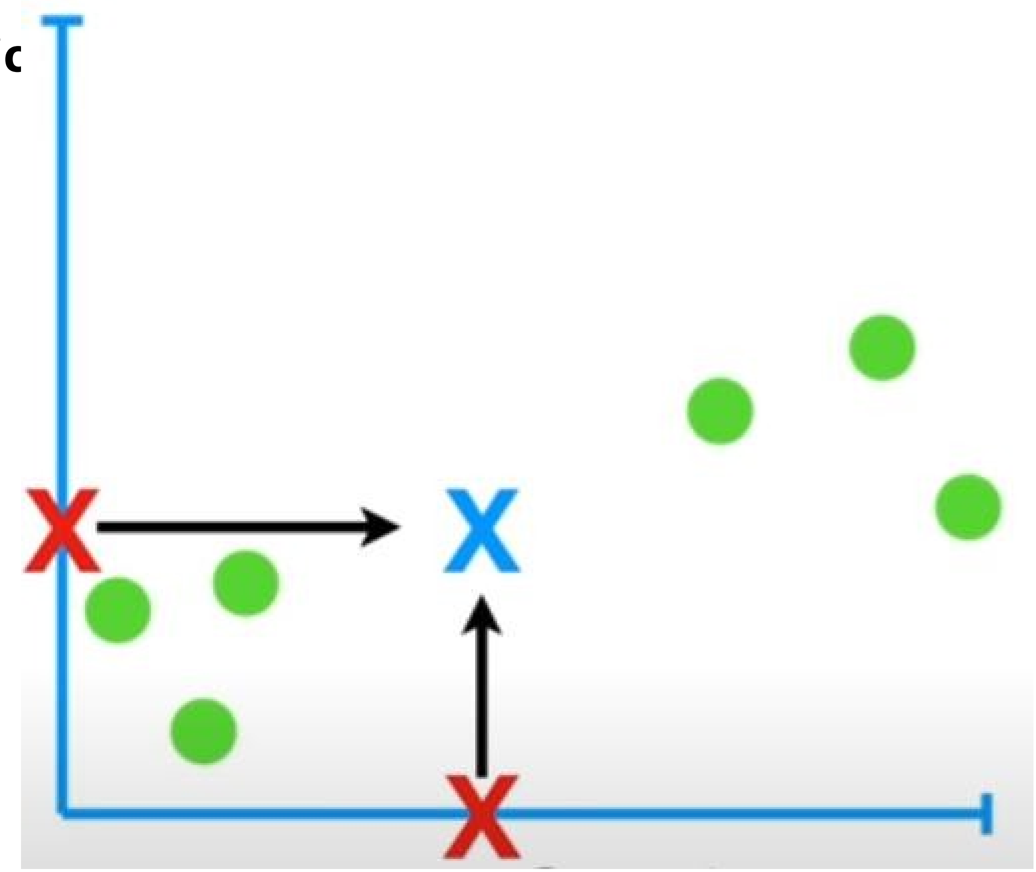


# PCA passo a passo

1.) Centro médio dos dados;

Característic

| | Característica 1 | Característica 2 |
|---|---|---|
| **Amostra 1** | 10 | 6 |
| **Amostra 2** | 11 | 4 |
| **Amostra 3** | 8 | 5 |
| **Amostra 4** | 3 | 3 |
| **Amostra 5** | 1 | 2.8 |
| **Amostra 6** | 2 | 1 |

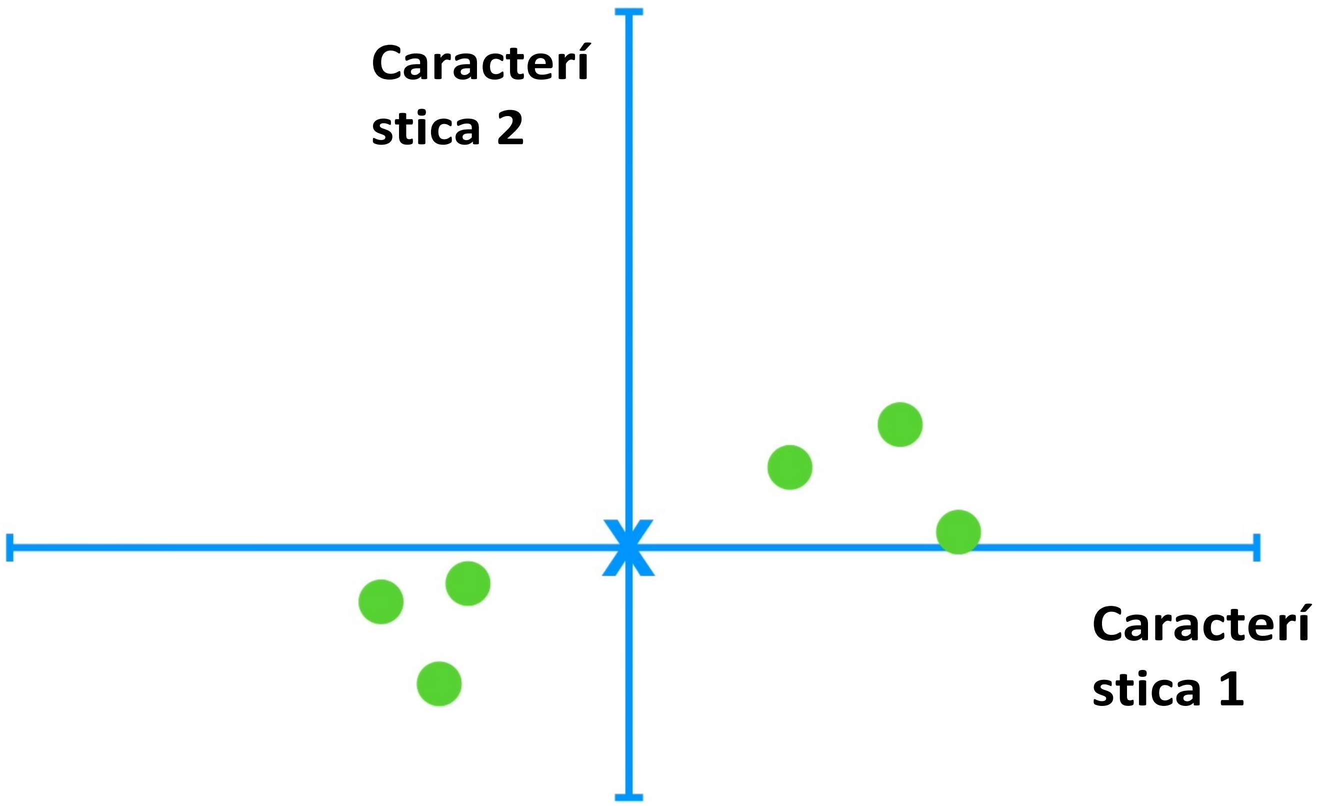


Característica 1

# PCA passo a passo

1.) Centro médio dos dados;

| | Característica 1 | Característica 2 |
|---|---|---|
| **Amostra 1** | 10 | 6 |
| **Amostra 2** | 11 | 4 |
| **Amostra 3** | 8 | 5 |
| **Amostra 4** | 3 | 3 |
| **Amostra 5** | 1 | 2.8 |
| **Amostra 6** | 2 | 1 |

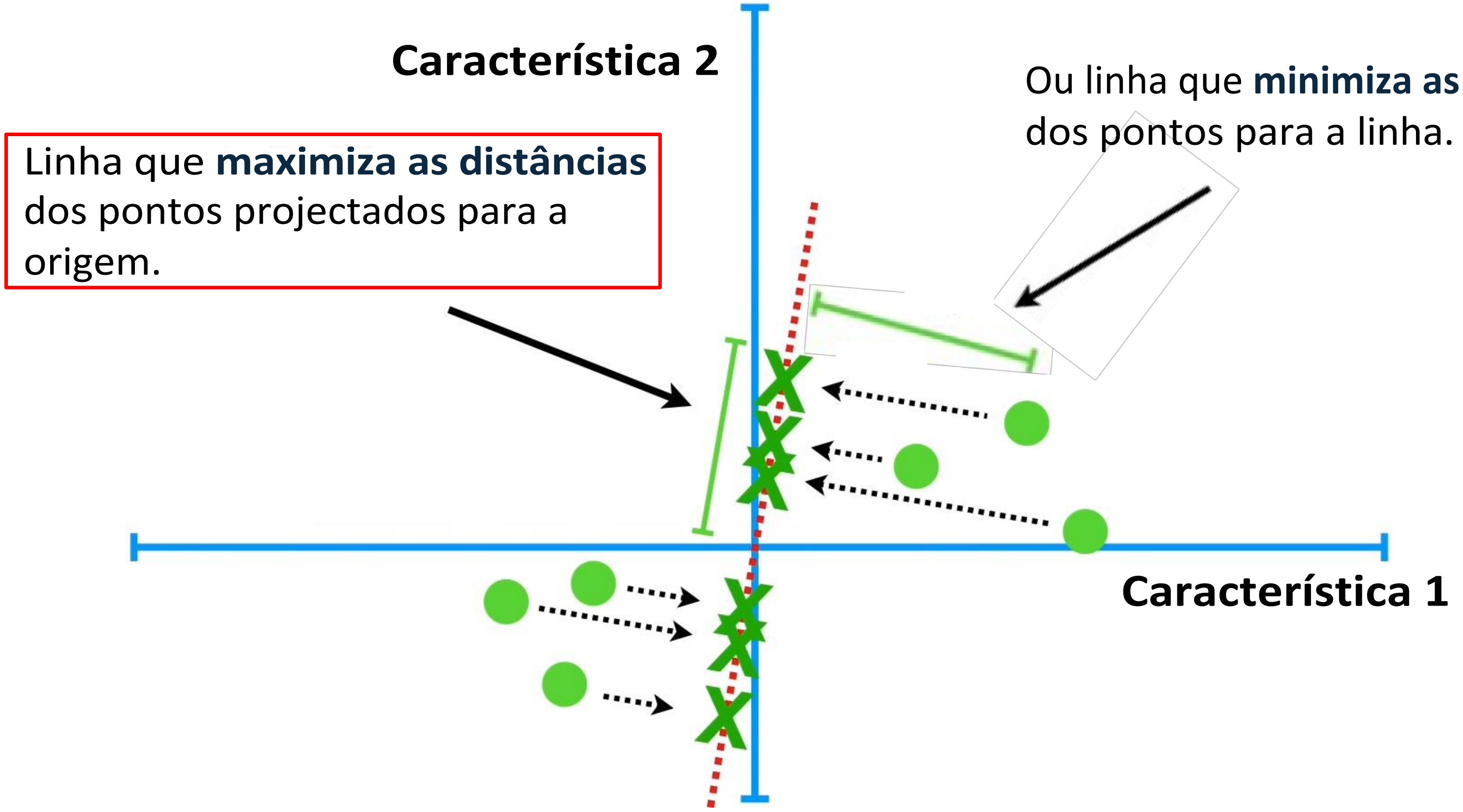


# PCA passo a passo

2.) Calcular PC1 utilizando a decomposição do valor singular (SVD).

# PCA passo a passo

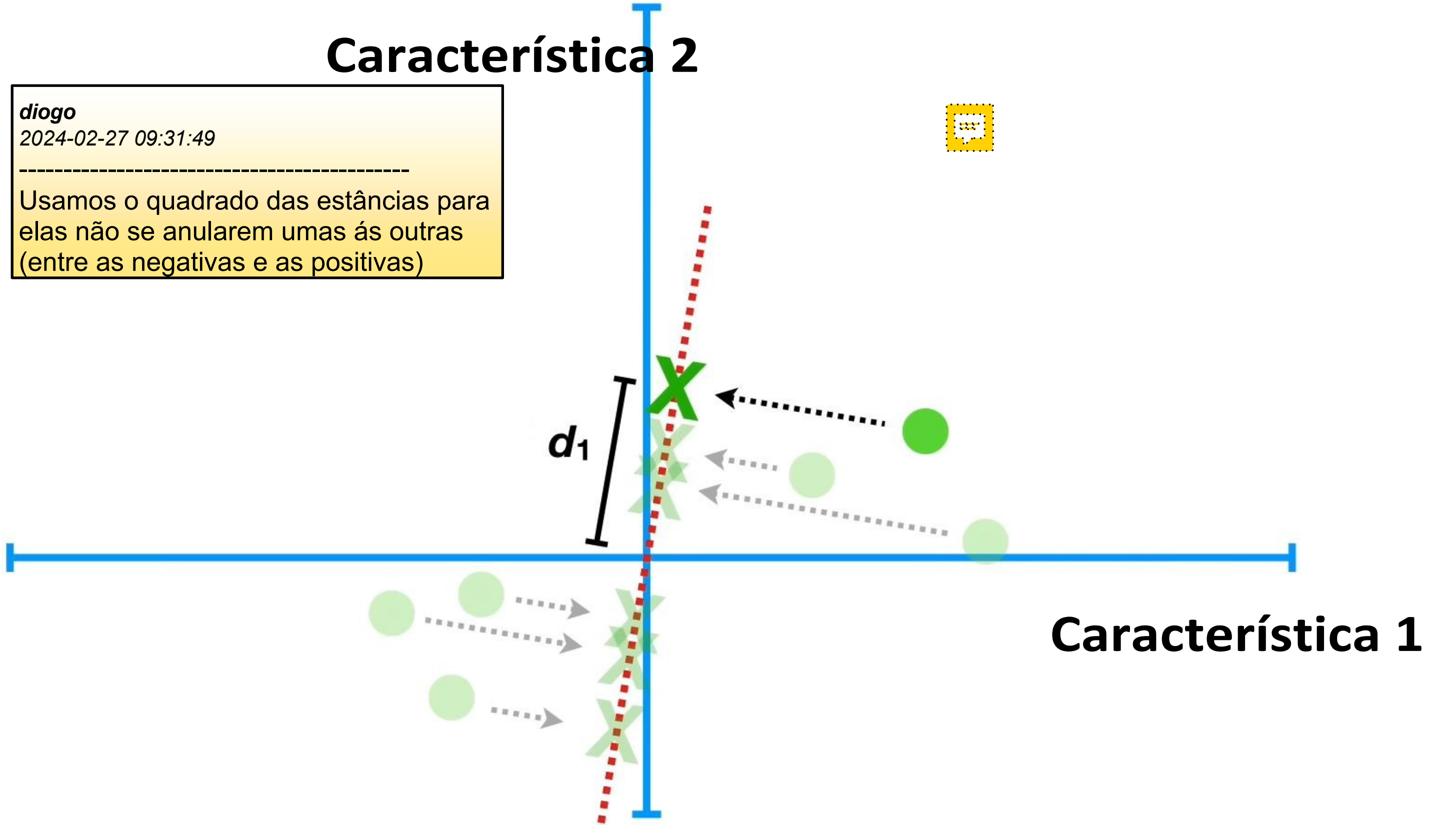
2.) Calcular PC1 utilizando a decomposição em valores singulares (SVD).



# PCA passo a passo

2.) Calcular PC1 utilizando a decomposição em valores singulares (SVD).

*diogo*
*2024-02-27 09:31:49*
--------------------------------------------
Usamos o quadrado das estâncias para elas não se anularem umas ás outras (entre as negativas e as positivas)



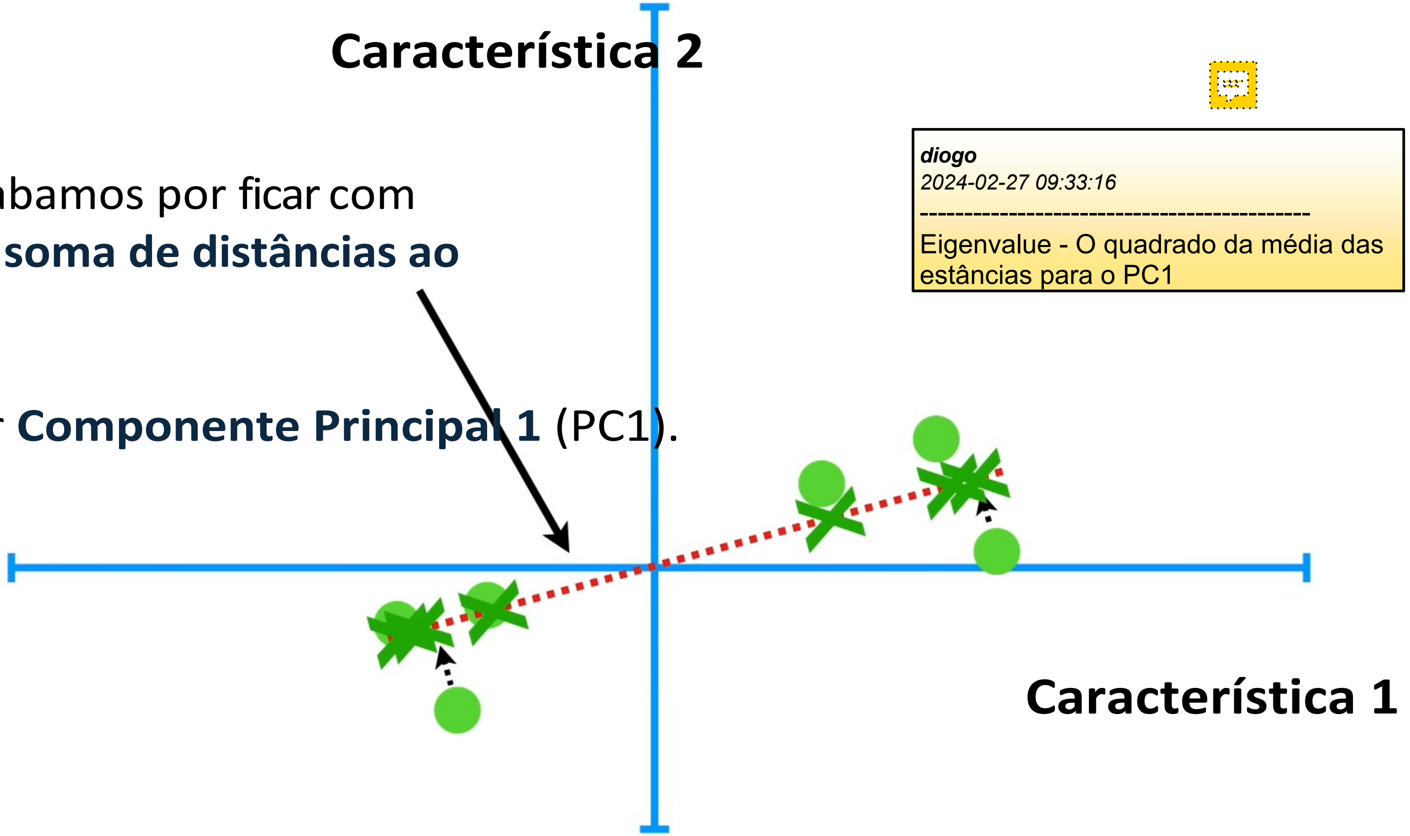

$d1^2 + d2^2 + d3^2 + d4^2 + d5^2 + d6^2$ = **soma das distâncias ao quadrado**

# PCA passo a passo

2.) Calcular PC1 utilizando a decomposição em valores singulares (SVD).

Acabamos por ficar com esta linha. Tem a **maior soma de distâncias ao quadrado**.

Esta linha é designada por **Componente Principal 1** (PC1).




*diogo*
*2024-02-27 09:33:16*
-------------------------------------------
Eigenvalue - O quadrado da média das estâncias para o PC1

$d1^2 + d2^2 + d3^2 + d4^2 + d5^2 + d6^2$ = soma das distâncias ao quadrado

# PCA passo a passo

A média da soma das distâncias ao quadrado para PC1 é designada por **Eigenvalue** para PC1.

# PCA passo a passo

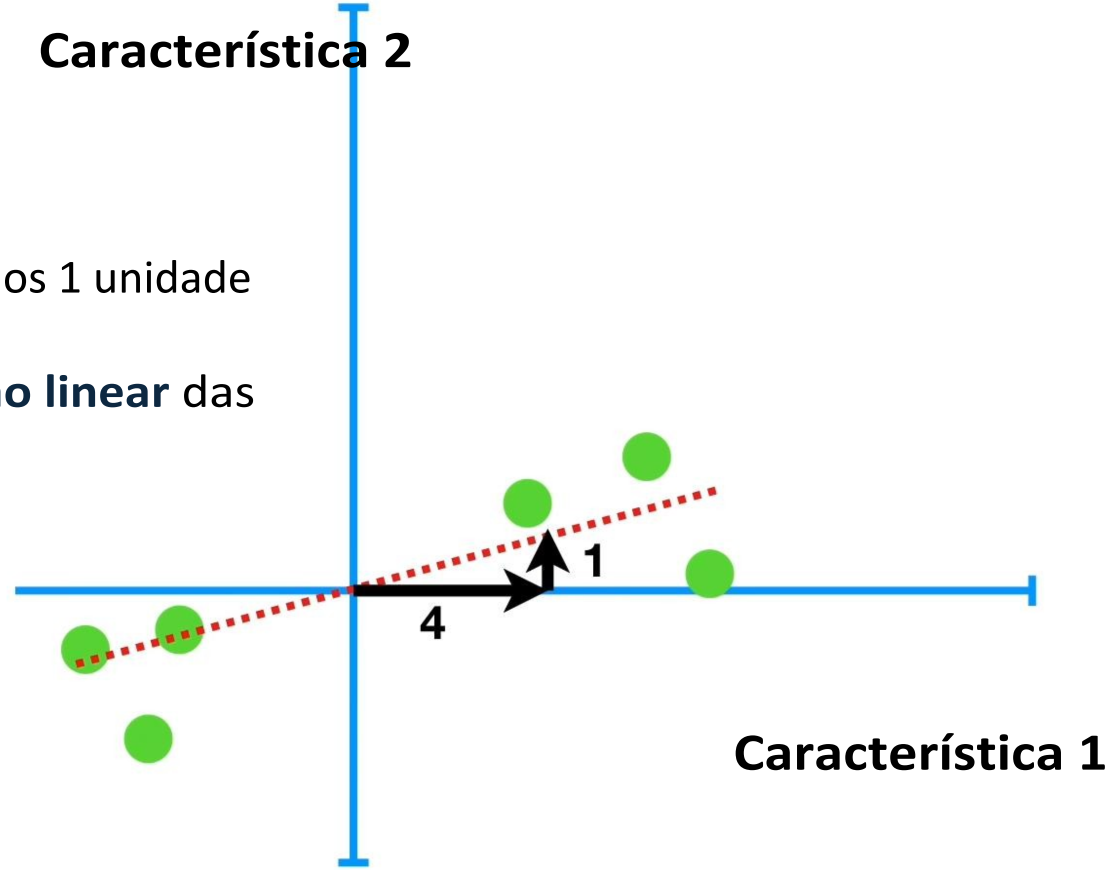
## 2.) Calcular PC1 utilizando a decomposição em valores singulares (SVD).

Digamos que o declive da reta é 0,25.

Para 4 unidades da caraterística 1 aumentamos 1 unidade

na caraterística 2. O PC1 é uma **combinação linear** das

características 1 e 2.

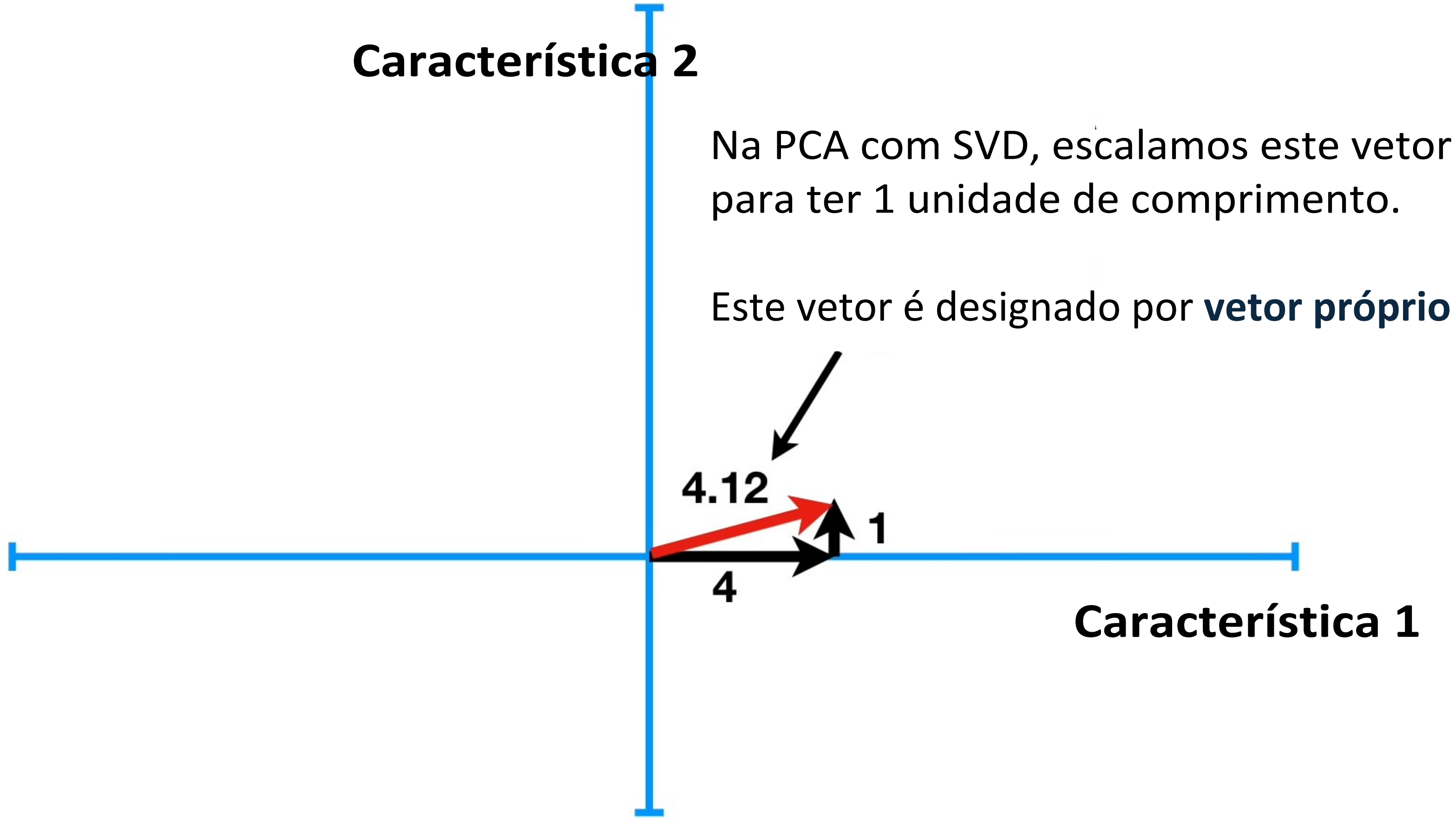


# PCA passo a passo

2.) Calcular PC1 utilizando a decomposição em valores singulares (SVD).

Na PCA com SVD, escalamos este vetor para ter 1 unidade de comprimento.

Este vetor é designado por **vetor próprio** do PC1.

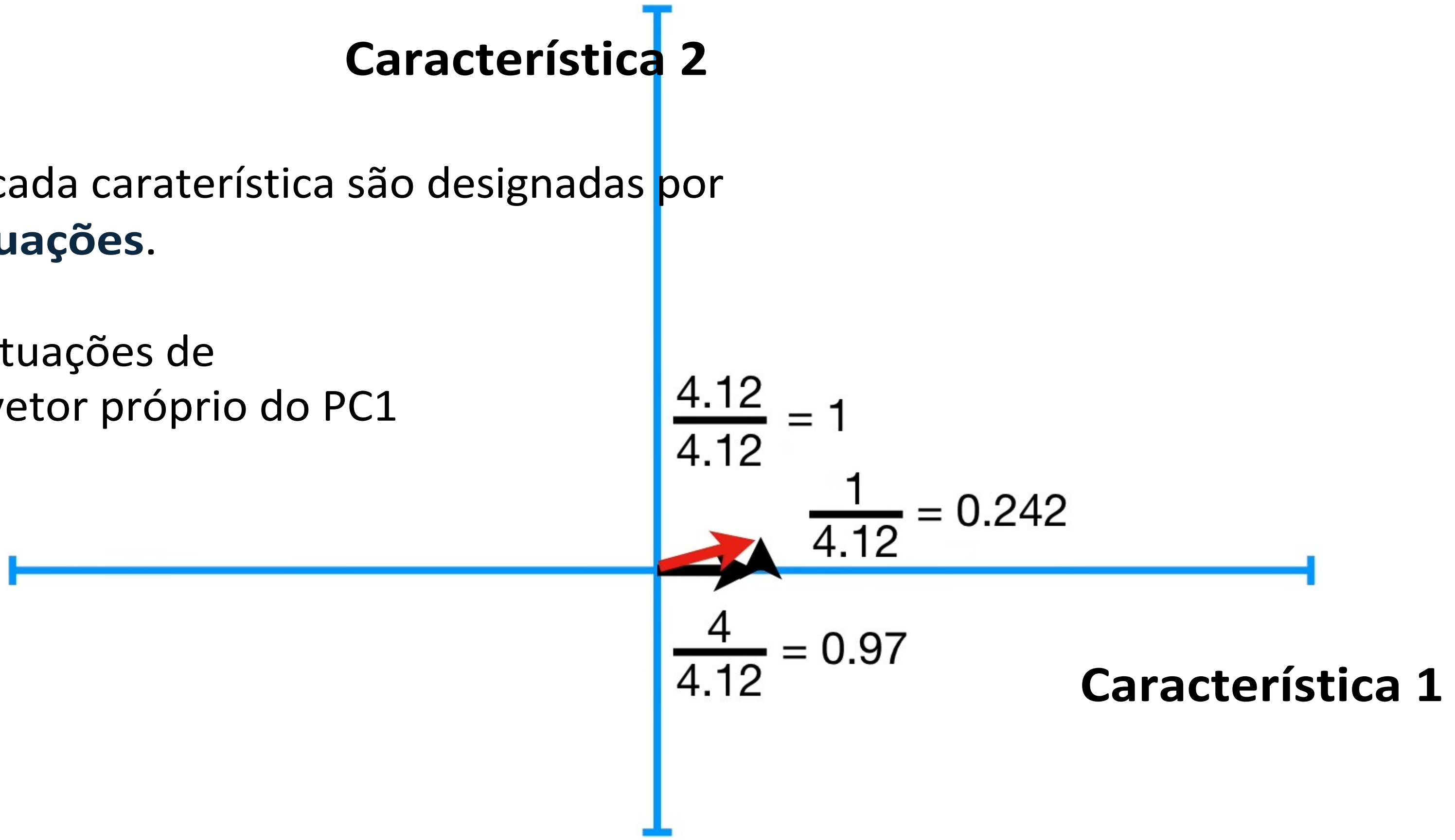


# PCA passo a passo

2.) Calcular PC1 utilizando a decomposição do valor singular (SVD).

As proporções de cada caraterística são designadas por **Carregando pontuações**.

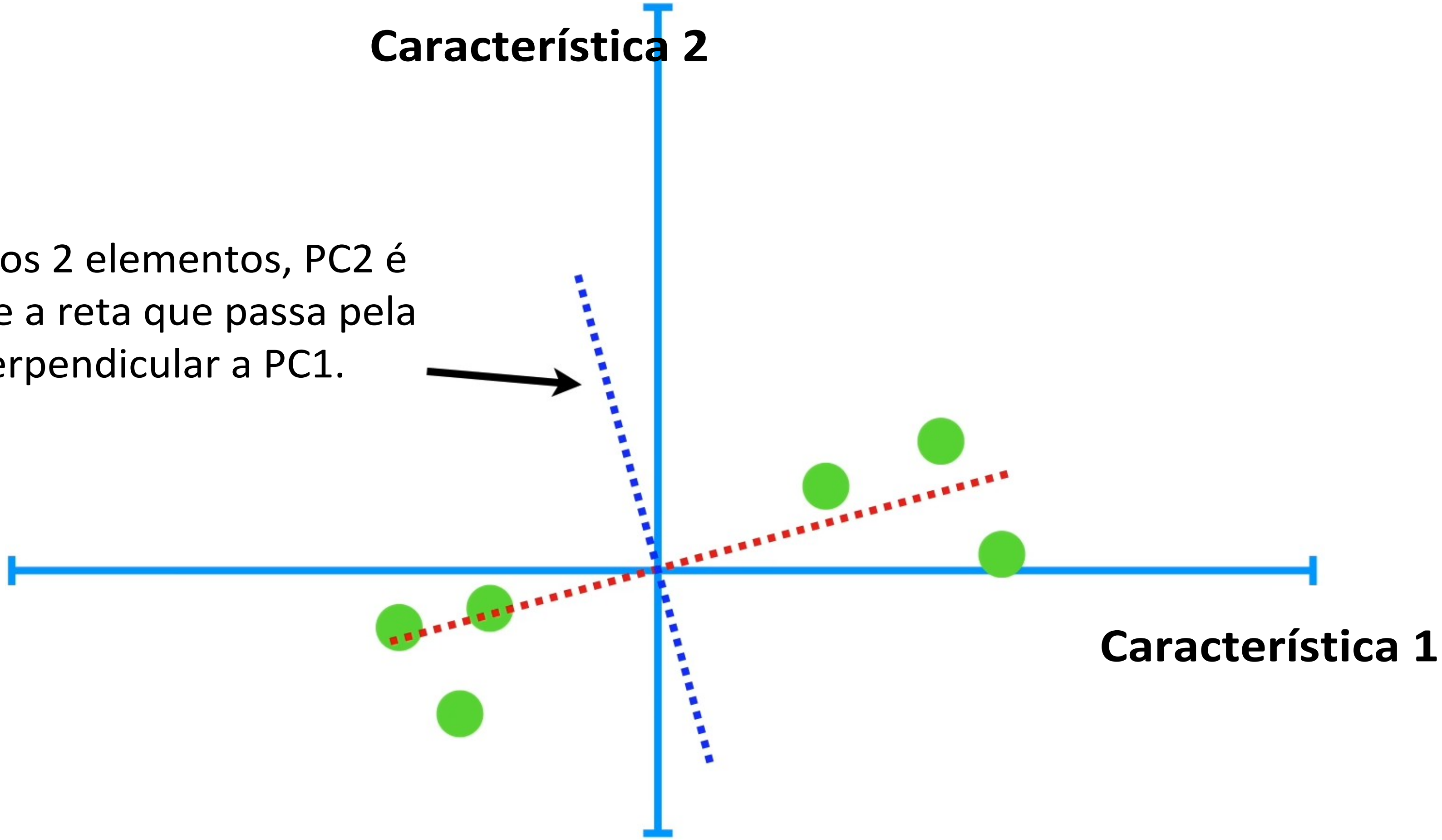
Neste caso, as pontuações de carregamento do vetor próprio do PC1 são 0,97 e 0,242.



# PCA passo a passo

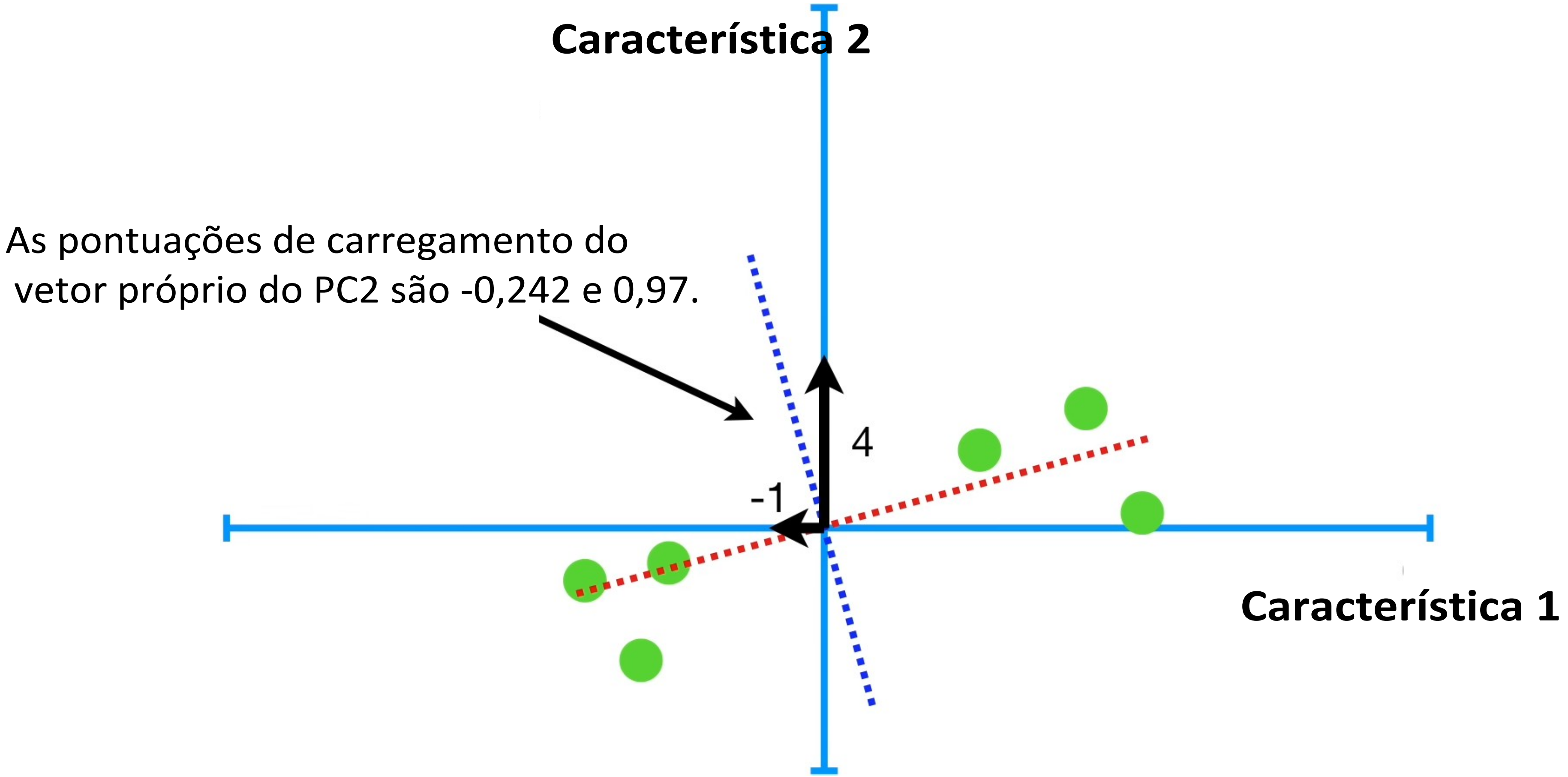
3.) Calcular o PC2 utilizando a decomposição do valor singular (SVD).

Como só temos 2 elementos, PC2 é simplesmente a reta que passa pela origem e é perpendicular a PC1.



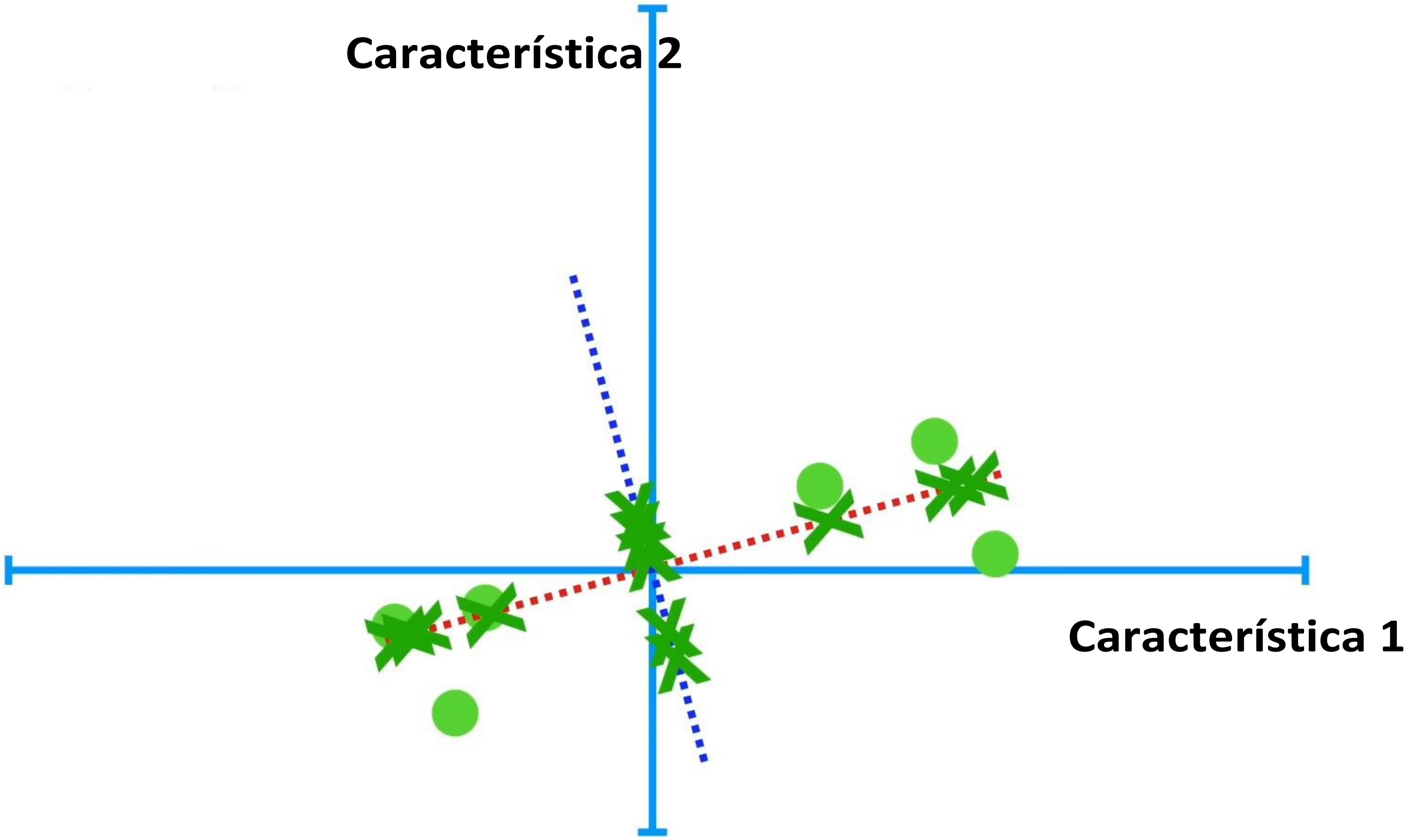
# PCA passo a passo

3.) Calcular o PC2 utilizando a decomposição do valor singular (SVD).



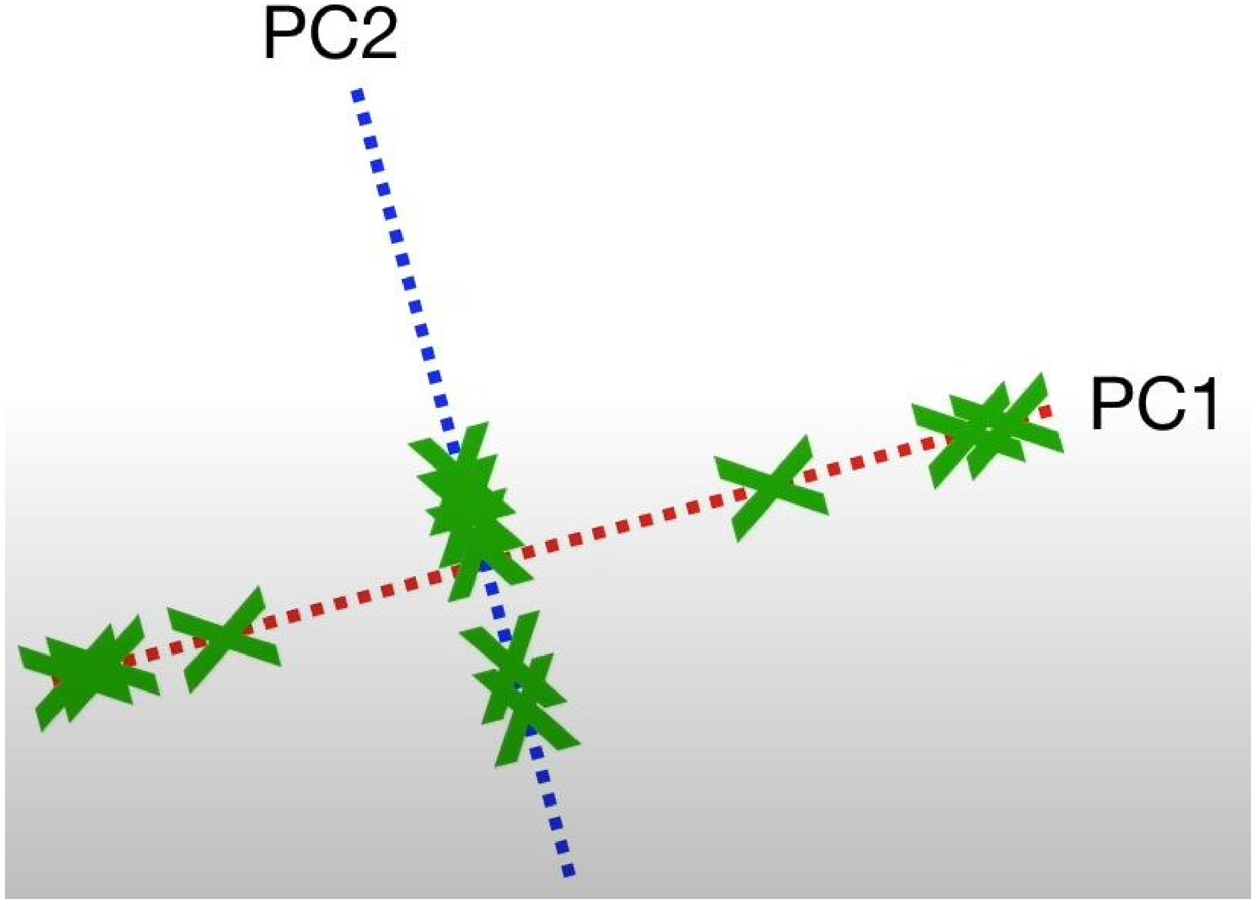
# PCA passo a passo

4.) Obter as coordenadas finais ao longo dos PCs.



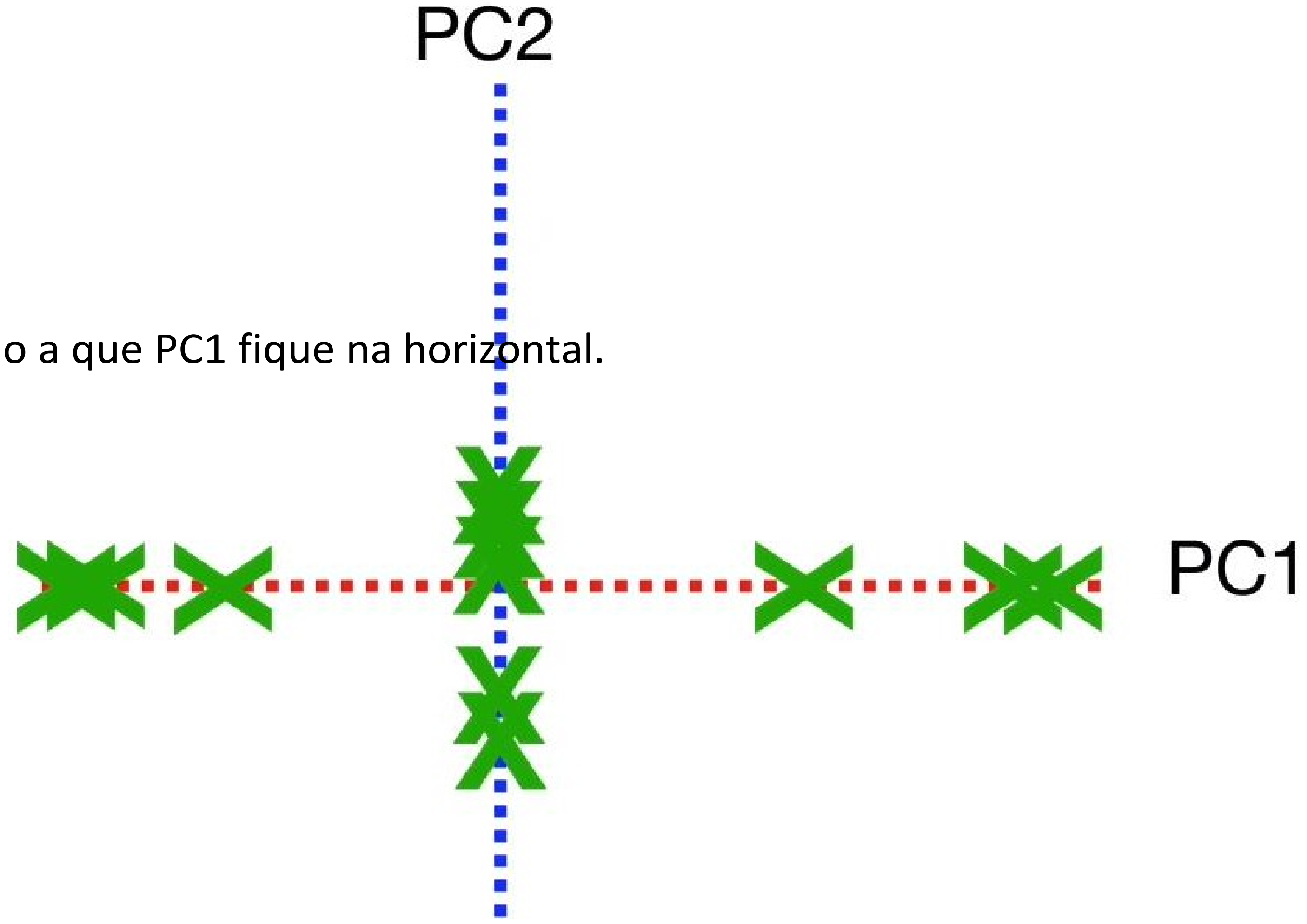
# PCA passo a passo

4.) Obter as coordenadas finais ao longo dos PCs.



# PCA passo a passo

4.) Obter as coordenadas finais ao longo dos PCs.

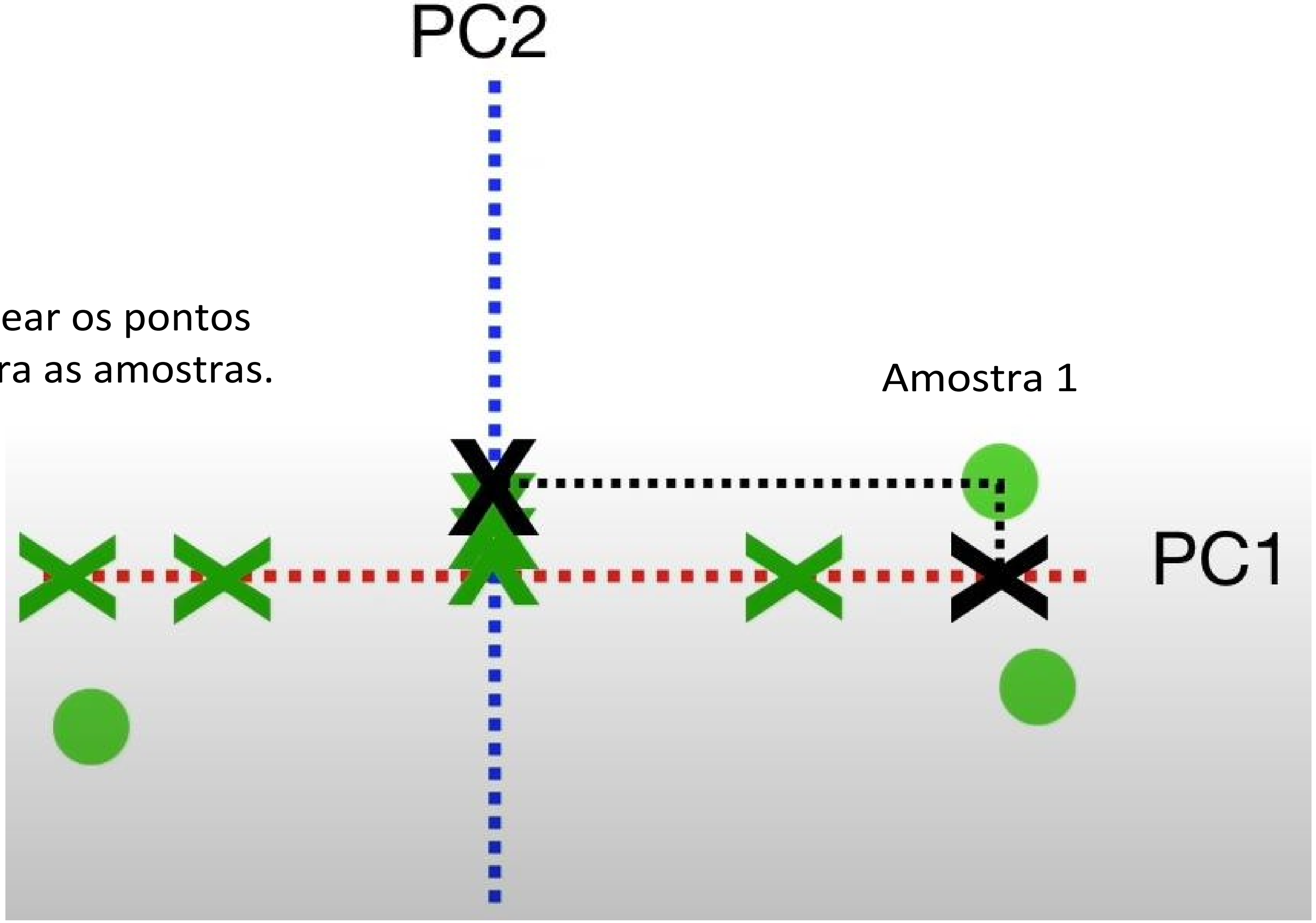



Rodar o gráfico de modo a que PC1 fique na horizontal.

# PCA passo a passo

4.) Obter as coordenadas finais ao longo dos PCs.

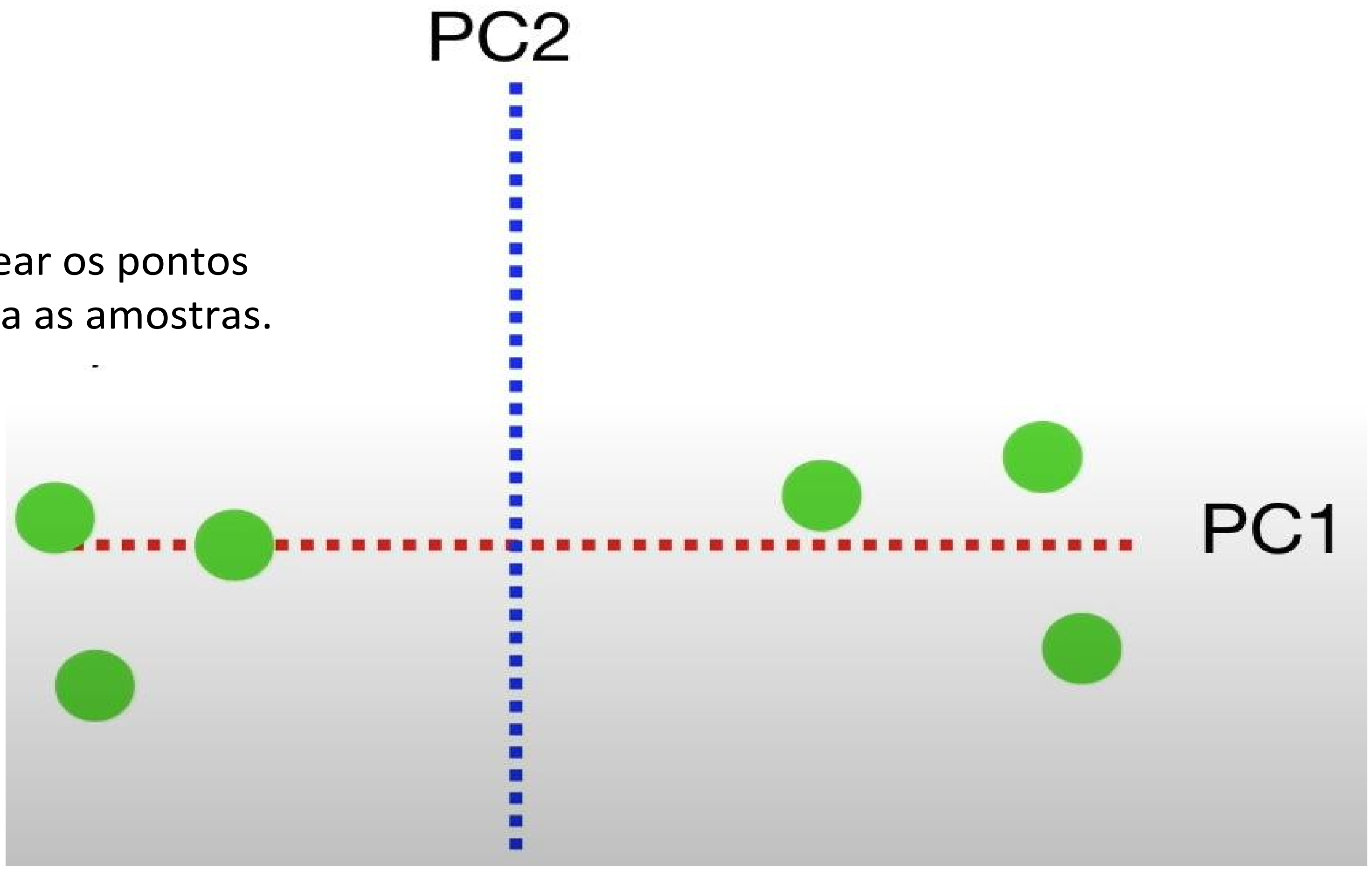
Podemos agora mapear os pontos projectados bach para as amostras.



# PCA passo a passo

4.) Obter as coordenadas finais ao longo dos PCs.

Podemos agora mapear os pontos projectados bach para as amostras.



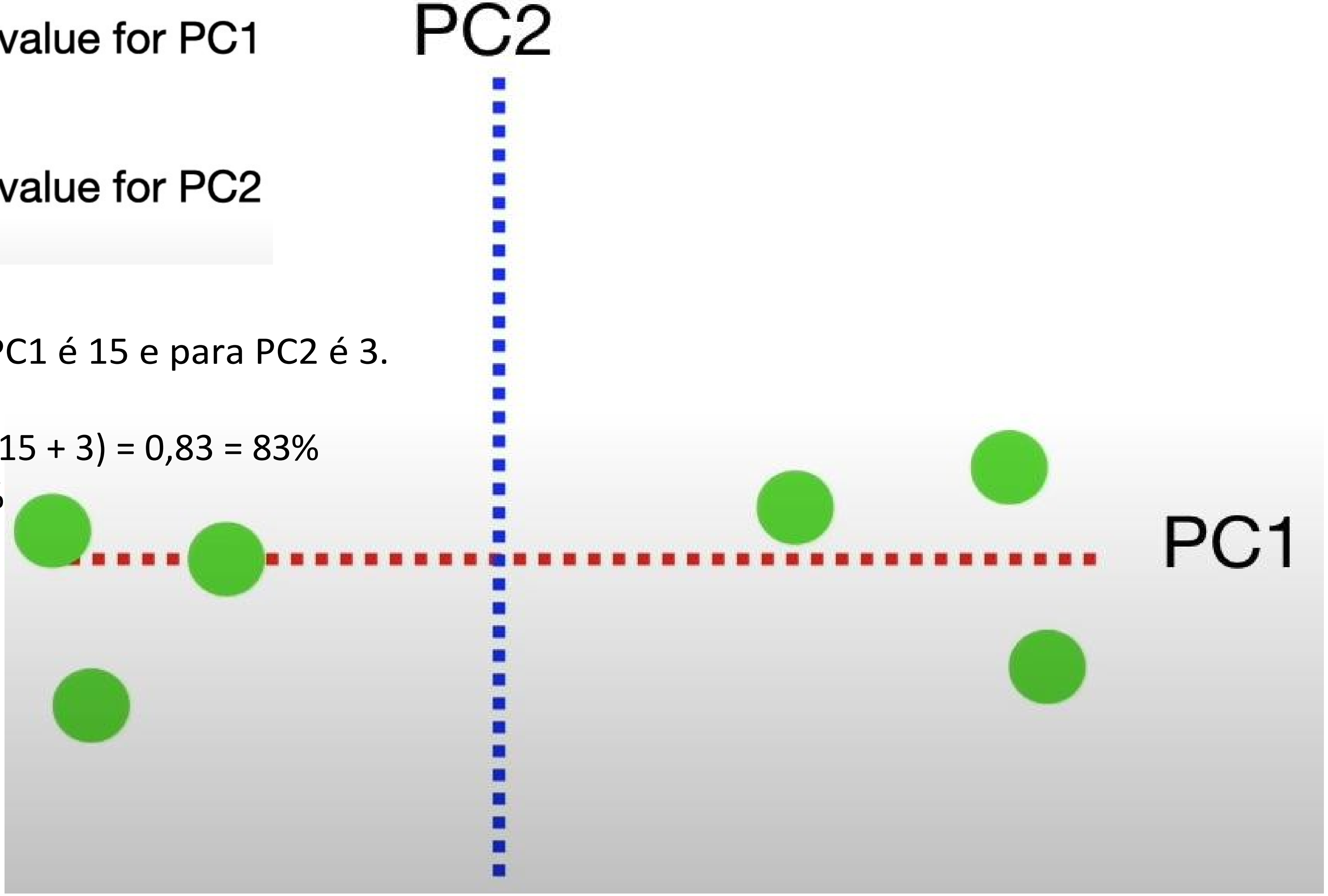
# PCA - Variância Explicada

$$\frac{\text{SS(distances for PC1)}}{n - 1} = \text{Eigenvalue for PC1}$$

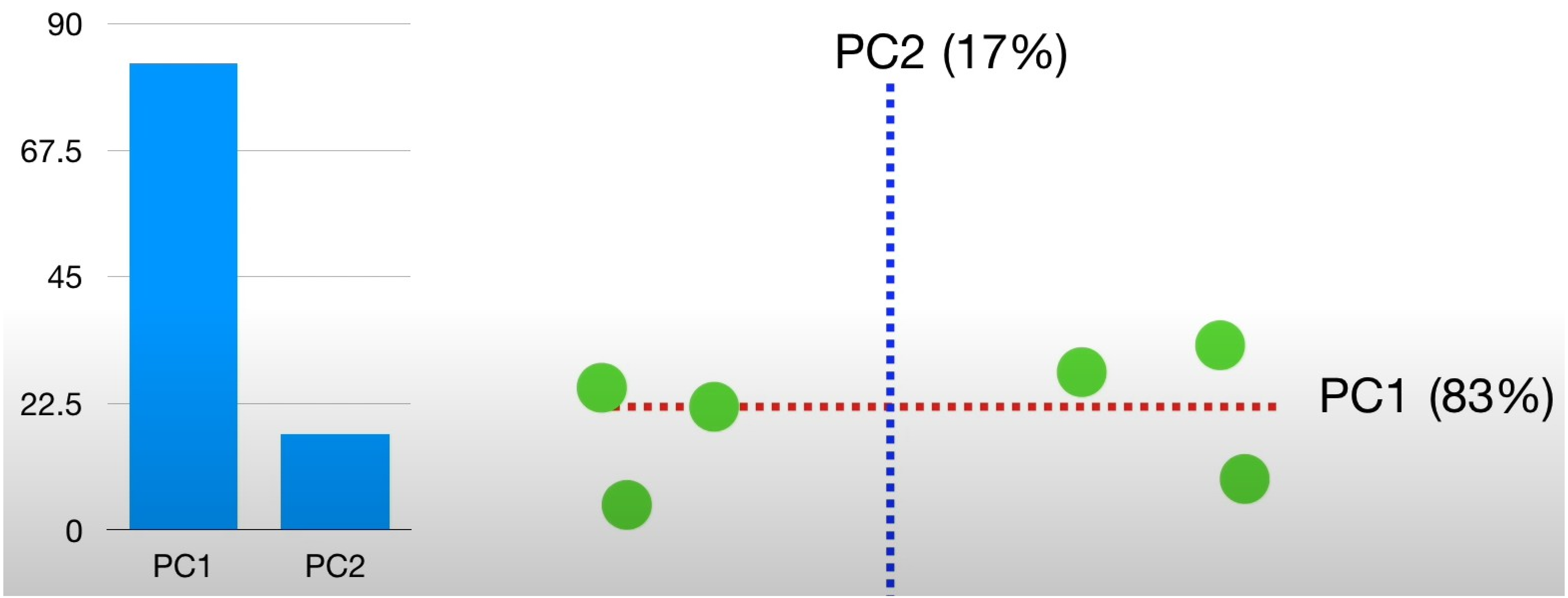
$$\frac{\text{SS(distances for PC2)}}{n - 1} = \text{Eigenvalue for PC2}$$

Digamos que o valor próprio para PC1 é 15 e para PC2 é 3.

A **variância explicada** do PC1 = 15 / (15 + 3) = 0,83 = 83%
Para PC2 = 3 / (15 + 3) = 0,17 = 17%



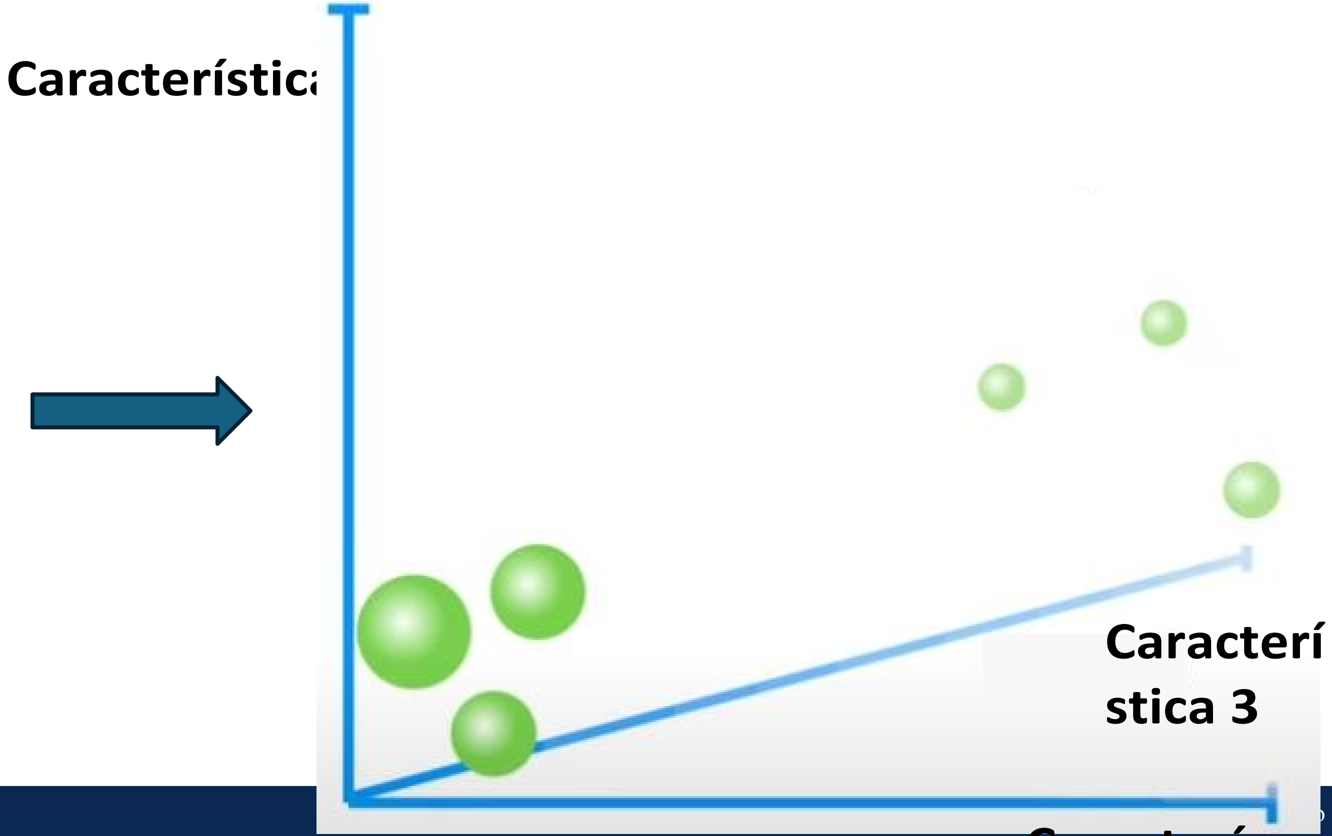
# PCA - Variância explicada - Scree Plot



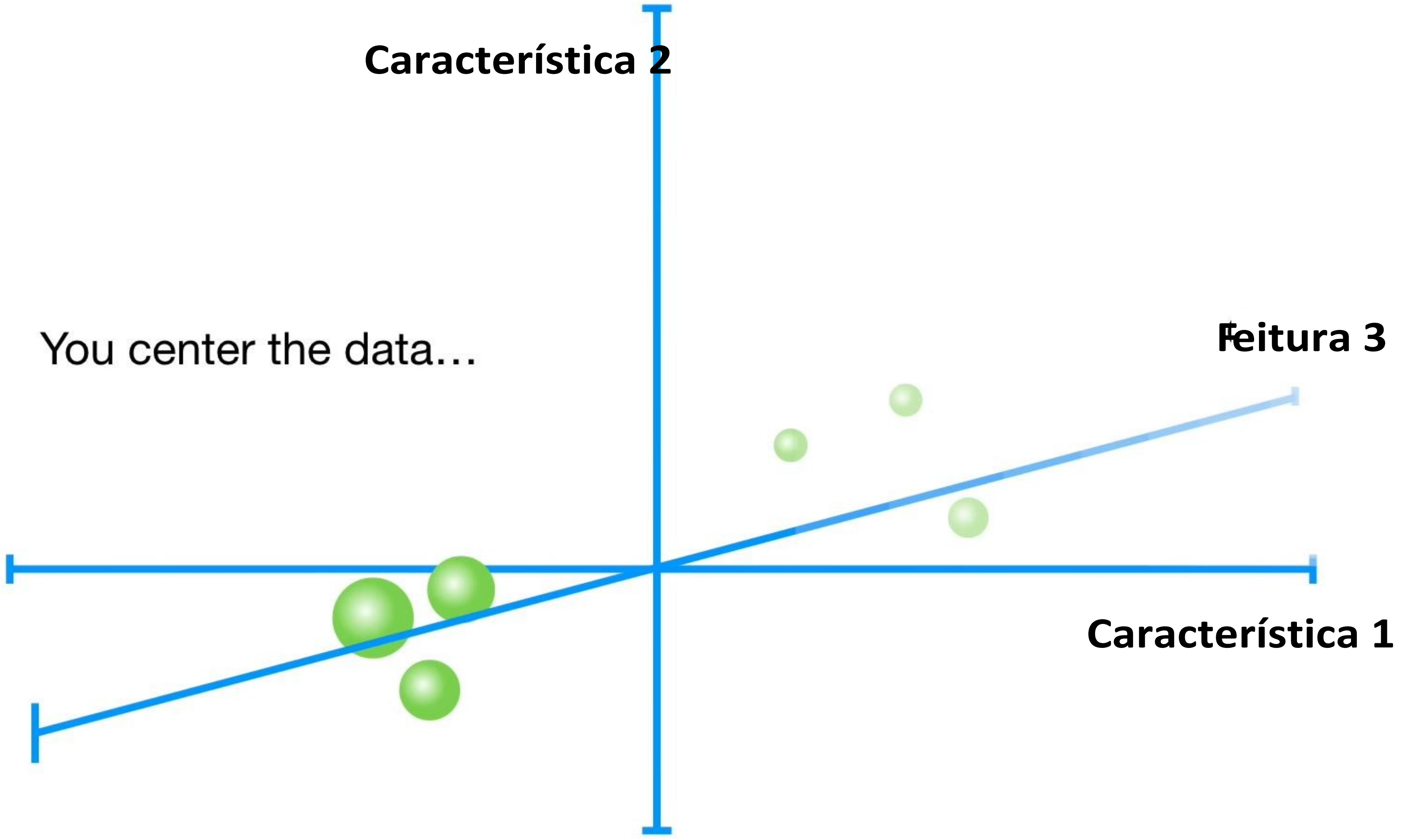
# PCA com mais recursos

- PCA com 3 ou mais características é praticamente o mesmo que 2 características...

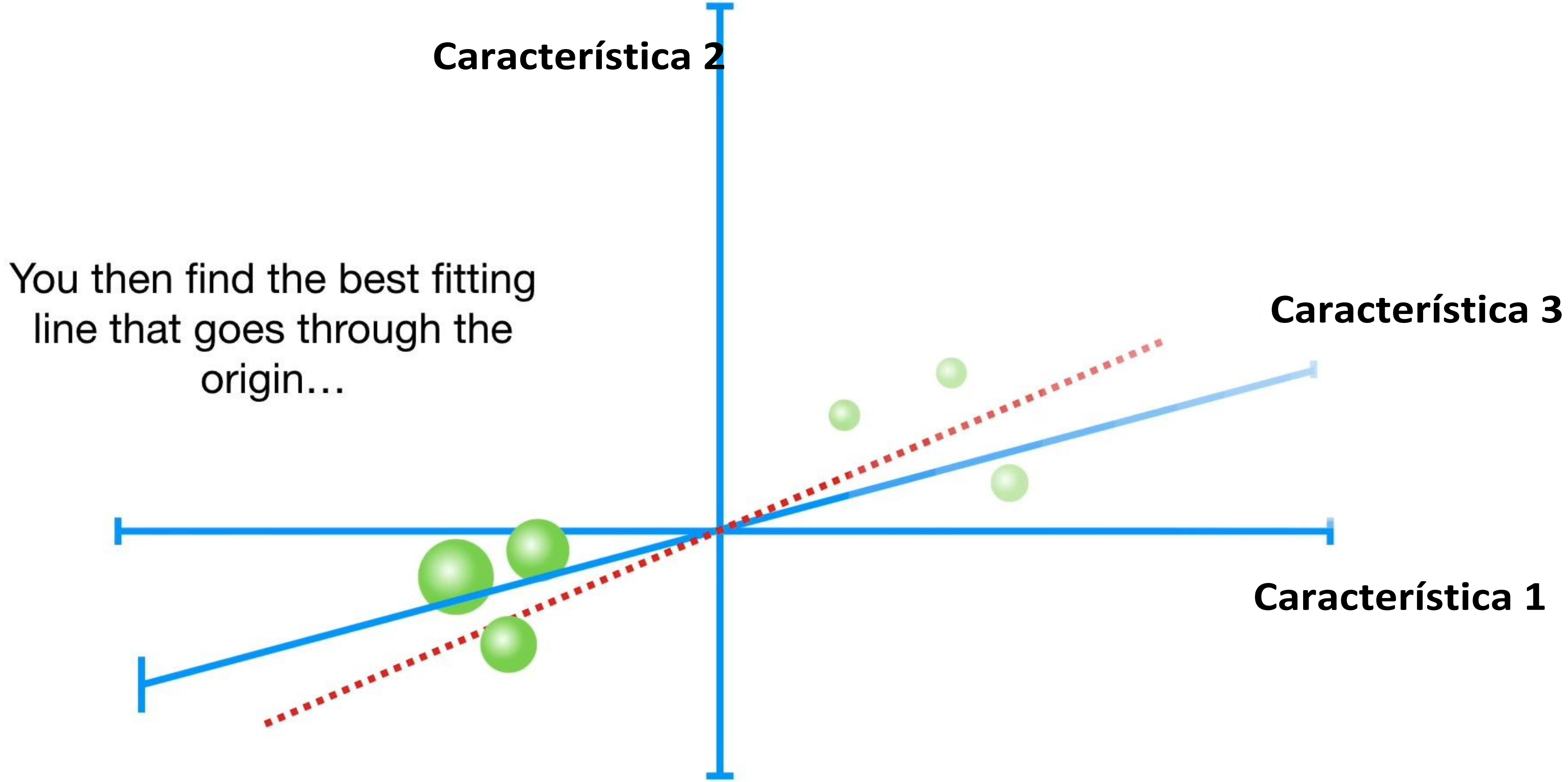
| | Característica 1 | Característica 2 | Característica 3 |
|---|---|---|---|
| **Amostra 1** | 10 | 6 | 12 |
| **Amostra 2** | 11 | 4 | 9 |
| **Amostra 3** | 8 | 5 | 10 |
| **Amostra 4** | 3 | 3 | 2.5 |
| **Amostra 5** | 1 | 2.8 | 1.3 |
| **Amostra 6** | 2 | 1 | 2 |



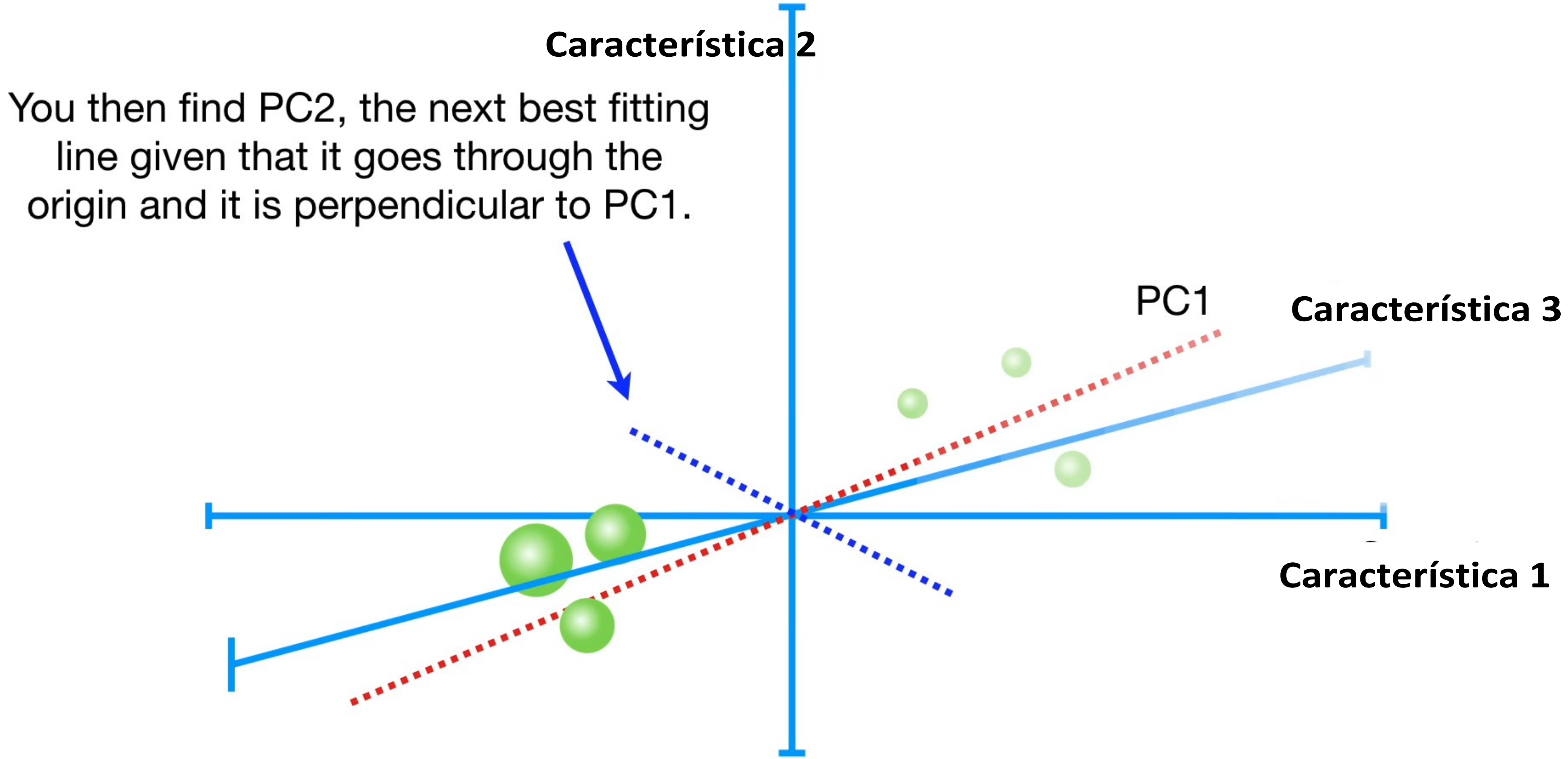
# PCA com mais recursos



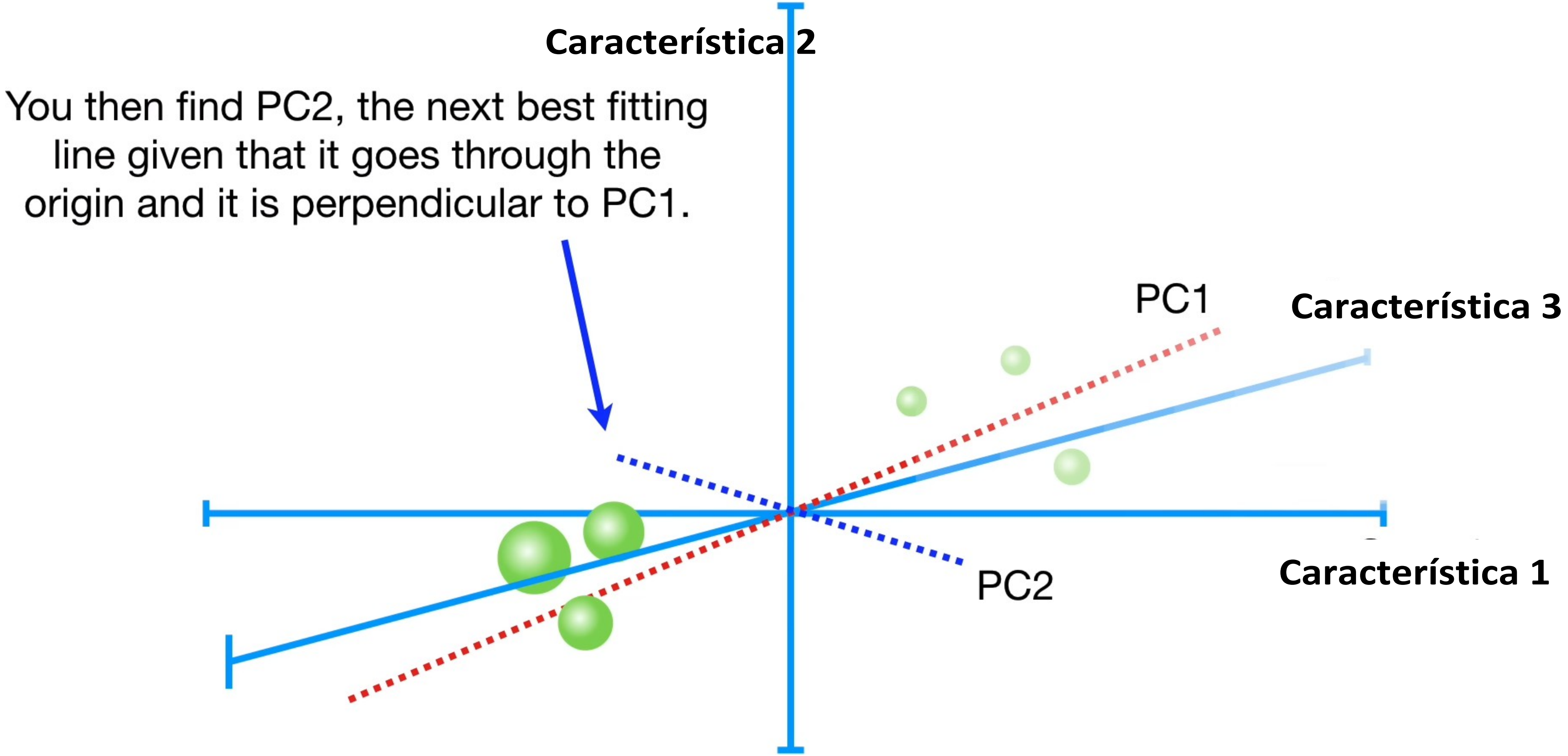
# PCA com mais recursos



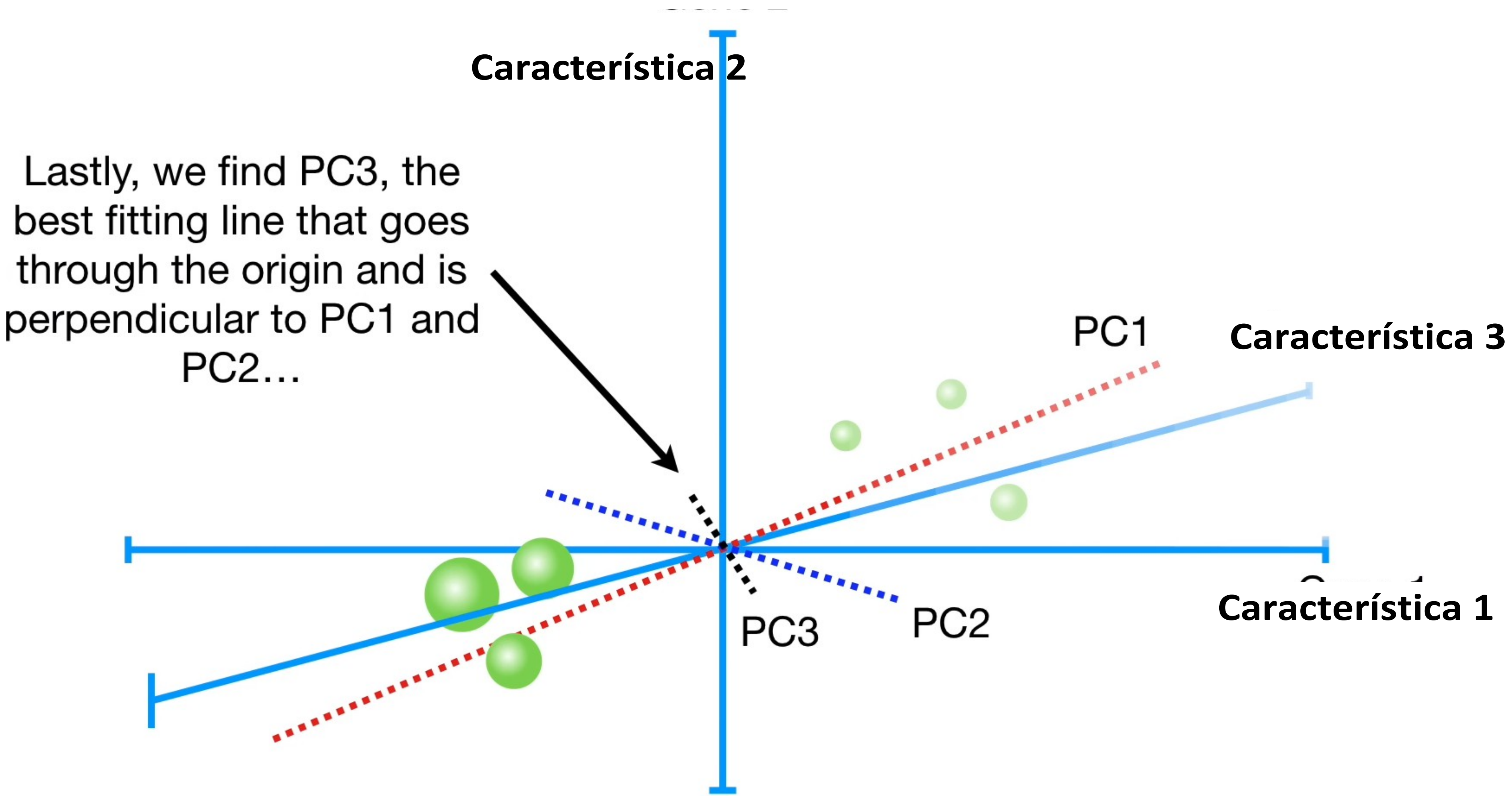
# PCA com mais recursos



# PCA com mais recursos

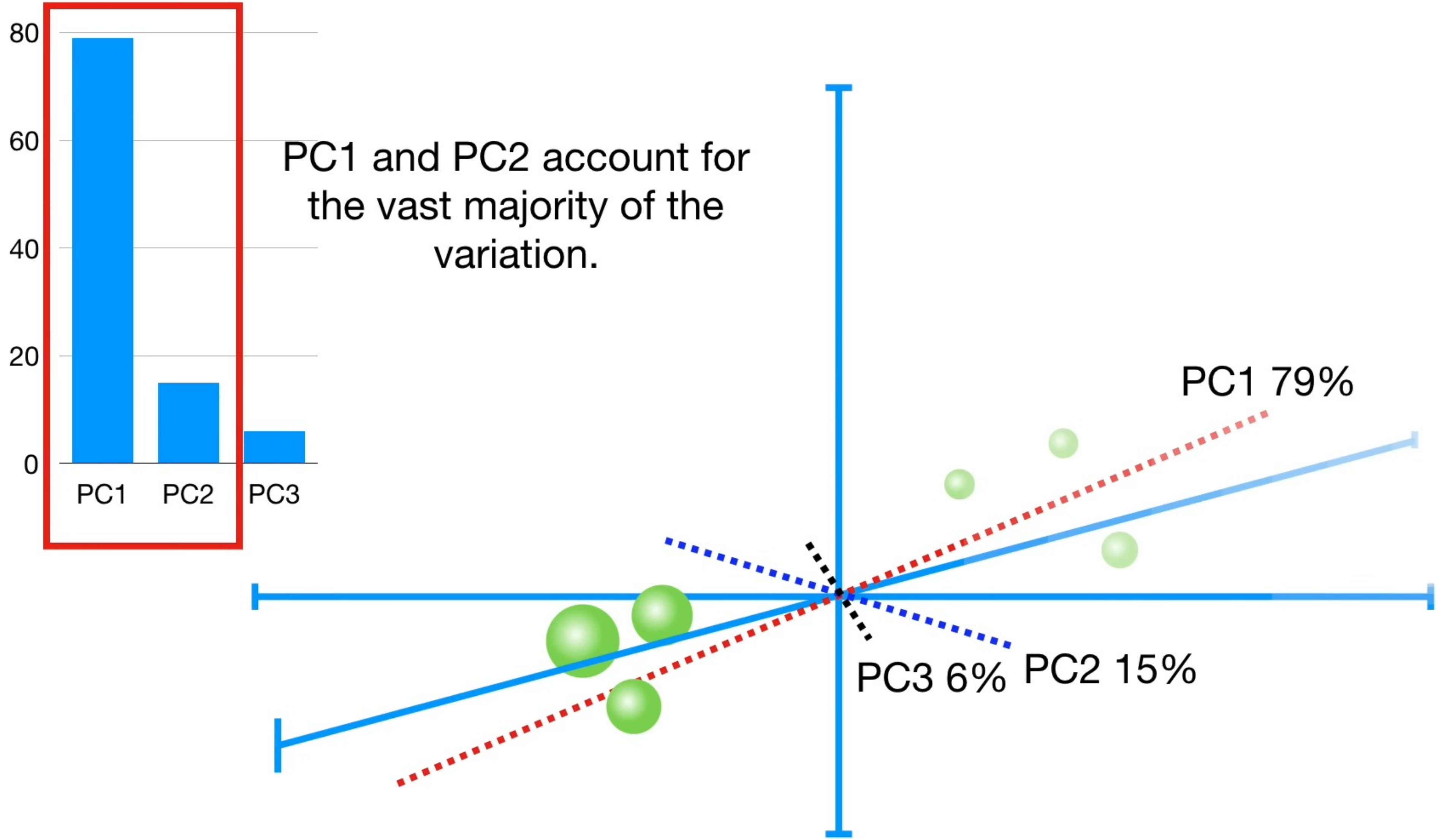


# PCA com mais recursos

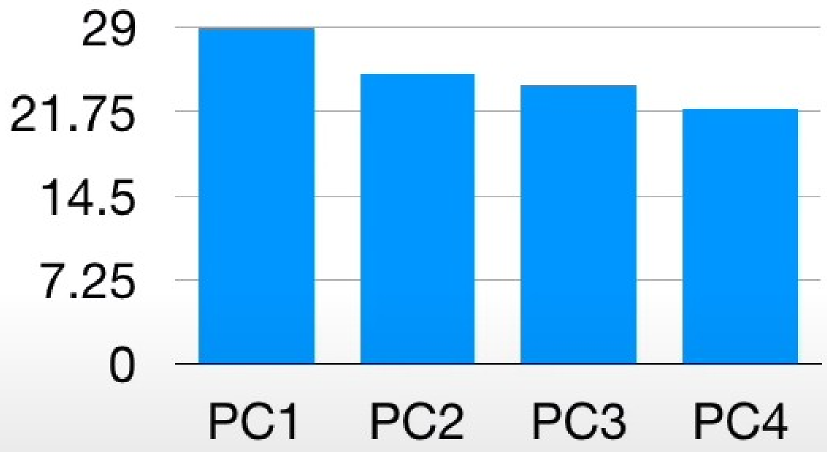


# PCA com mais recursos



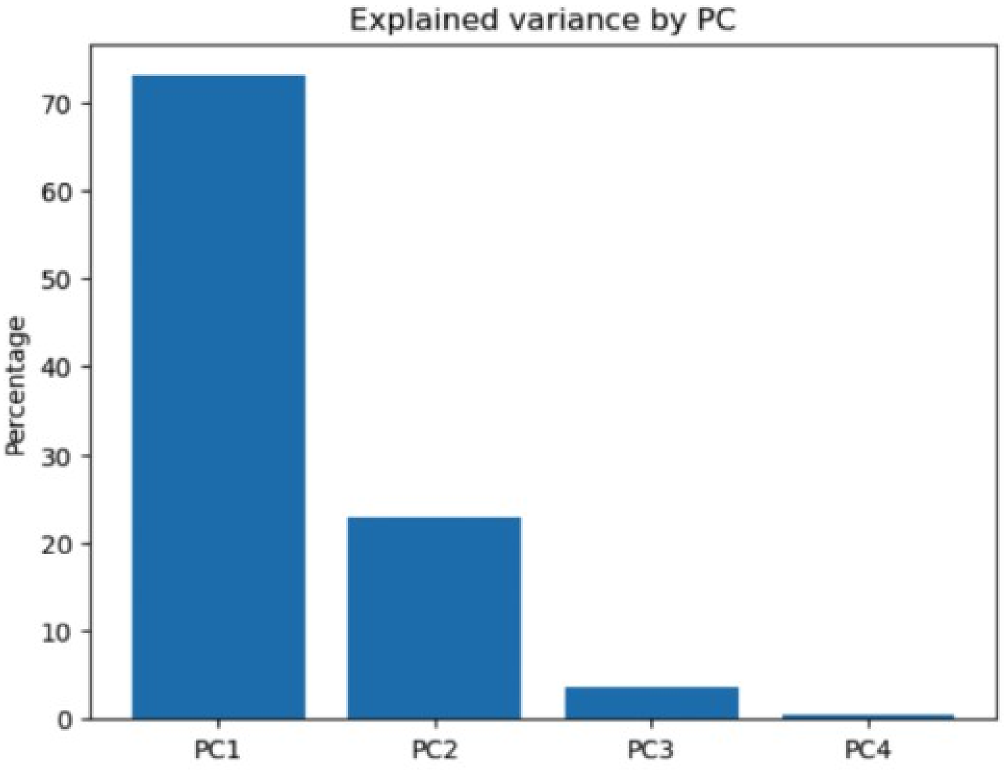


# PCA com mais recursos

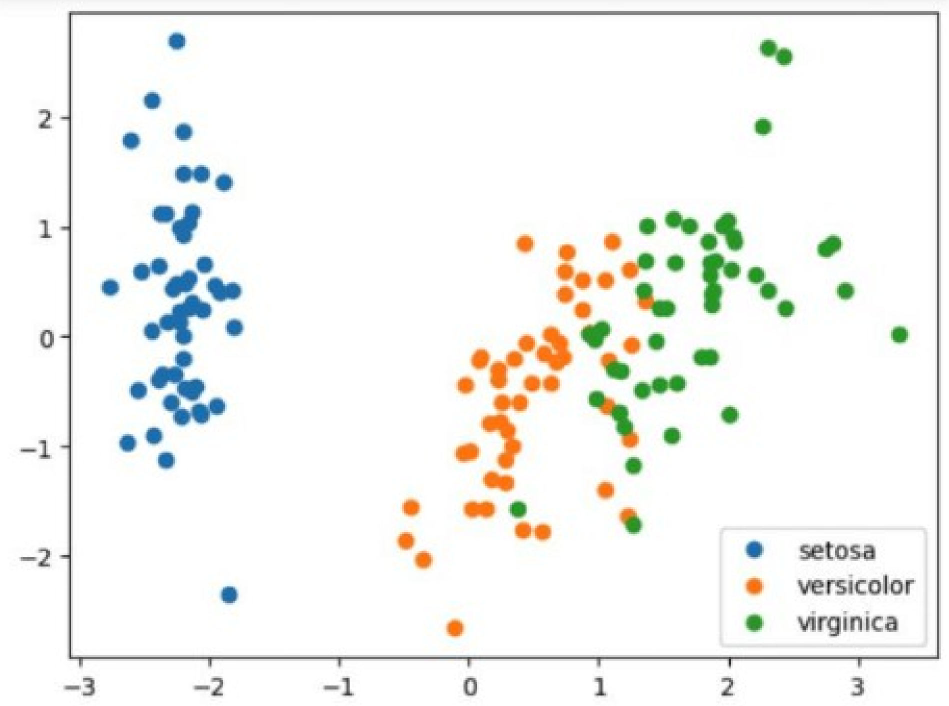



**NOTE:** If the scree plot looked like this, where PC3 and PC4 account for a substantial amount of variation, then just using the first 2 PCs would not create a very accurate representation of the data.

# Exemplo de PCA com o conjunto de dados da íris




Graph showing the variance explained by each PC




Scores plot showing coordinates of the different flowers in the PC1/ PC2. Colours represent species, **which were not used in the PCA computations.**

# Terminologia PCA

- **Componente principal:** Combinações lineares de variáveis originais que captam a variância máxima nos dados.

- **Vetor próprio:** Direção no da caraterística espaço que define o principal componentes.

- **Valor próprio:** Escalar indicando a quantidade de variância explicada pelo pelo seu vetor próprio correspondente.

- **Variância explicada:** Proporção da variância total dos dados explicada por cada componente principal.

- **Pontuação de carregamento:** Peso ou coeficiente atribuído a cada variável original na construção de componentes principais.

# Considerações finais da AdC

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

- **Relações não lineares:** A ACP pressupõe relações lineares entre variáveis, o que limita a sua eficácia na captação de padrões não lineares complexos; para esses casos, considerar técnicas não lineares como t-SNE ou UMAP.

- **Interpretabilidade:** A PCA cria novas combinações de características que podem ser difíceis de interpretar, tornando-a menos adequada quando é crucial manter a interpretabilidade das características individuais.

- **Dados esparsos: O** desempenho da PCA pode ser afetado por dados esparsos em que a maioria dos valores são zero ou estão em falta, o que leva a distorções na estrutura dos dados.

- **Excedentes:** A PCA é sensível a valores aberrantes, que podem distorcer os resultados; para conjuntos de dados com valores aberrantes, podem ser necessárias técnicas robustas de PCA para atenuar o seu impacto.

- **Escala dos dados:** A PCA é sensível à escala das características, por isso é importante padronizar ou normalizar os dados antes de aplicar a PCA para garantir que todas as características contribuem igualmente para a análise.

# Outras técnicas de redução da dimensionalidade

## Escalonamento multidimensional (MDS)

- **Objetivo:** Reduzir dados de elevada dimensão para um espaço de dimensão inferior, preservando as relações entre pares, permitindo a visualização e a análise.

- **Abordagens:**
  - **MDS métrico:** Procura preservar as distâncias reais entre os pontos de dados no espaço de dimensão inferior, utilizando frequentemente algoritmos de otimização como a descida do gradiente.
  - **MDS não métrico:** centra-se na preservação da ordem de classificação das distâncias e não nos seus valores absolutos, o que o torna adequado para dados ordinais ou não quantitativos.

# Outras técnicas de redução da dimensionalidade

- Computacionalmente intensivo para grandes conjuntos de dados e sensível à entrada.

# Outras técnicas de redução da dimensionalidade

## Incorporação de vizinhança estocástica distribuída t (t-SNE)

- **Objetivo:** Reduzir dados dados para a espaço de dimensão inferior espaço, dando ênfase às **relações locais entre os pontos de dados**.

- **Abordagem:**
  - O t-SNE utiliza uma abordagem de mapeamento **não linear** que visa preservar **as semelhanças locais** no espaço de alta dimensão, modelando-as como probabilidades condicionais.
  - Minimiza a divergência entre as probabilidades condicionais nos espaços de alta dimensão e de baixa dimensão utilizando a otimização de descida gradiente.

- **Características principais:**
  - Enfatiza a preservação de estruturas locais, tornando-o eficaz para visualizar agrupamentos e estruturas múltiplas nos dados.
  - Particularmente útil para explorar relações complexas e não lineares em conjuntos de dados de elevada dimensão.

# Outras técnicas de redução da dimensionalidade

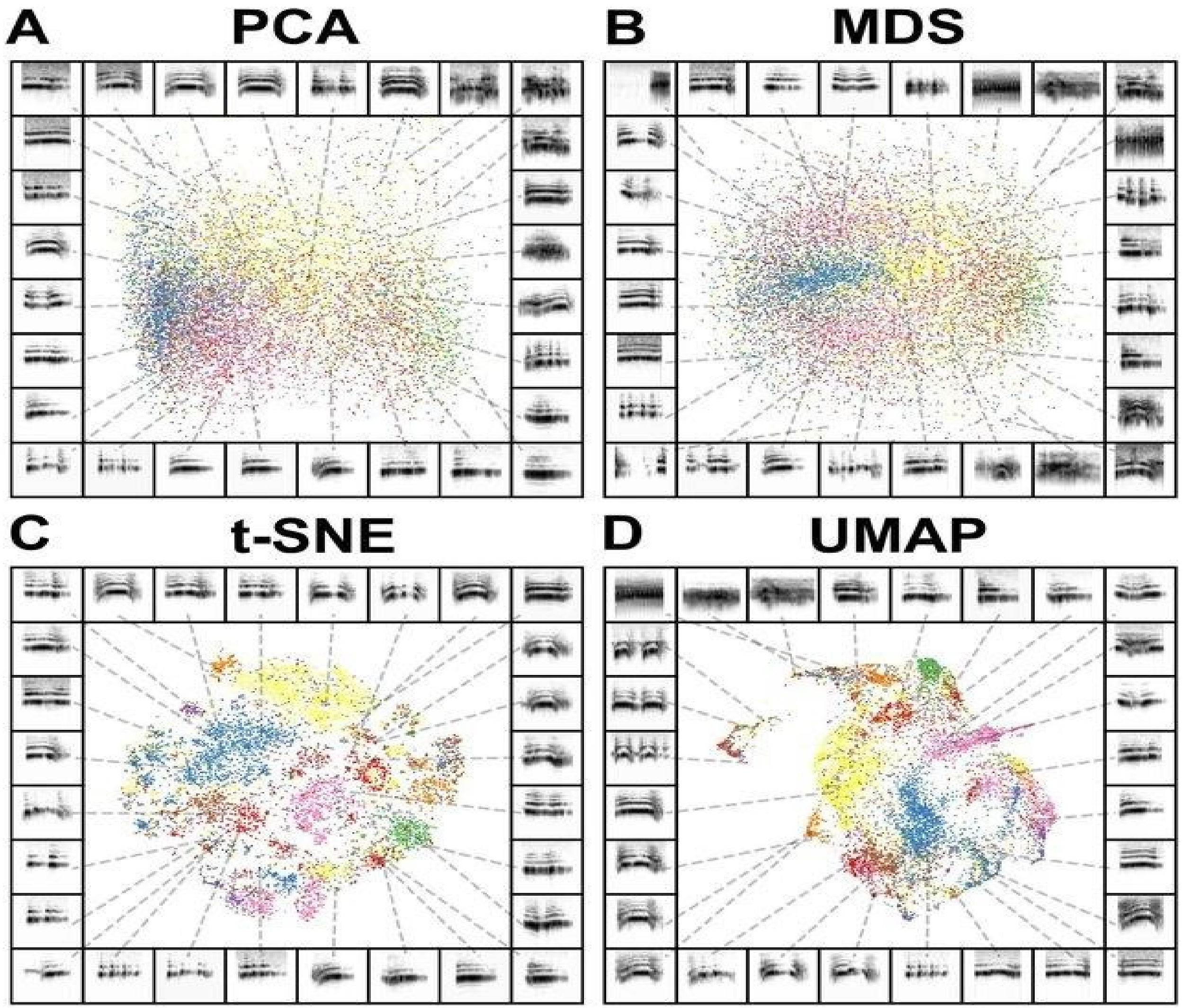
## Aproximação e Projeção Uniforme de Manifolds (UMAP)

- **Objetivo:** Reduzir a dimensionalidade de dados de elevada dimensão, preservando a estrutura local e global, oferecendo um equilíbrio entre a preservação de **detalhes locais e a captura de padrões globais**.

- **Abordagem:**
  - O UMAP constrói um gráfico de alta dimensão que representa as relações locais entre os pontos de dados e optimiza a incorporação num espaço de dimensão inferior para corresponder à topologia do gráfico.
  - Utiliza uma combinação da teoria dos conjuntos difusos e da geometria Riemanniana para modelar a estrutura múltipla dos dados.

- **Características principais:**
  - Preserva a estrutura local e global, permitindo uma representação mais abrangente dos dados.
  - Oferece flexibilidade no equilíbrio entre a preservação de detalhes locais e a captura de padrões globais através da afinação de parâmetros.
  - Conhecido pela sua escalabilidade e eficiência, tornando-o adequado para grandes conjuntos de dados.

# PCA vs MDS vs tSNE vs UMAP




Sainburg, T., Thielk, M., C Gentner, T. Q. (2020). Encontrar, visualizar e quantificar a estrutura latente em diversos repertórios vocais de animais. Em F. E. Theunissen (Ed.), PLOS Computational Biology (Vol. 16, Issue 10, p. e1008228). Biblioteca Pública da Ciência (PLoS). https://doi.org/10.1371/journal.pcbi.1008228

# Recursos

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

- Redução de dimensão: Uma visita guiada:

(https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/FnT_dimensionReduction.pdf)


- Oskolkov, N. (2022). Redução da dimensionalidade. Em Applied Data Science in Tourism (pp. 151-167). Springer International Publishing. https://doi.org/10.1007/978-3-030-88389-8_9