UNIVERSIDADE CATÓLICA PORTUGUESA

BRAGA

Machine Learning

Session 7 - T

Unsupervised Learning – Dimensionality Reduction

Degree in Applied Data Science

2024/2025

Dimensionality Reduction

- **Objective:** transforming high-dimensional data into a lower-dimensional representation, aiming to capture the essential patterns and relationships within the data while minimizing redundancy and noise.

- **Why reduce dimensions?**
  - Simplifies **analysis and visualization** of complex data;
  - Reduces **computational complexity** and memory requirements;
  - Helps mitigate the **curse of dimensionality**;
  - Improves **model performance and generalization**;
  - Enhances **interpretability** and understanding of underlying data patterns;

Dimensionality Reduction

- If we only have one feature, we can easily plot the data on a number line.
- Even with this simple graph, we can see diferences between samples 1, 2 and 3 and 4, 5 and 6.

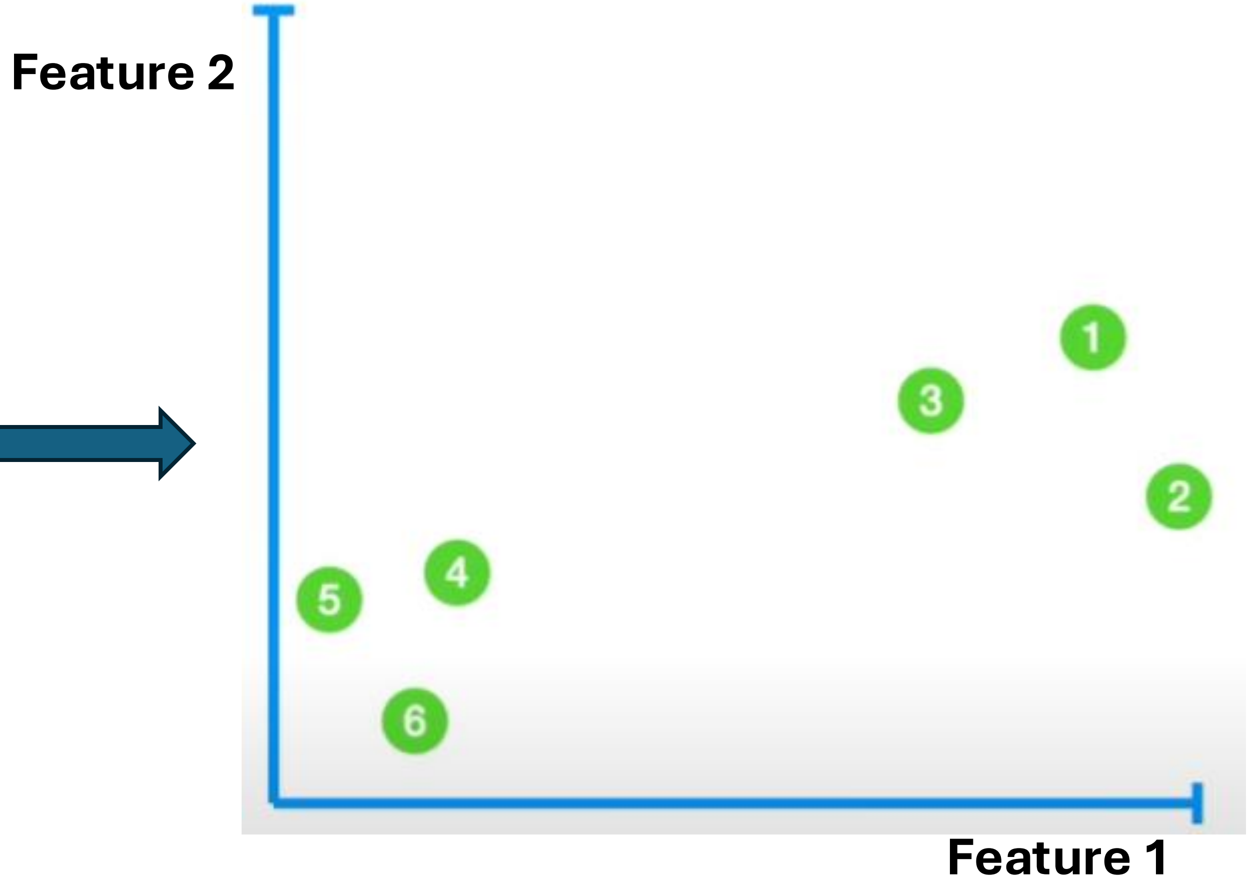
	Feature 1
Sample 1	10
Sample 2	11
Sample 3	8
Sample 4	3
Sample 5	1
Sample 6	2

Feature 1

Dimensionality Reduction

- Now we can plot the data on a 2-Dimensional graph.

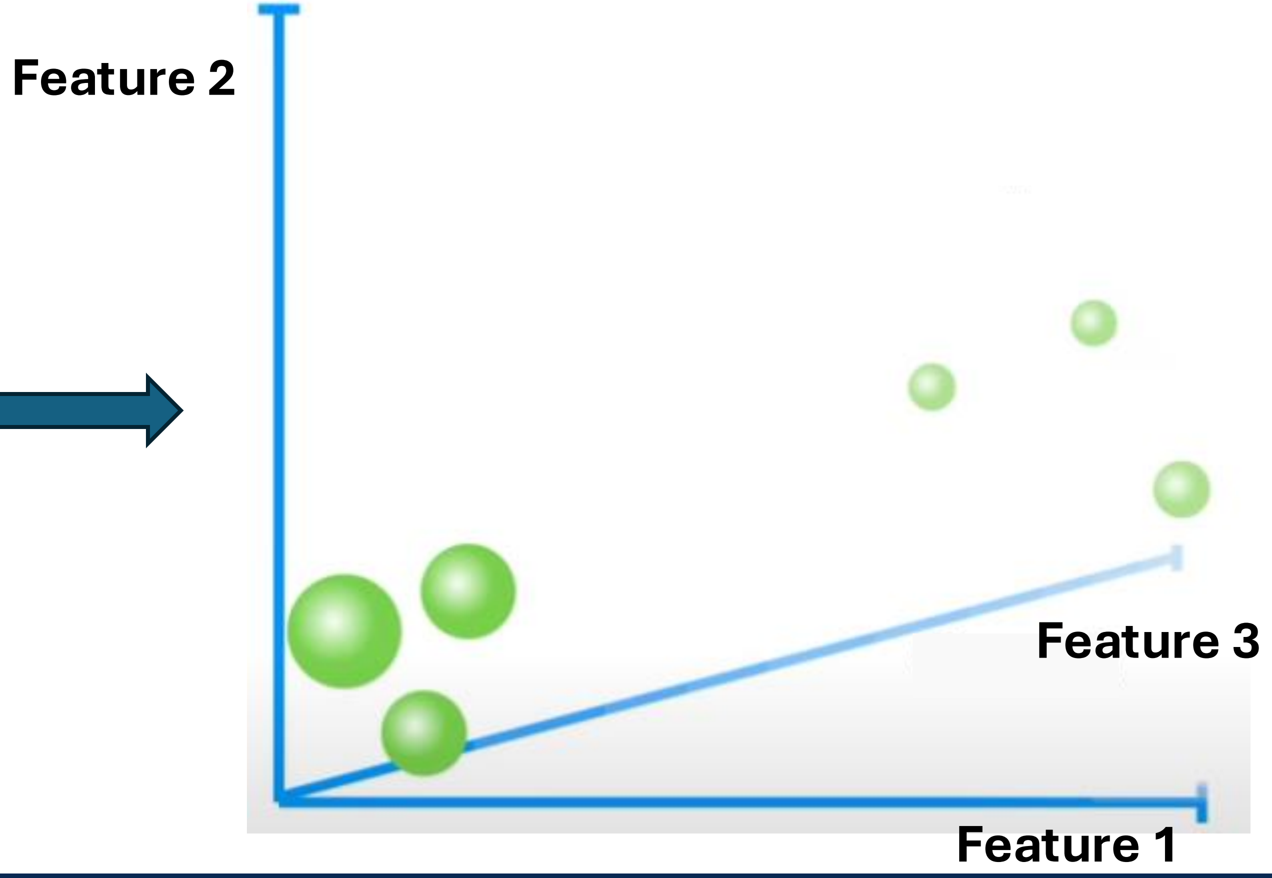
	Feature 1	Feature 2
Sample 1	10	6
Sample 2	11	4
Sample 3	8	5
Sample 4	3	3
Sample 5	1	2.8
Sample 6	2	1



Dimensionality Reduction

- Now we can plot the data on a 3-Dimensional graph.

	Feature 1	Feature 2	Feature 3
Sample 1	10	6	12
Sample 2	11	4	9
Sample 3	8	5	10
Sample 4	3	3	2.5
Sample 5	1	2.8	1.3
Sample 6	2	1	2



Dimensionality Reduction

- What about 4 or more dimensions?

	Feature 1	Feature 2	Feature 3	Feature 4	...
Sample 1	10	6	12	5	...
Sample 2	11	4	9	7	...
Sample 3	8	5	10	6	...
Sample 4	3	3	2.5	2	...
Sample 5	1	2.8	1.3	4	...
Sample 6	2	1	2	7	...

→ ?

Principal Component Analysis (PCA)

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

- It identifies the **principal components**, which are new variables that **capture the most variance in the data**;

- These **components are orthogonal** (uncorrelated) to each other, allowing for efficient reduction of dimensions;

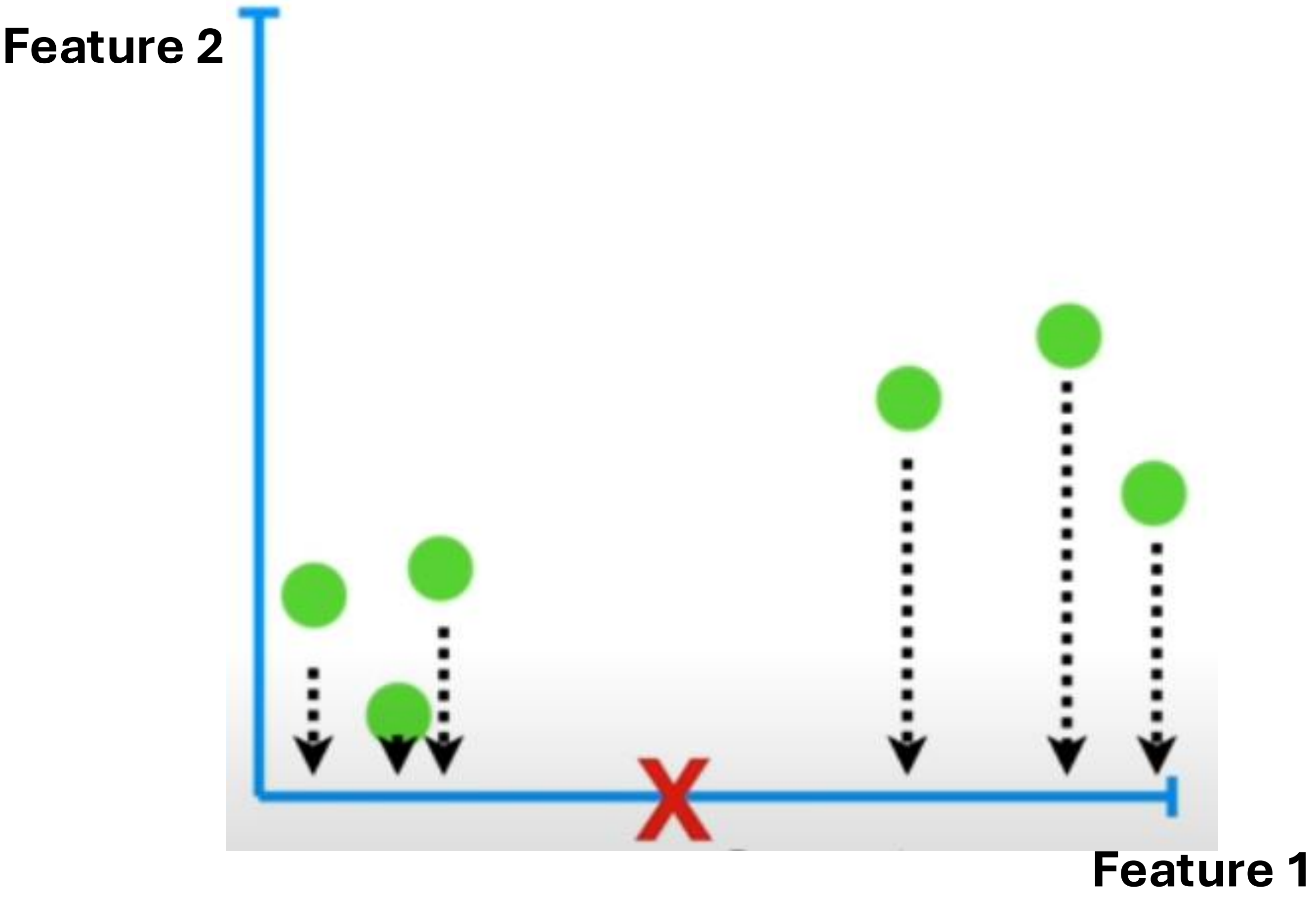
- The **first principal component explains the maximum amount of variance** in the data, followed by subsequent components in descending order of variance explained.

PCA Step by Step

- Let's start with a simple example with 2 features.

1.) Mean center the data;

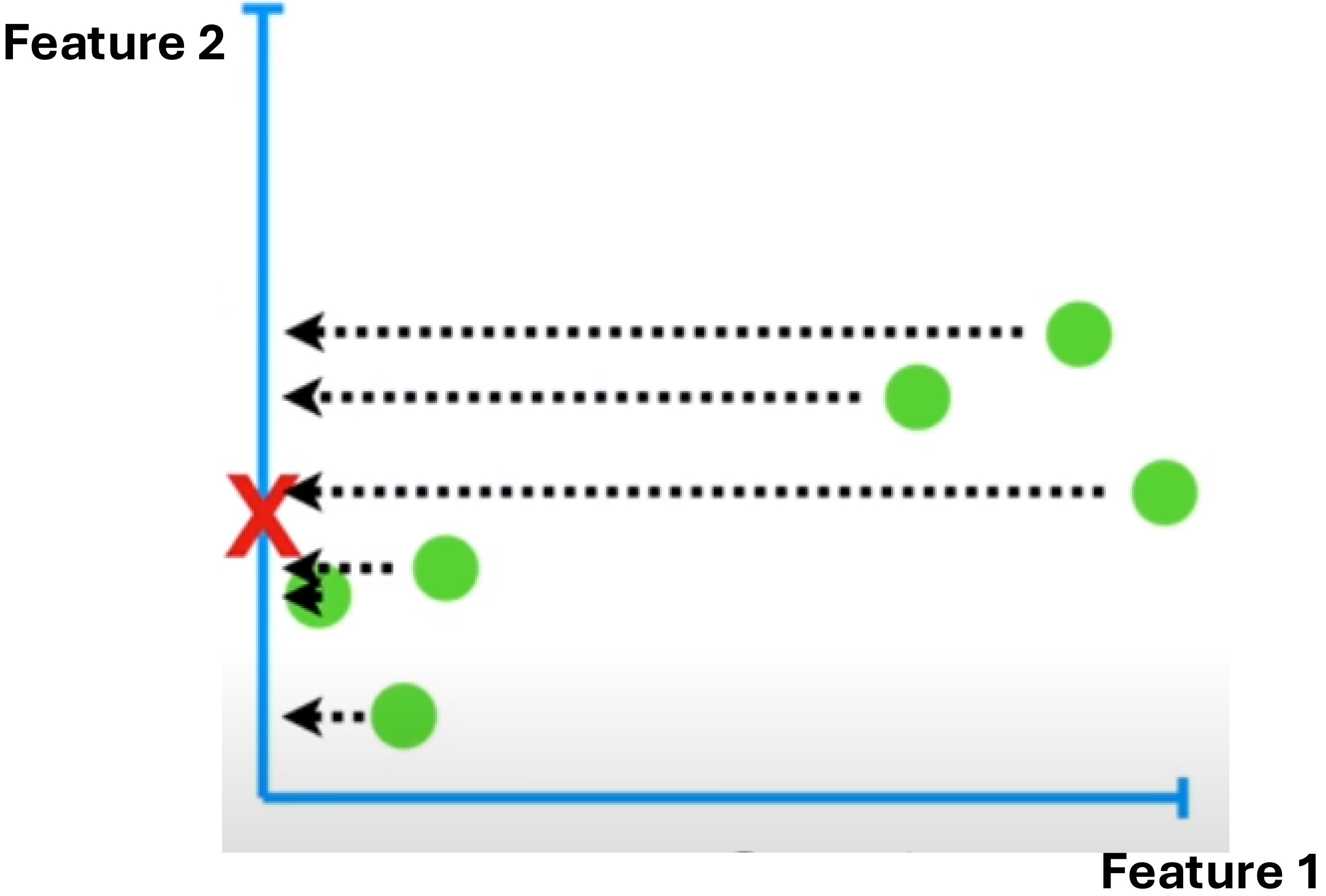
	Feature 1	Feature 2
Sample 1	10	6
Sample 2	11	4
Sample 3	8	5
Sample 4	3	3
Sample 5	1	2.8
Sample 6	2	1



PCA Step by Step

1.) Mean center the data;

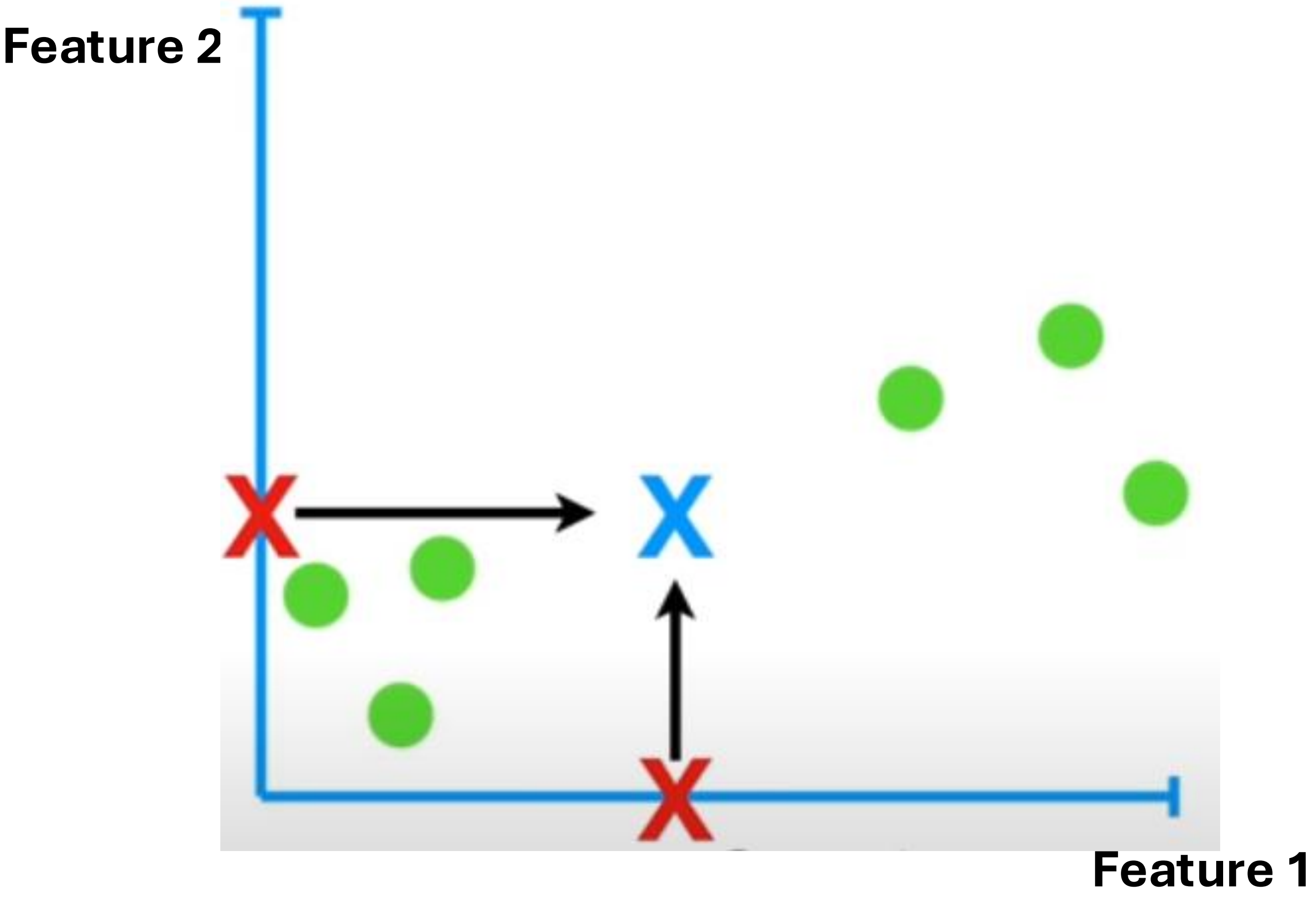
	Feature 1	Feature 2
Sample 1	10	6
Sample 2	11	4
Sample 3	8	5
Sample 4	3	3
Sample 5	1	2.8
Sample 6	2	1



PCA Step by Step

1.) Mean center the data;

	Feature 1	Feature 2
Sample 1	10	6
Sample 2	11	4
Sample 3	8	5
Sample 4	3	3
Sample 5	1	2.8
Sample 6	2	1

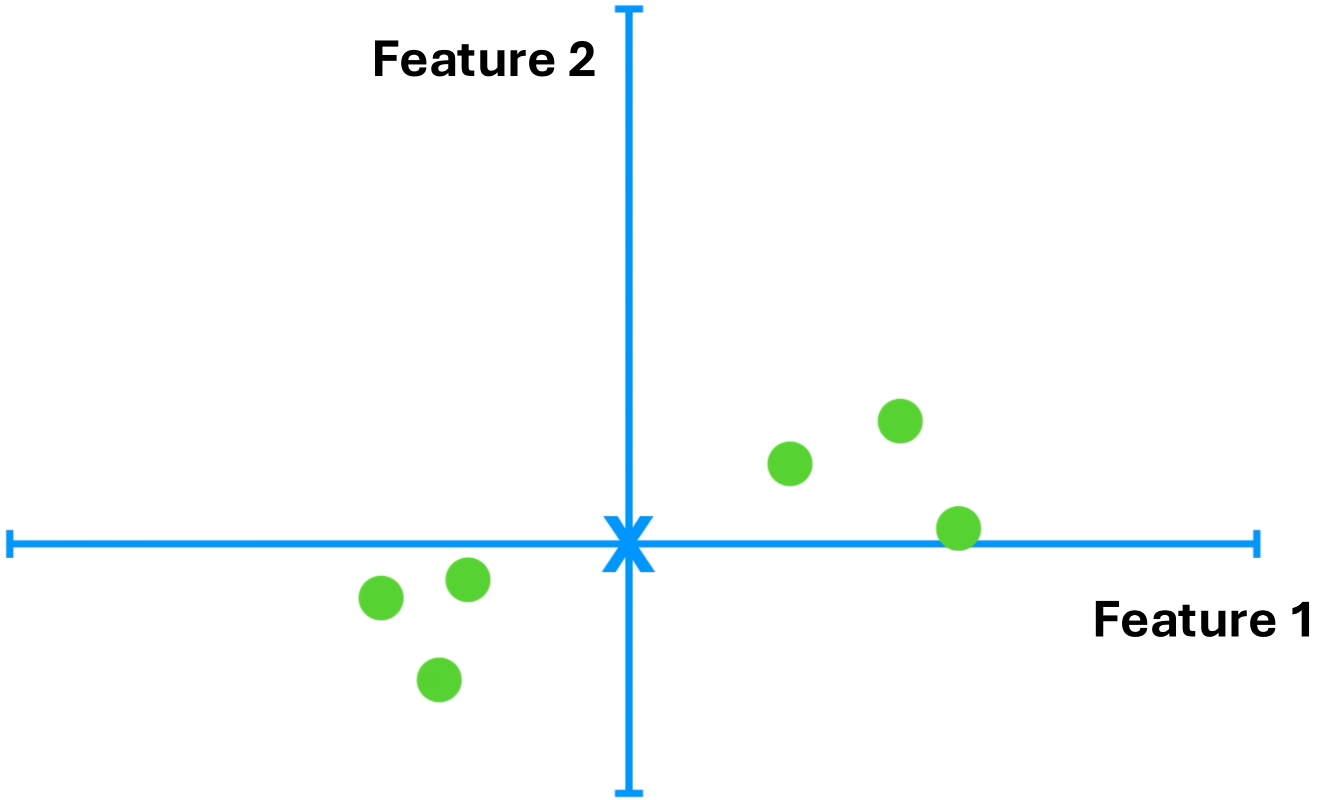


PCA Step by Step

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

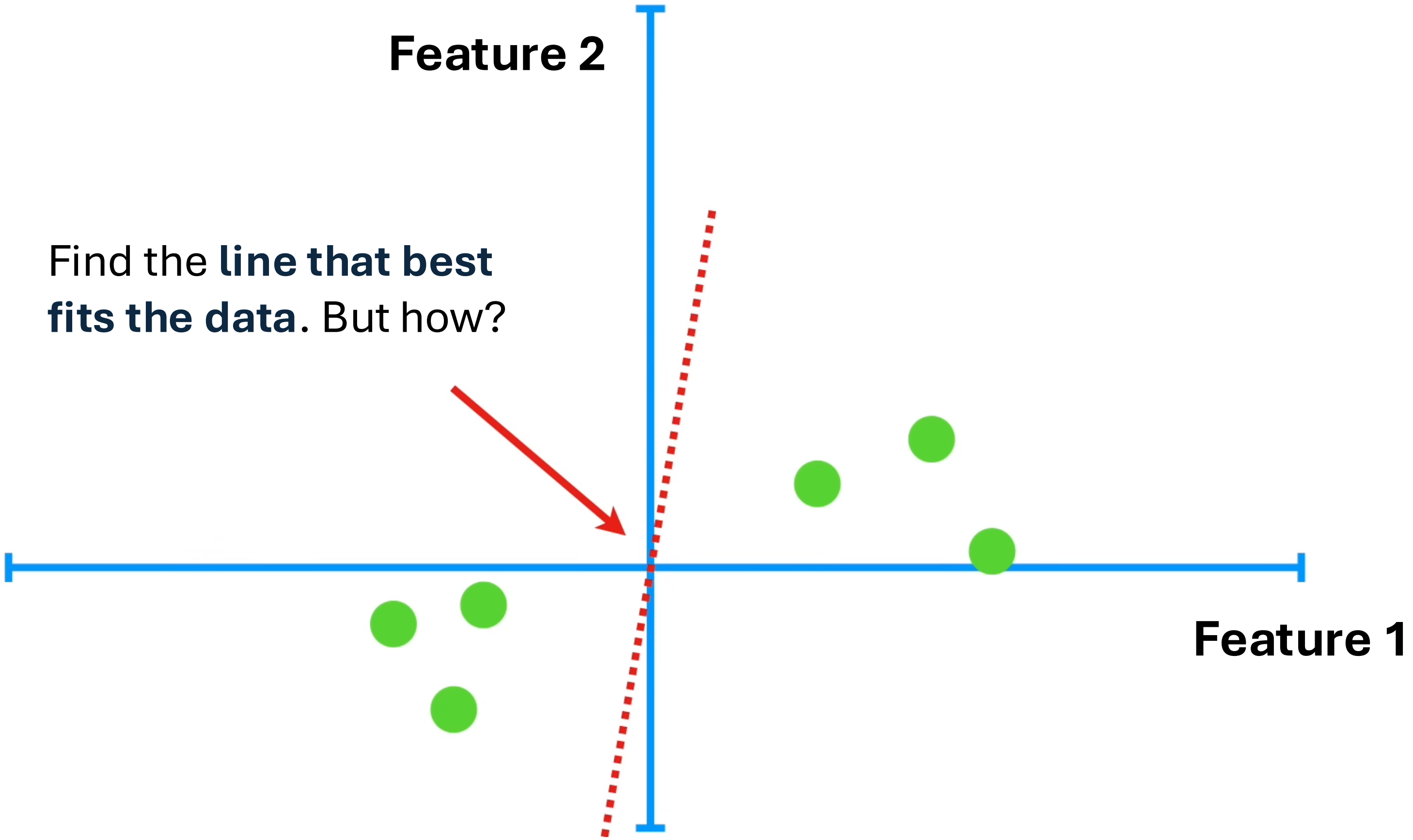
1.) Mean center the data;

	Feature 1	Feature 2
Sample 1	10	6
Sample 2	11	4
Sample 3	8	5
Sample 4	3	3
Sample 5	1	2.8
Sample 6	2	1



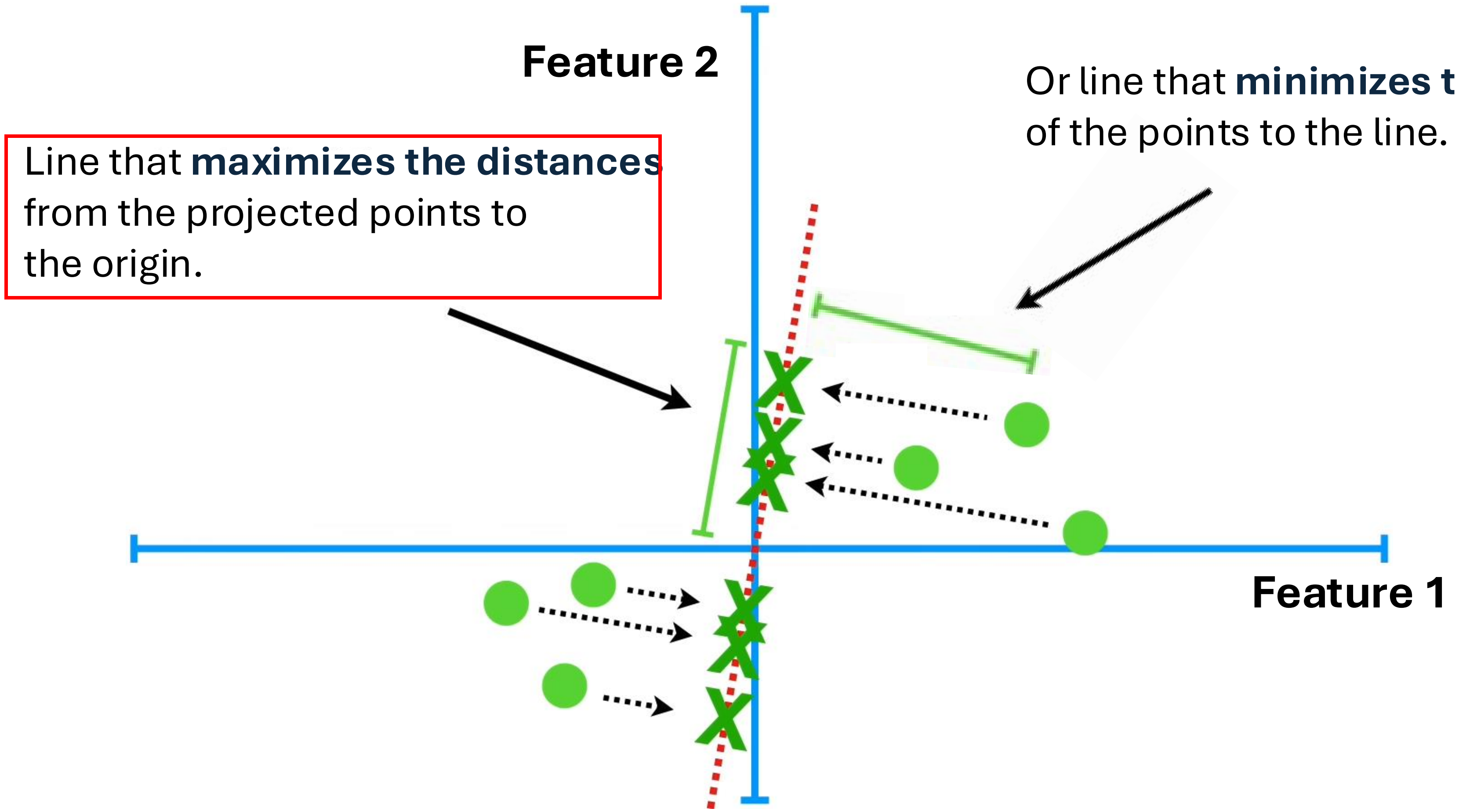
PCA Step by Step

2.) Calculate PC1 using singular value decomposition (SVD).



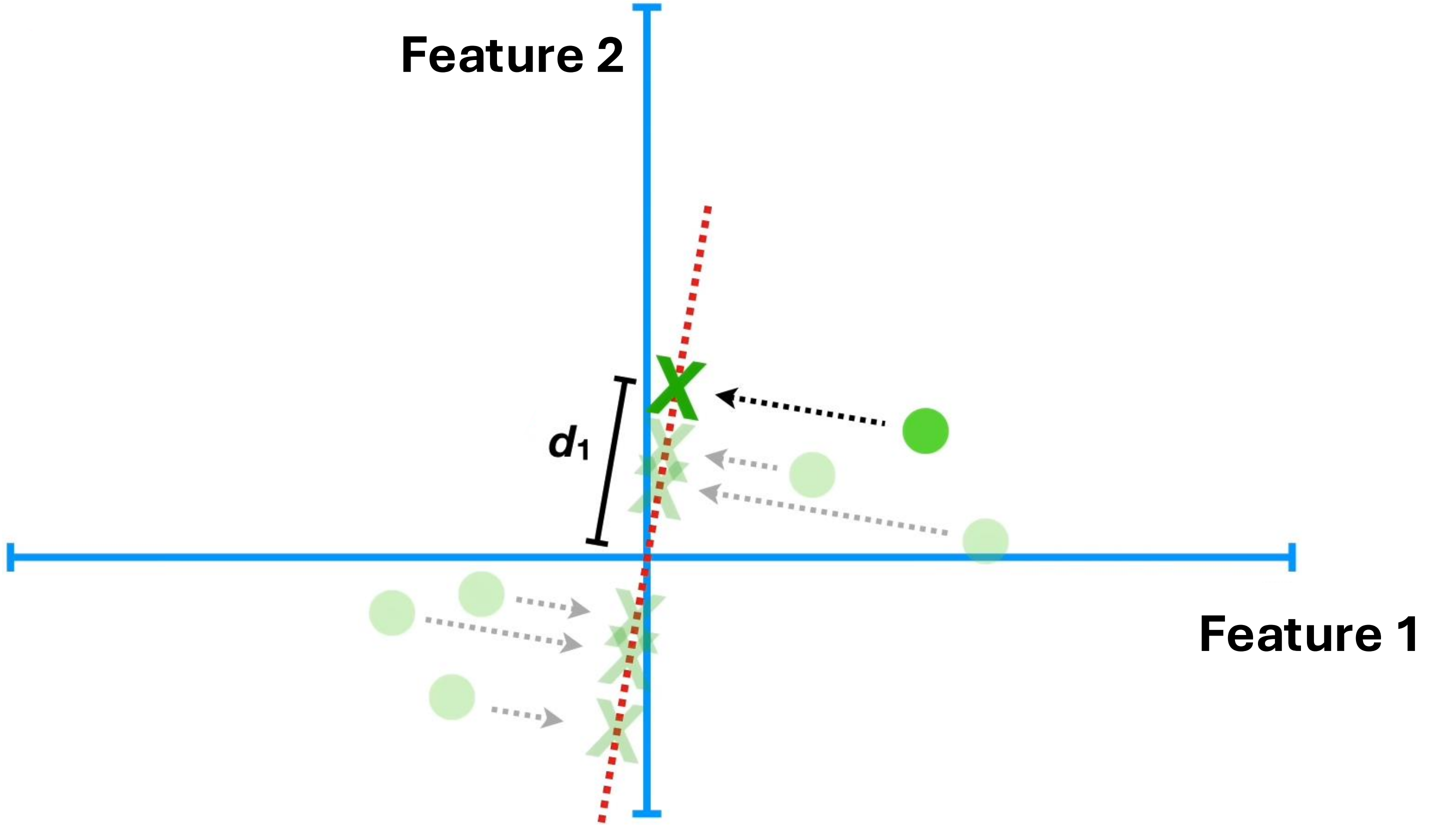
PCA Step by Step

2.) Calculate PC1 using singular value decomposition (SVD).



PCA Step by Step

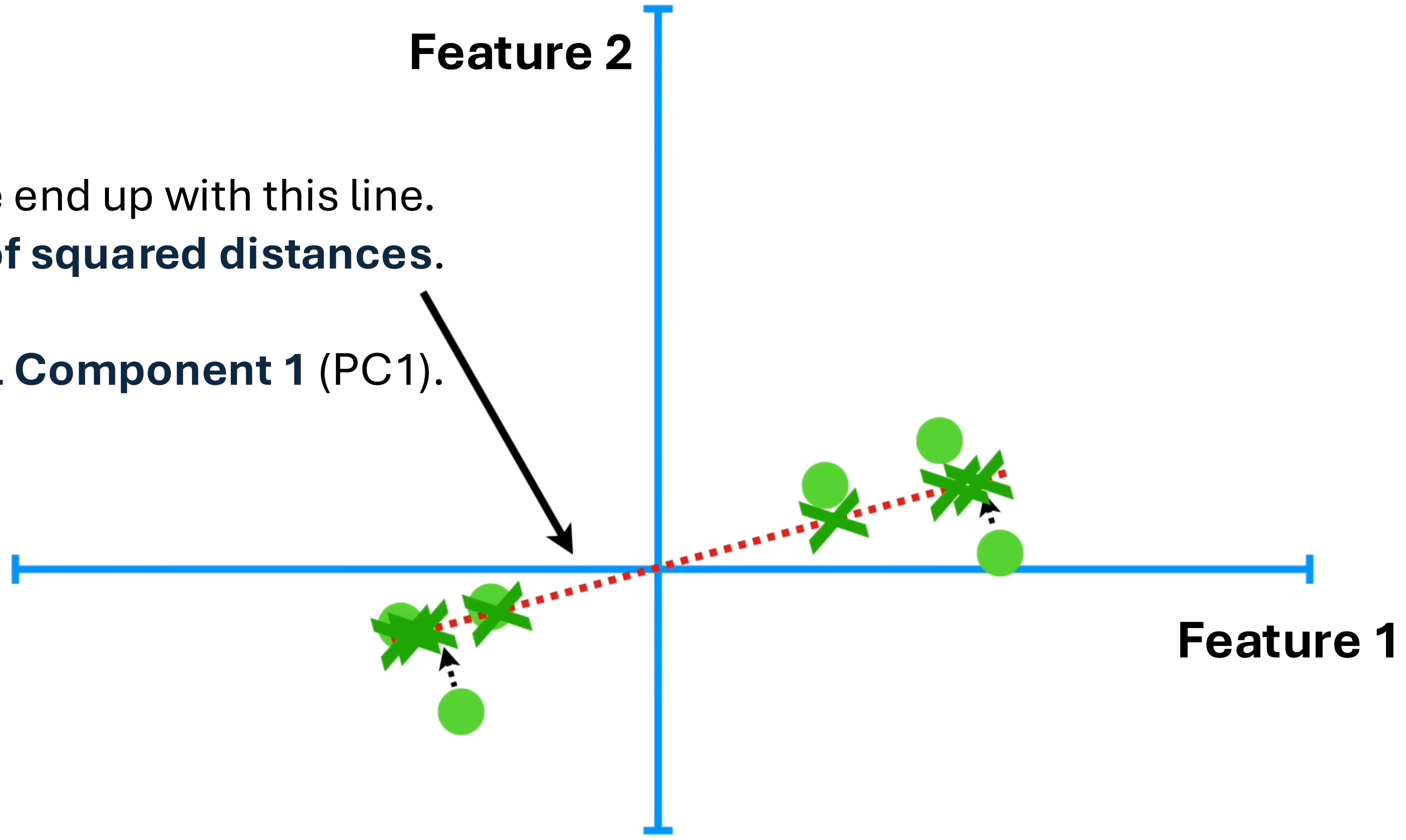
2.) Calculate PC1 using singular value decomposition (SVD).



$d1^2 + d2^2 + d3^2 + d4^2 + d5^2 + d6^2$ = **sum of squared distances**

PCA Step by Step

2.) Calculate PC1 using singular value decomposition (SVD).



We end up with this line.
It has the **largest sum of squared distances**.

This line is called **Principal Component 1** (PC1).

$d1^2 + d2^2 + d3^2 + d4^2 + d5^2 + d6^2$ = sum of squared distances

The average of the sum of squared distances for PC1 is named **Eigenvalue** for PC1.

PCA Step by Step

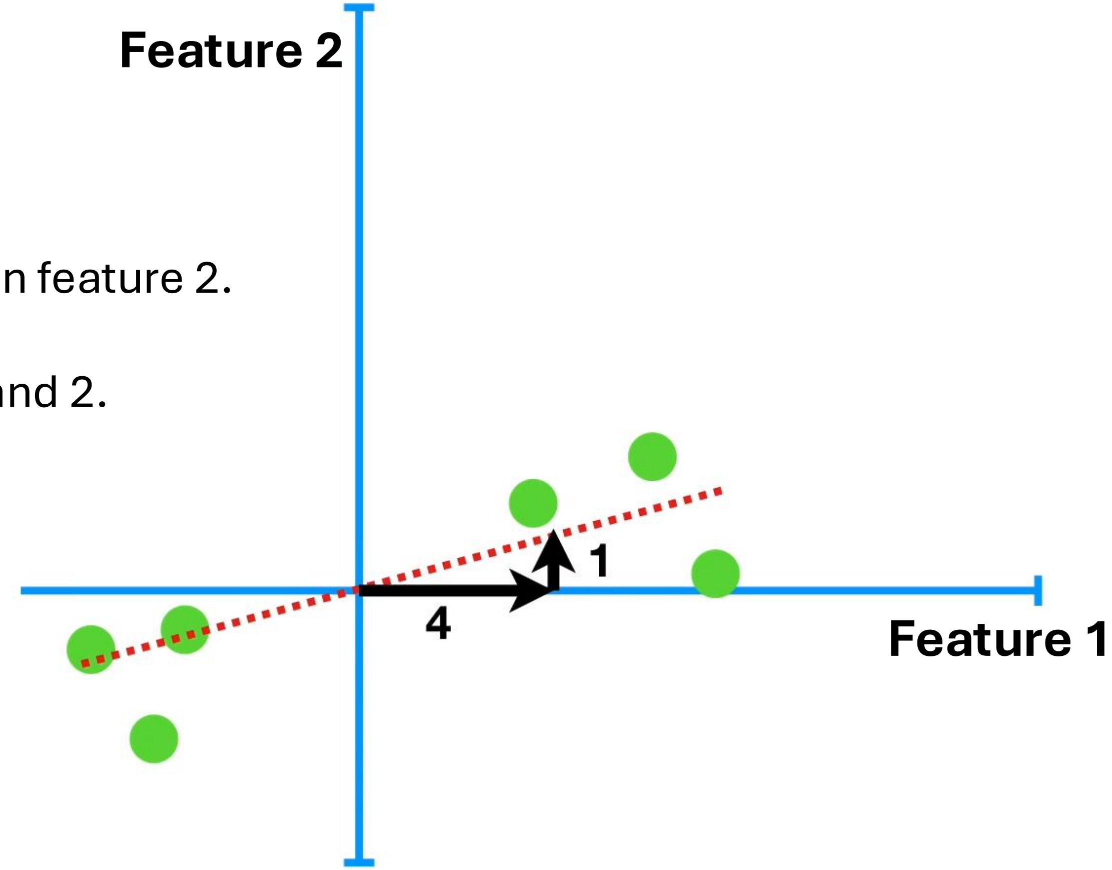
UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

2.) Calculate PC1 using singular value decomposition (SVD).

Let's say that the slope of the line is 0.25.

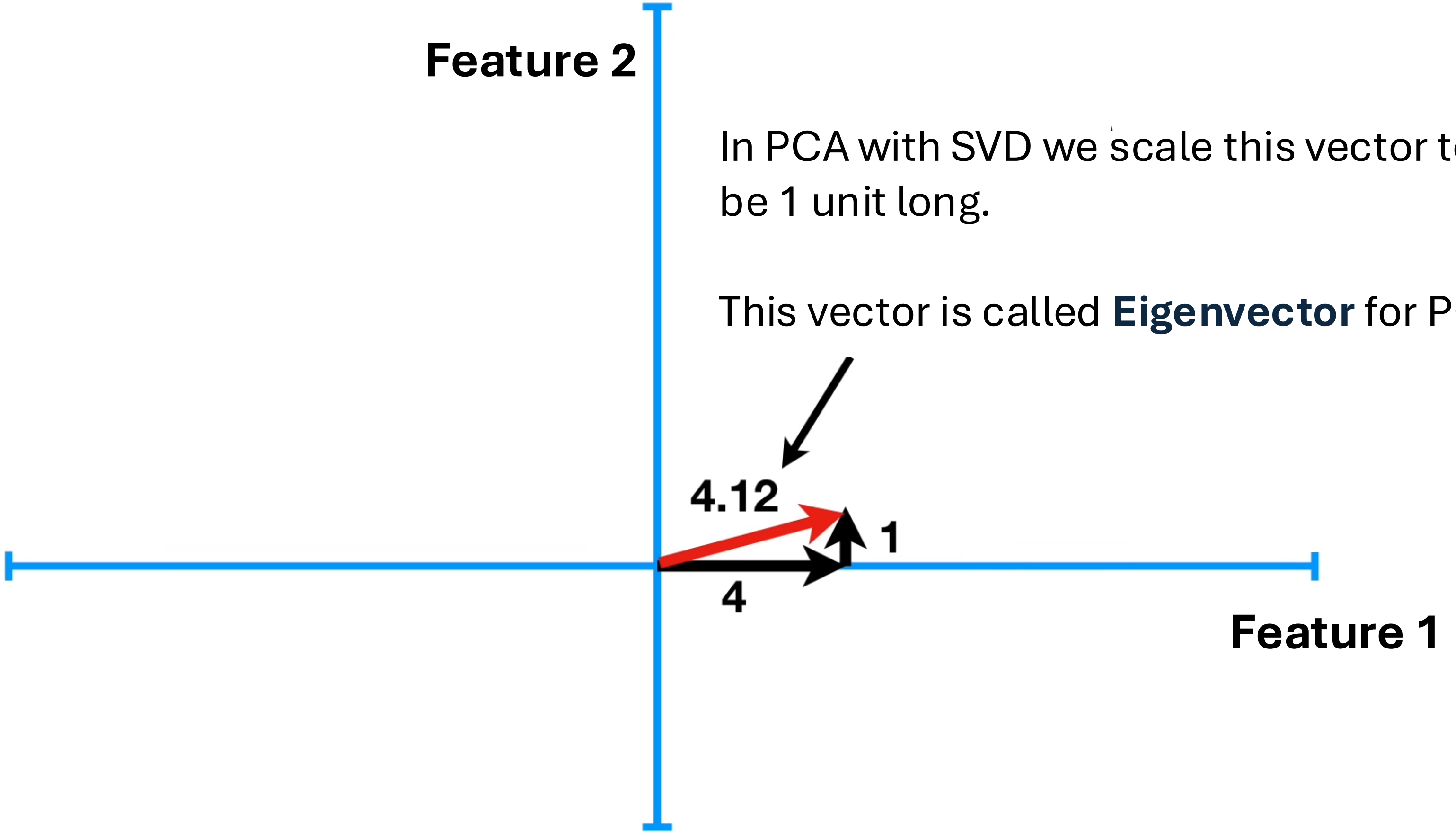
For 4 units of Feature 1 we increase 1 unit in feature 2.

PC1 is a **linear combination** of Feature 1 and 2.



PCA Step by Step

2.) Calculate PC1 using singular value decomposition (SVD).



In PCA with SVD we scale this vector to be 1 unit long.

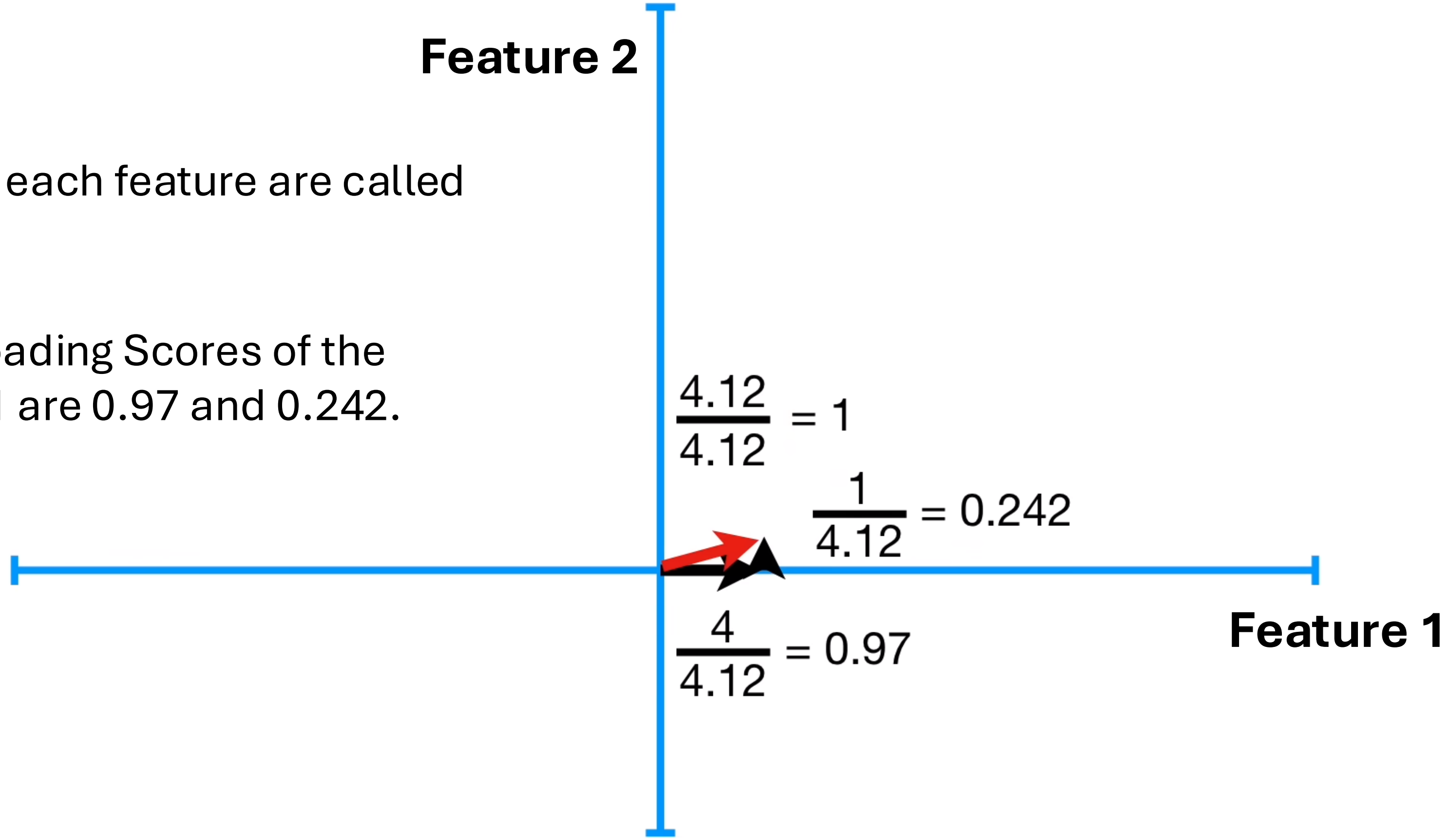
This vector is called **Eigenvector** for PC1.

PCA Step by Step

2.) Calculate PC1 using singular value decomposition (SVD).

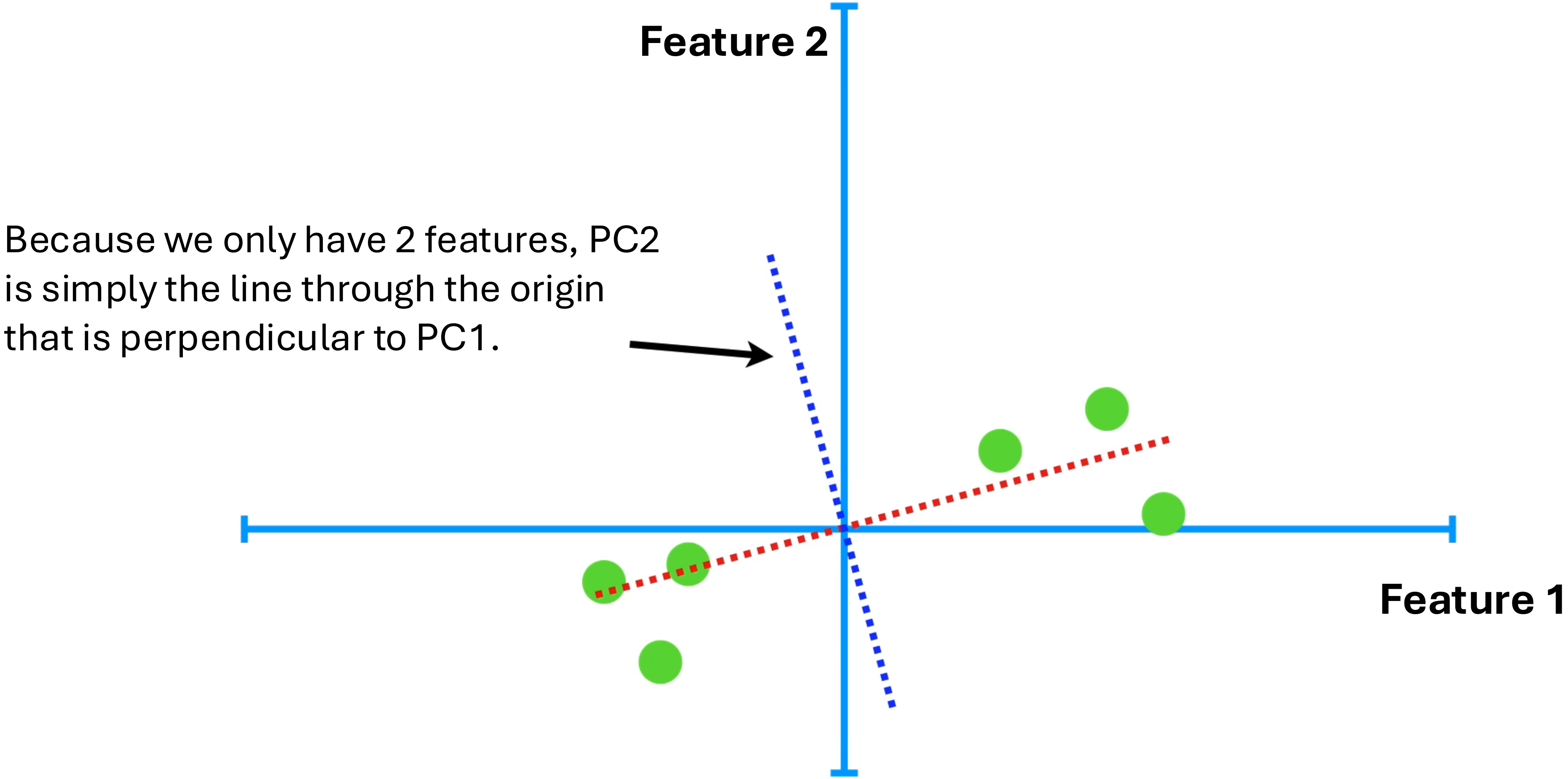
The proportions of each feature are called **Loading Scores**.

In this case, the Loading Scores of the Eigenvector of PC1 are 0.97 and 0.242.



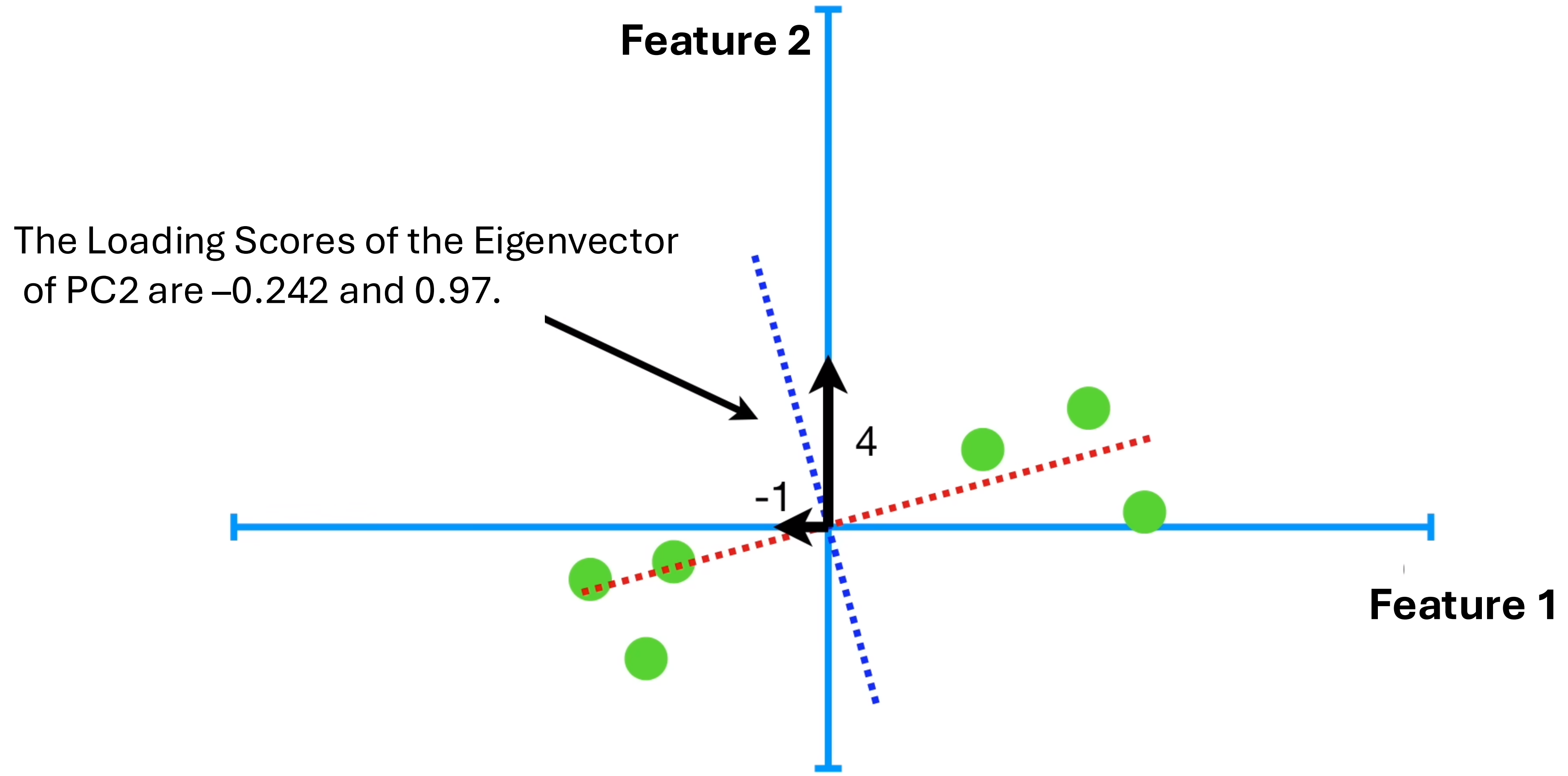
PCA Step by Step

3.) Calculate PC2 using singular value decomposition (SVD).



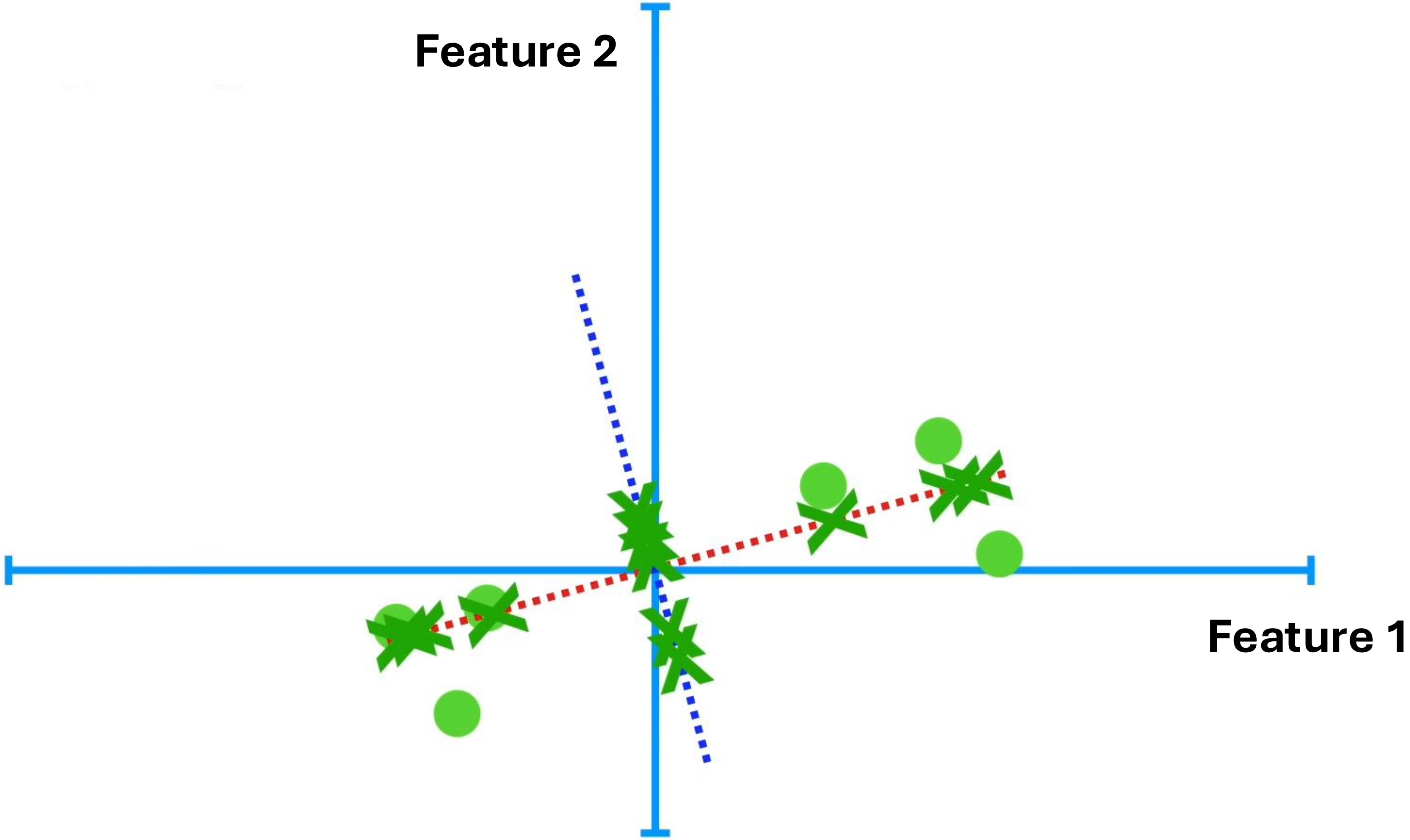
PCA Step by Step

3.) Calculate PC2 using singular value decomposition (SVD).



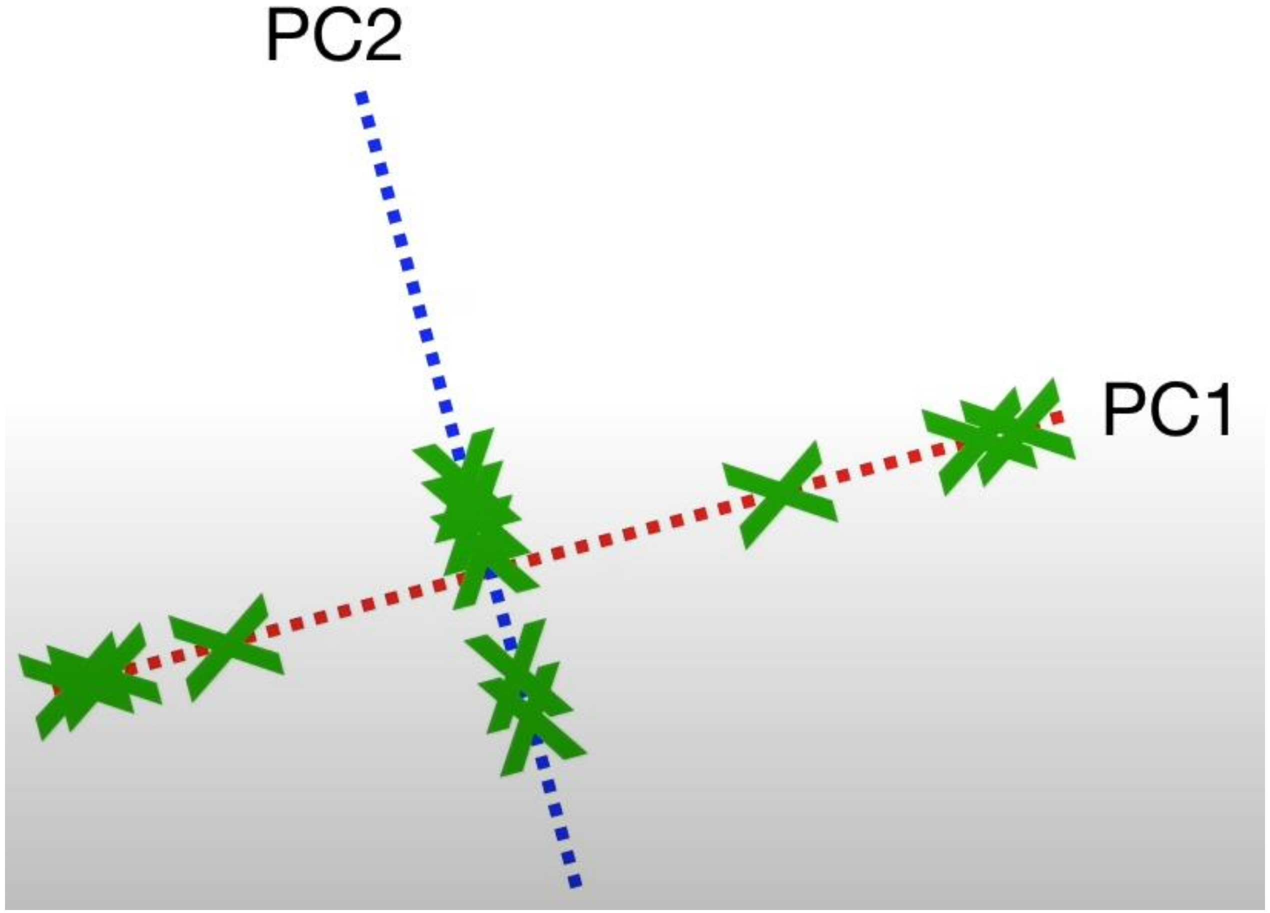
PCA Step by Step

4.) Get the final coordinates along the PCs.



PCA Step by Step

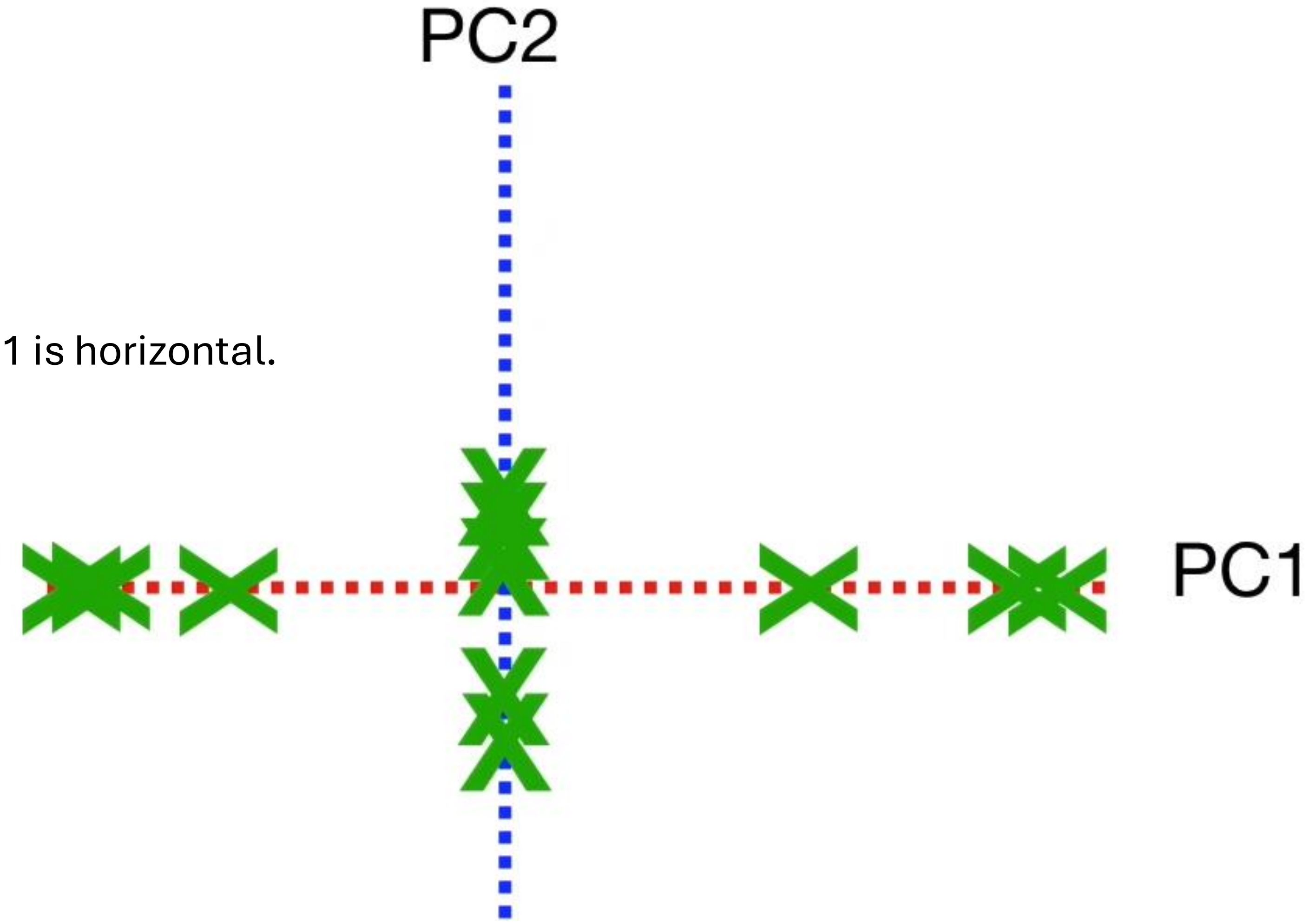
4.) Get the final coordinates along the PCs.



PCA Step by Step

4.) Get the final coordinates along the PCs.

Rotate the graph so PC1 is horizontal.

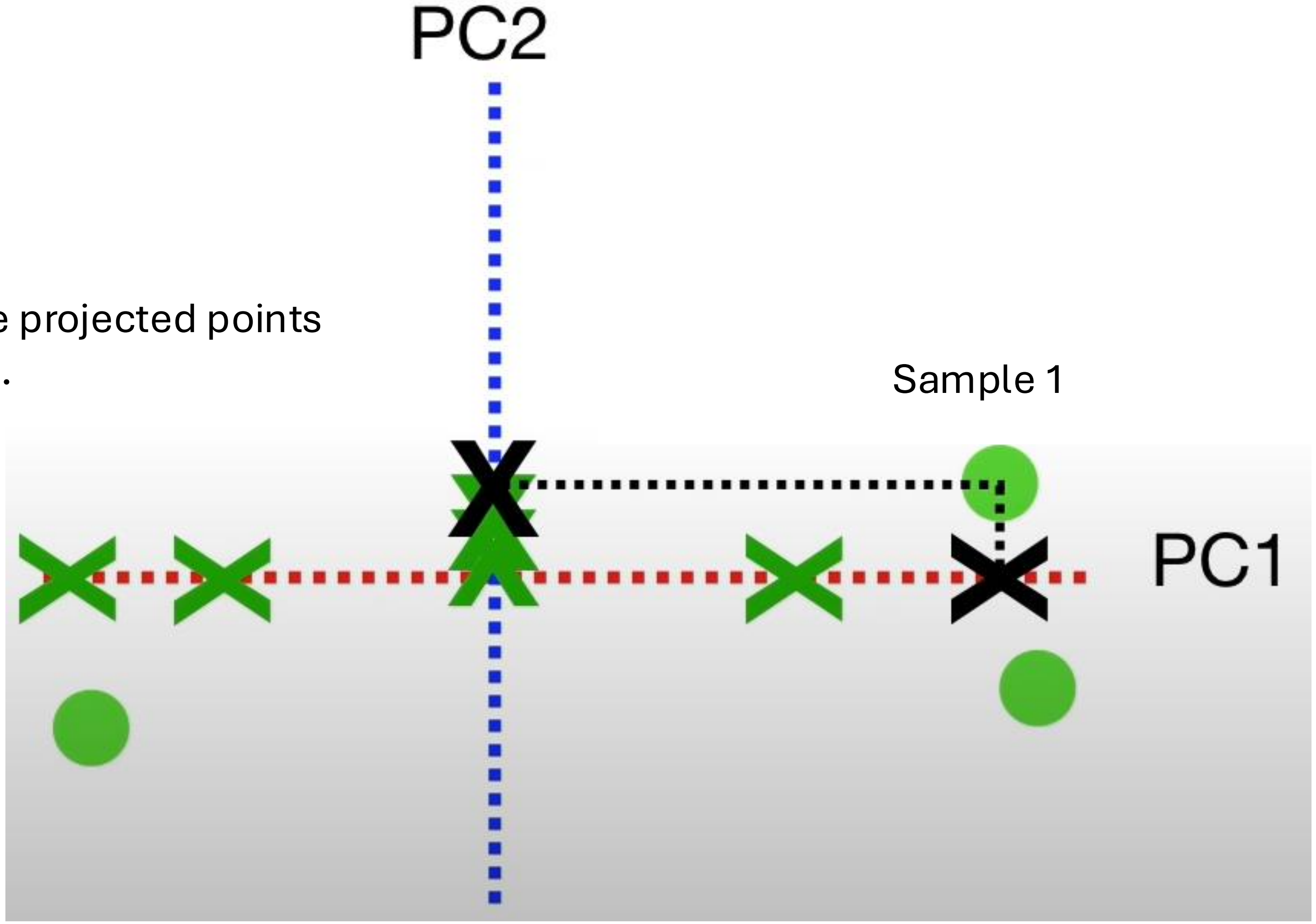


PCA Step by Step

4.) Get the final coordinates along the PCs.

We can now map the projected points bach to the samples.

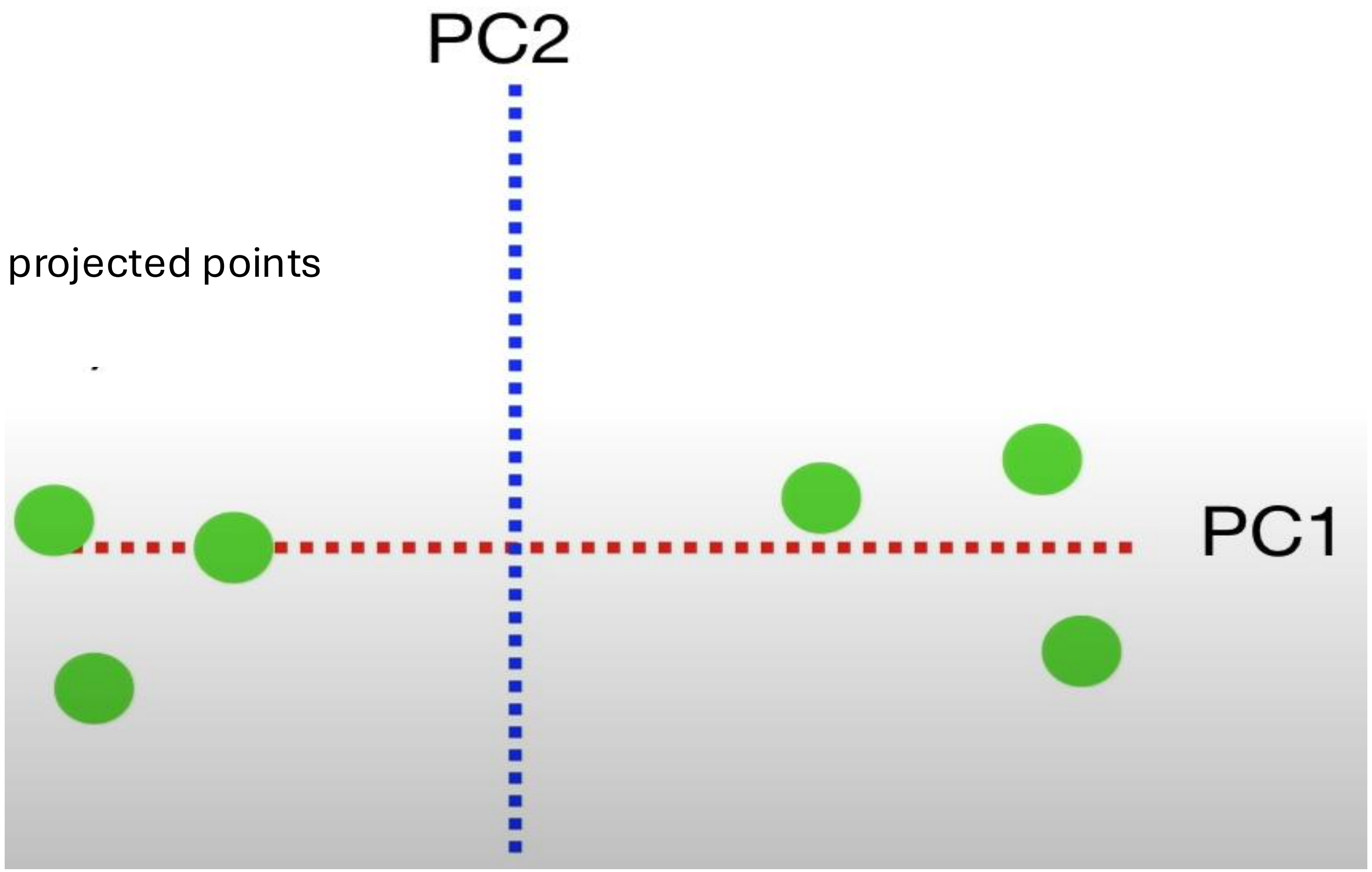
Sample 1



PCA Step by Step

4.) Get the final coordinates along the PCs.

We can now map the projected points bach to the samples.



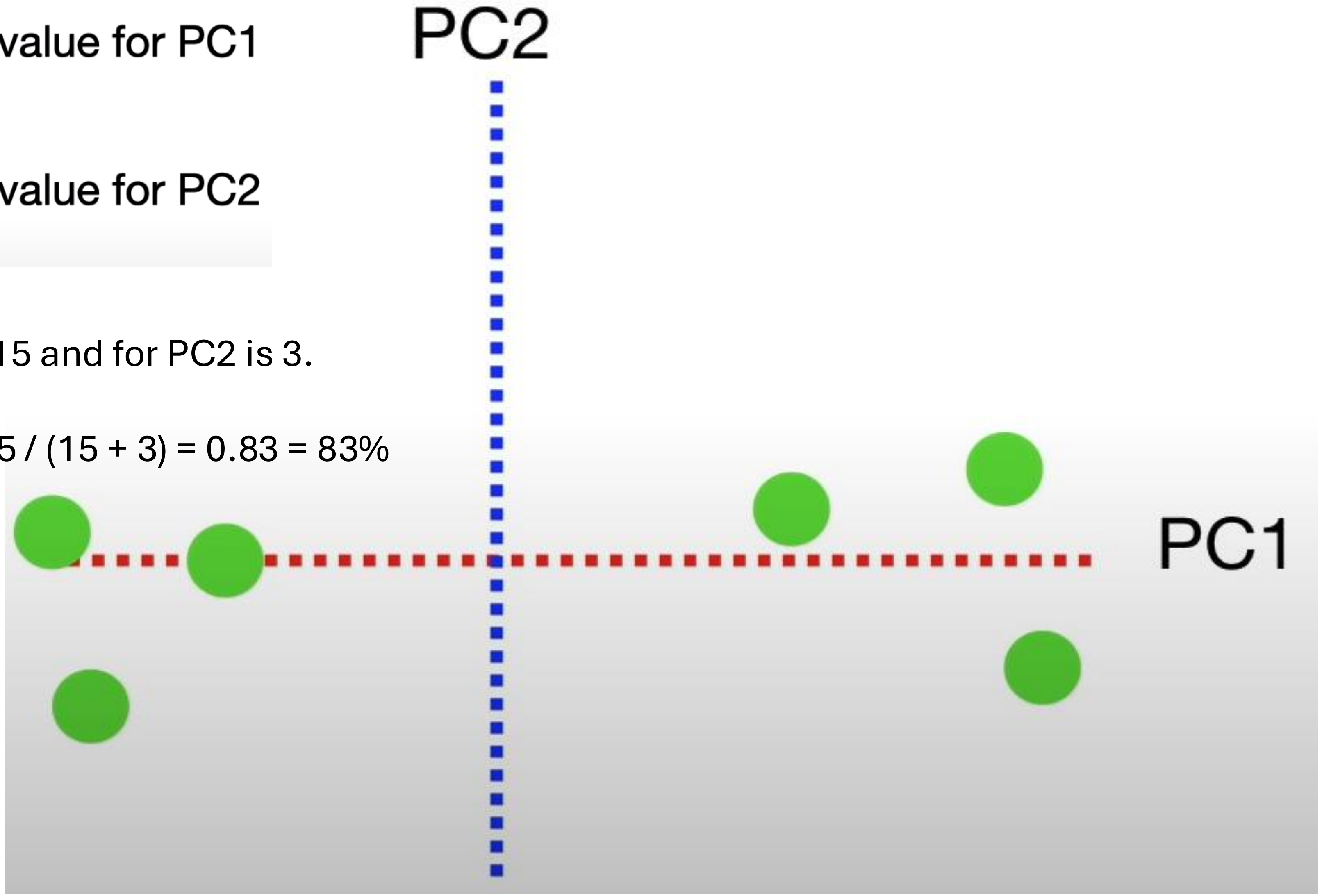
PCA – Explained Variance

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

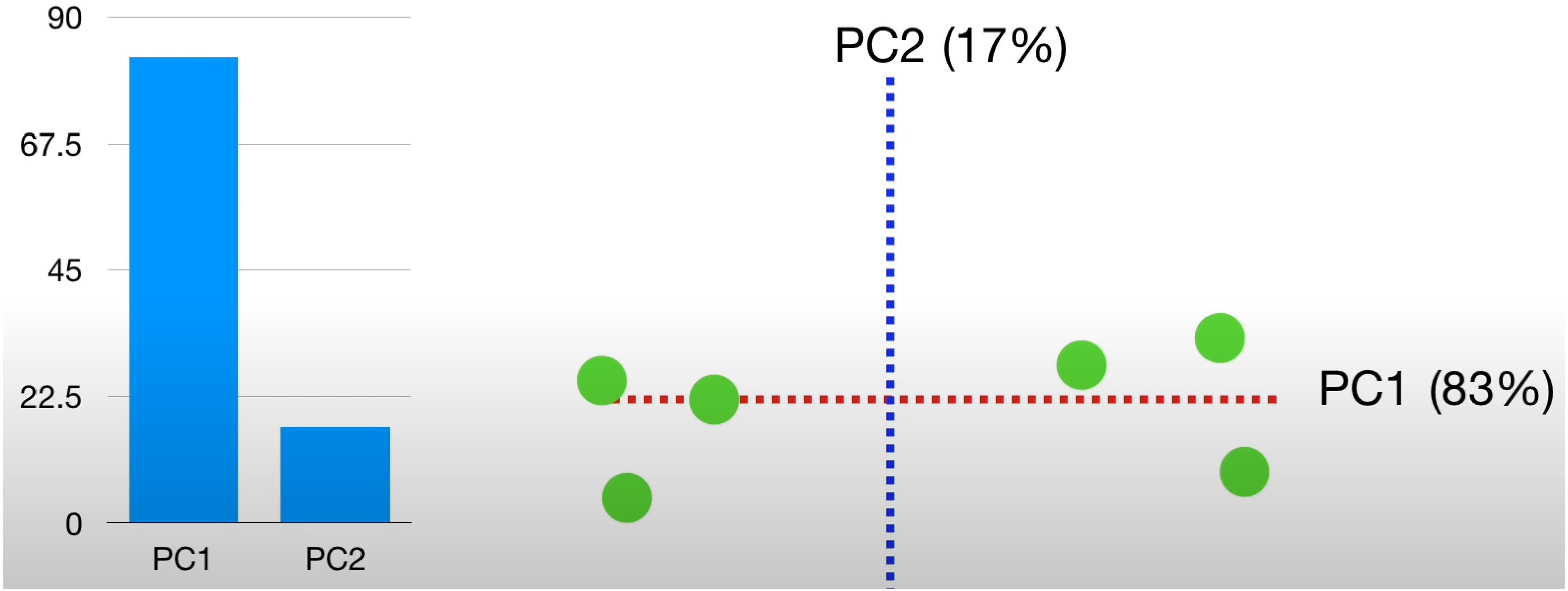
$$\frac{SS(\text{distances for PC2})}{n - 1} = \text{Eigenvalue for PC2}$$

Let's say the Eigenvalue for PC1 is 15 and for PC2 is 3.

The **explained variance** of PC1 = 15 / (15 + 3) = 0.83 = 83%
For PC2 = 3 / (15 + 3) = 0.17 = 17%



PCA – Explained Variance – Scree Plot

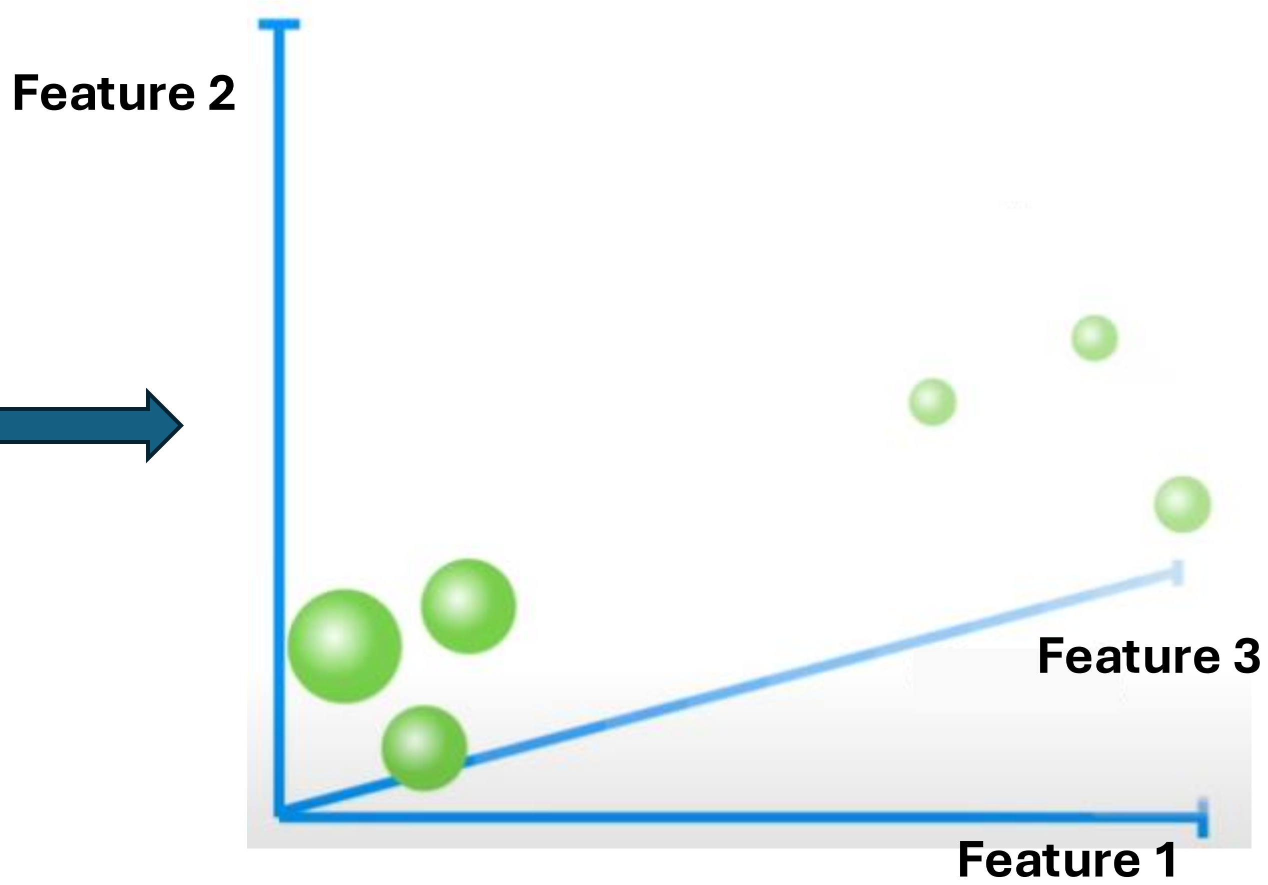


PCA with more Features

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

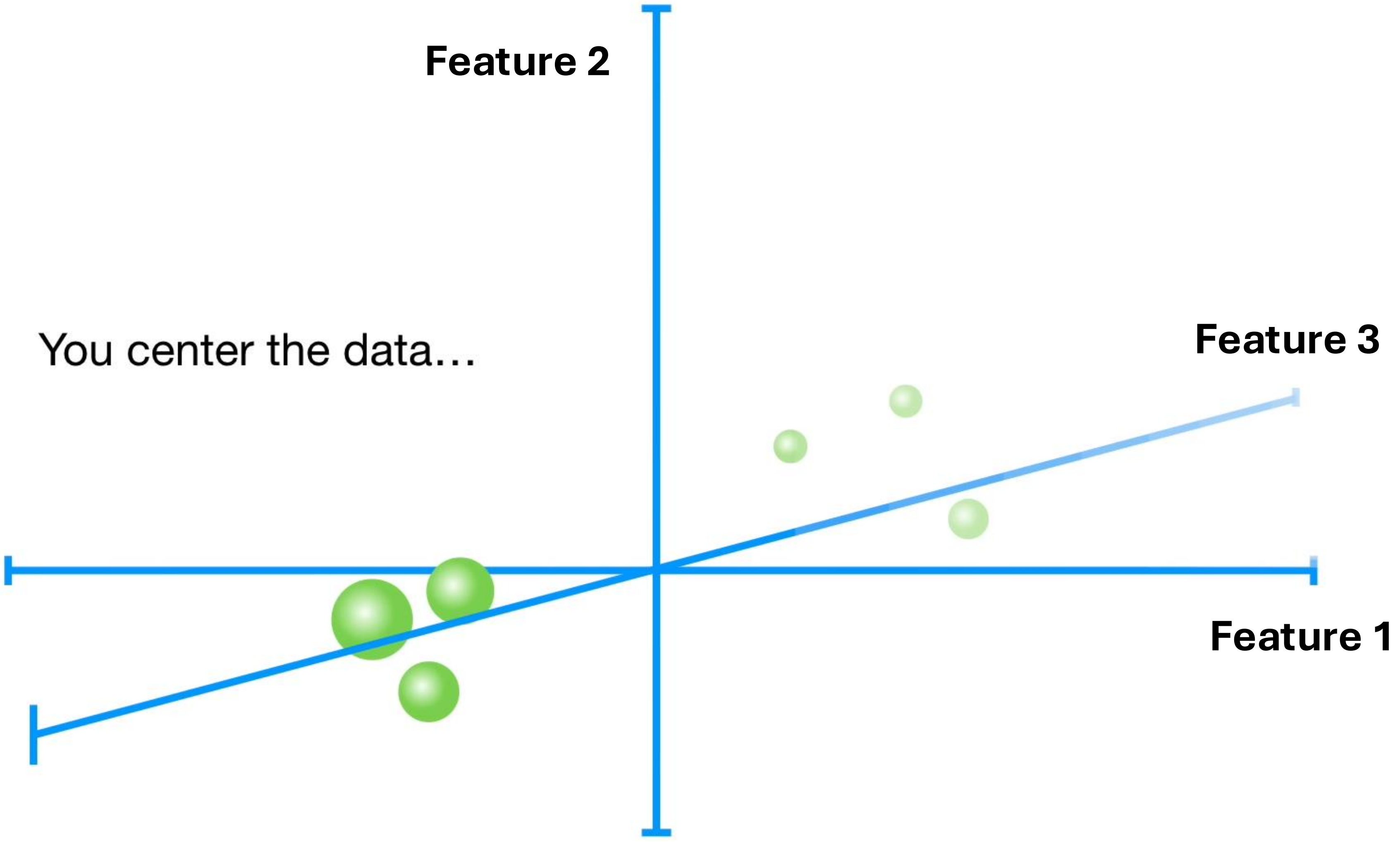
- PCA with 3 or more features is pretty much the same as 2 features...

	Feature 1	Feature 2	Feature 3
Sample 1	10	6	12
Sample 2	11	4	9
Sample 3	8	5	10
Sample 4	3	3	2.5
Sample 5	1	2.8	1.3
Sample 6	2	1	2



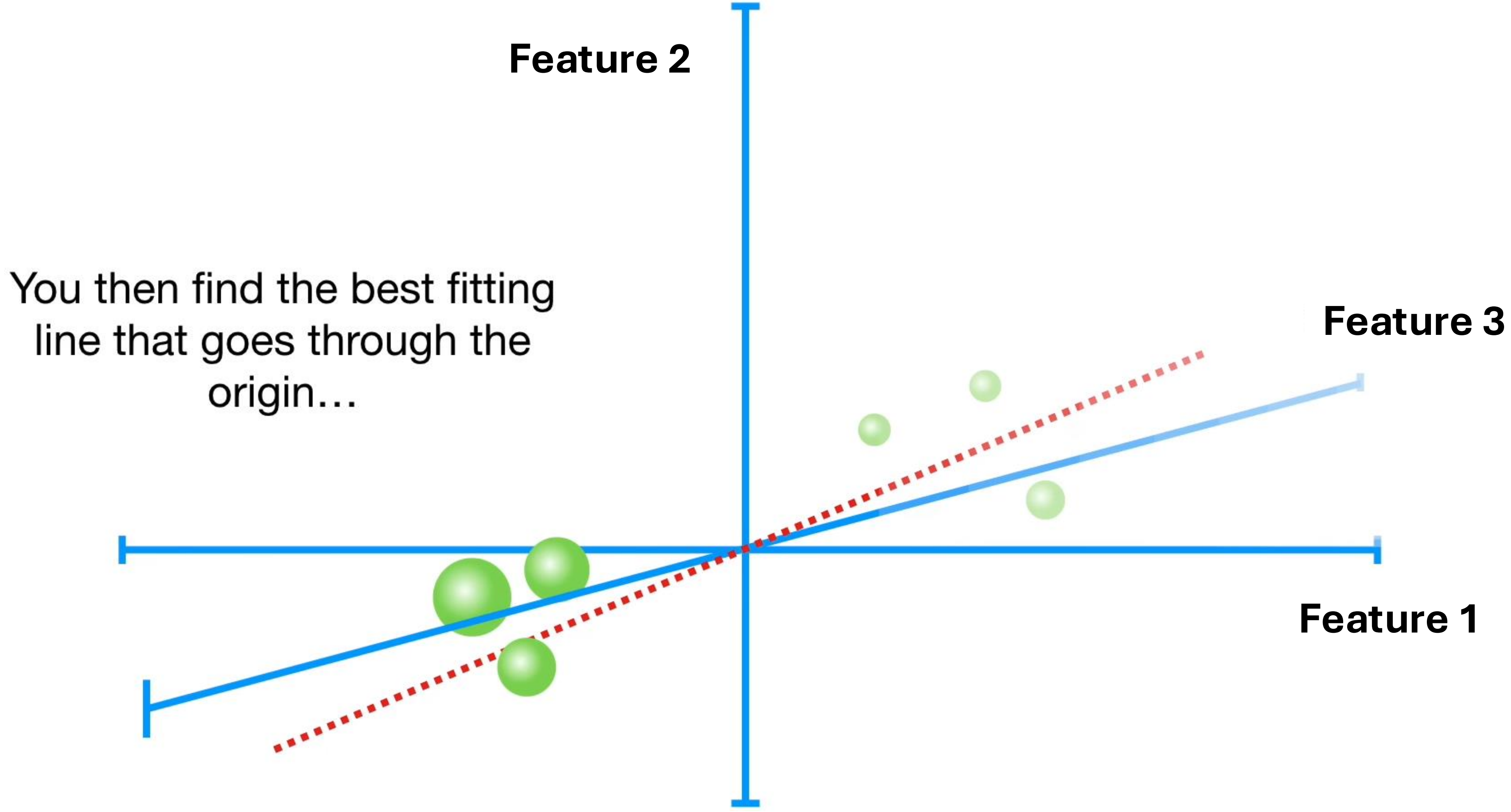
PCA with more Features

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

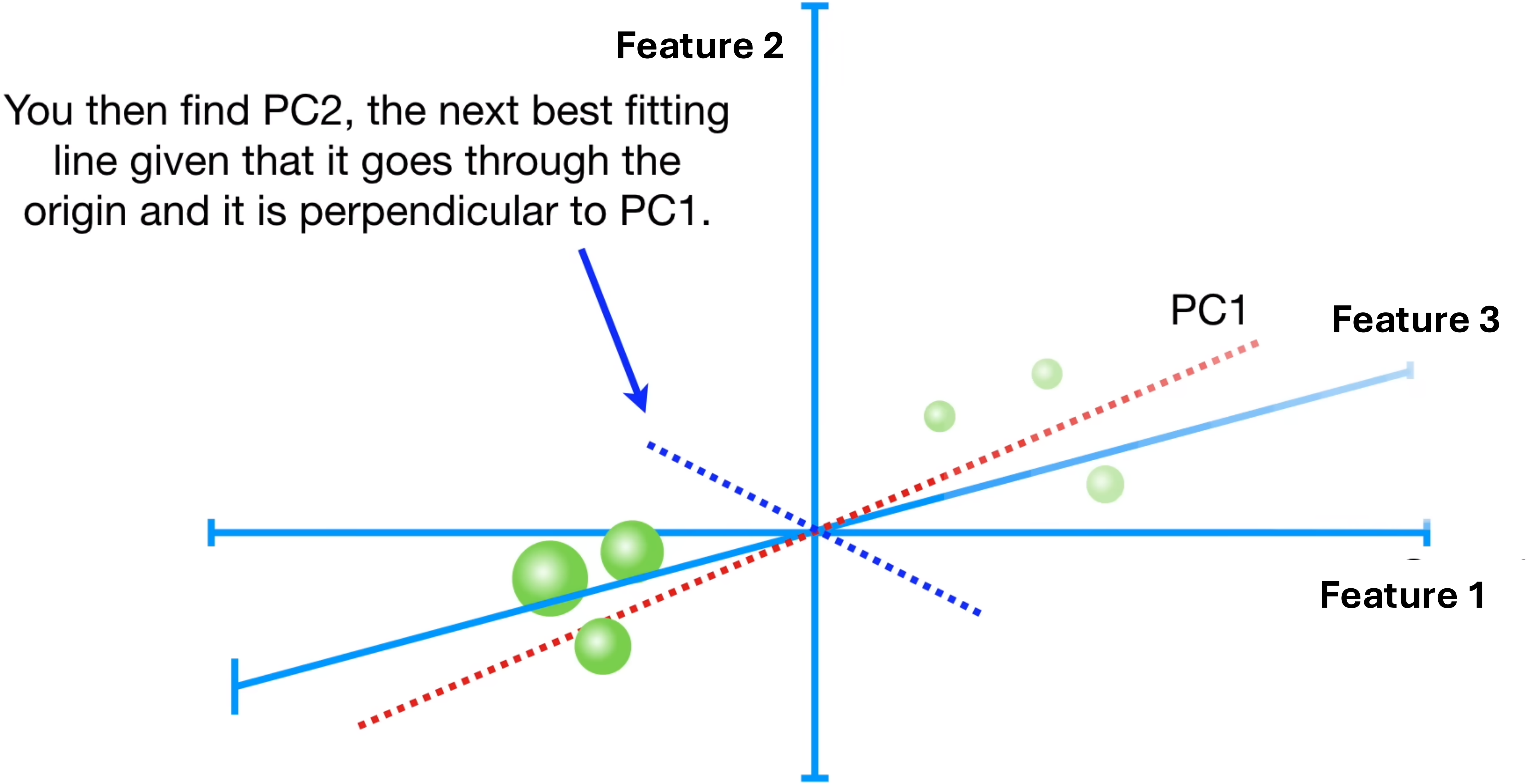


PCA with more Features

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

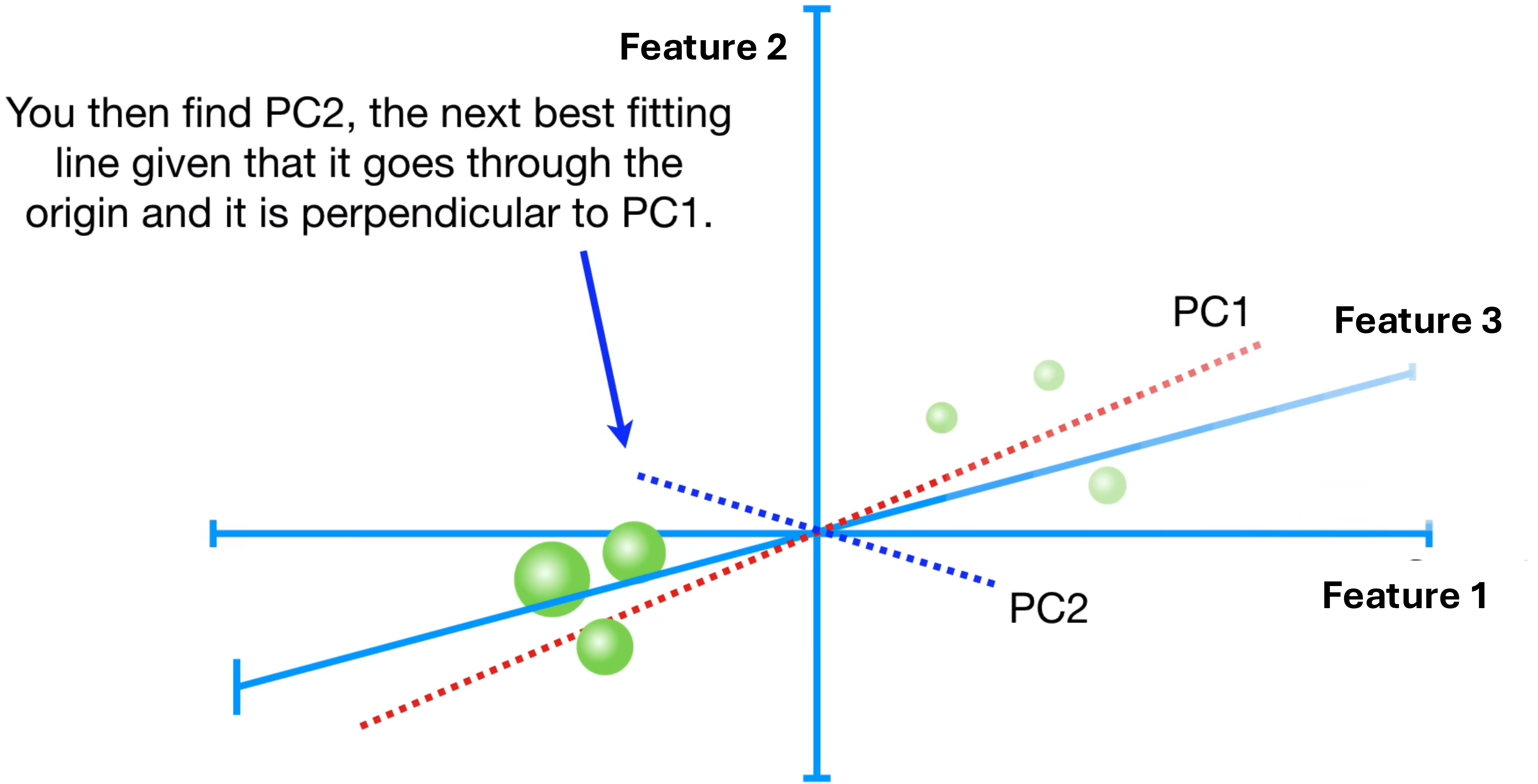


PCA with more Features

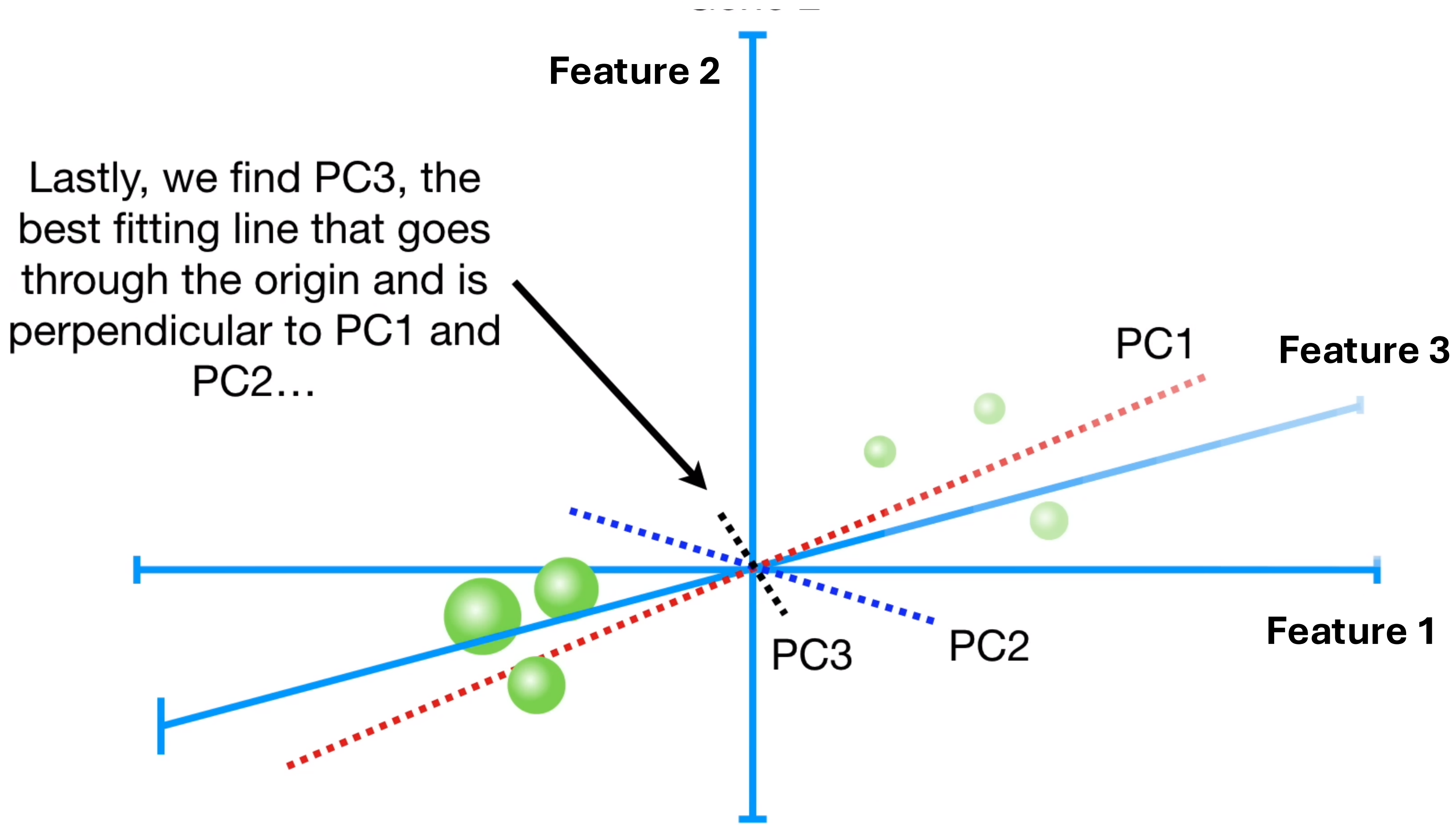


PCA with more Features

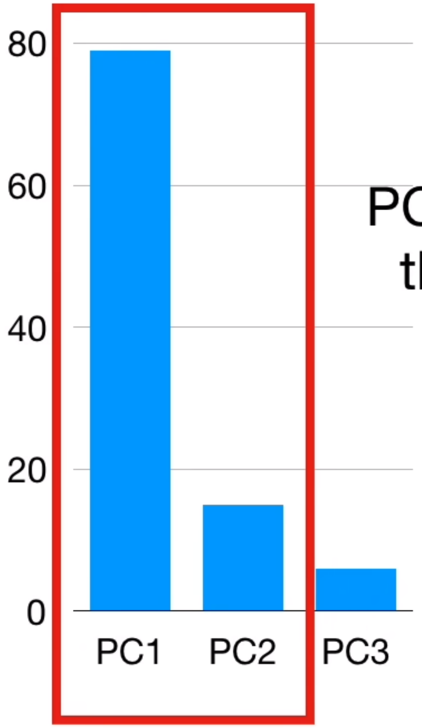
UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA



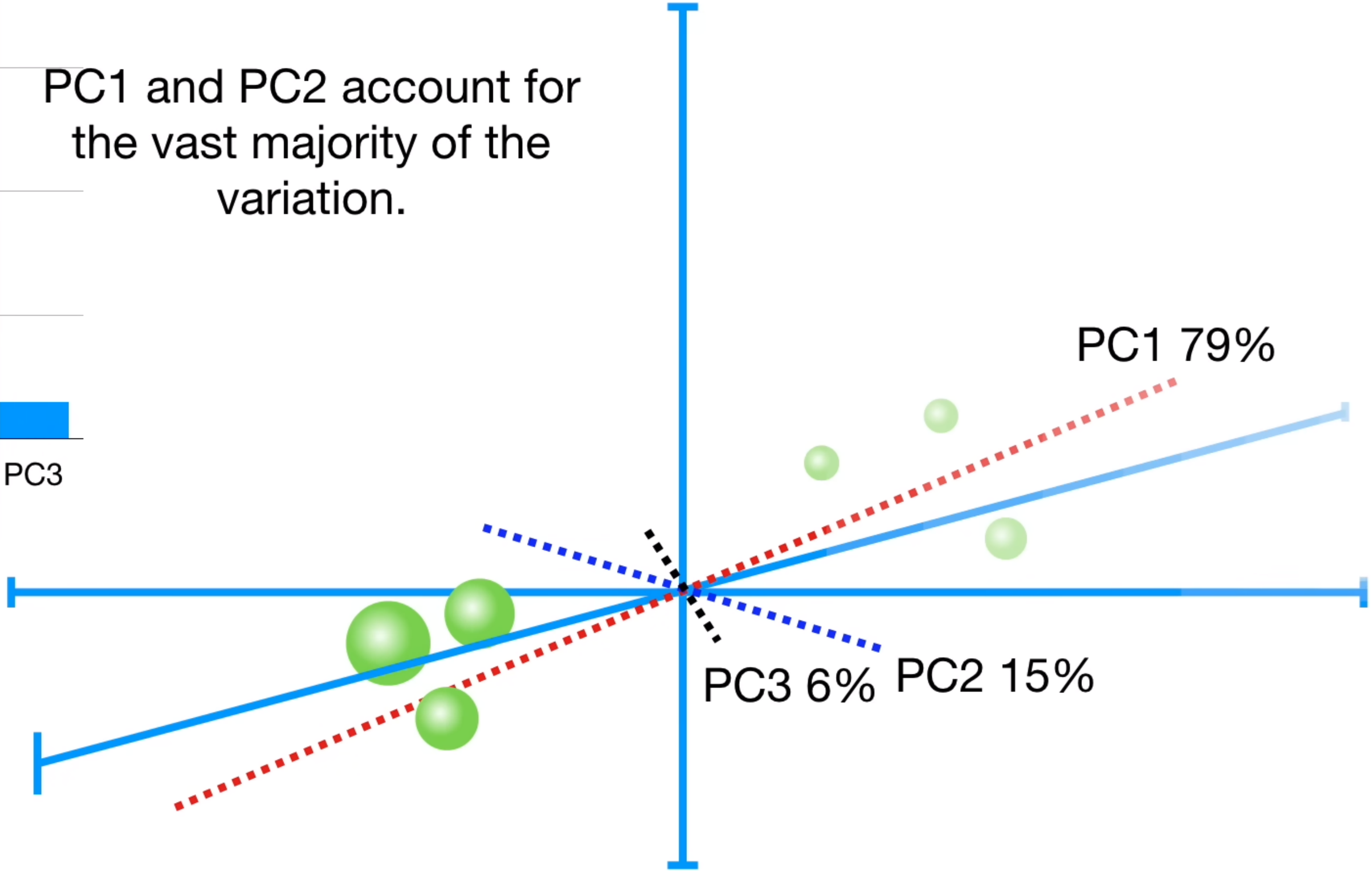
PCA with more Features



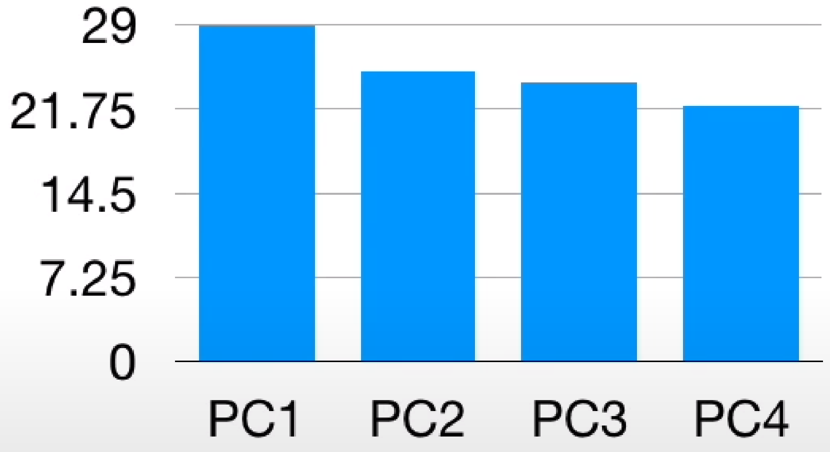
PCA with more Features



PC1 and PC2 account for the vast majority of the variation.

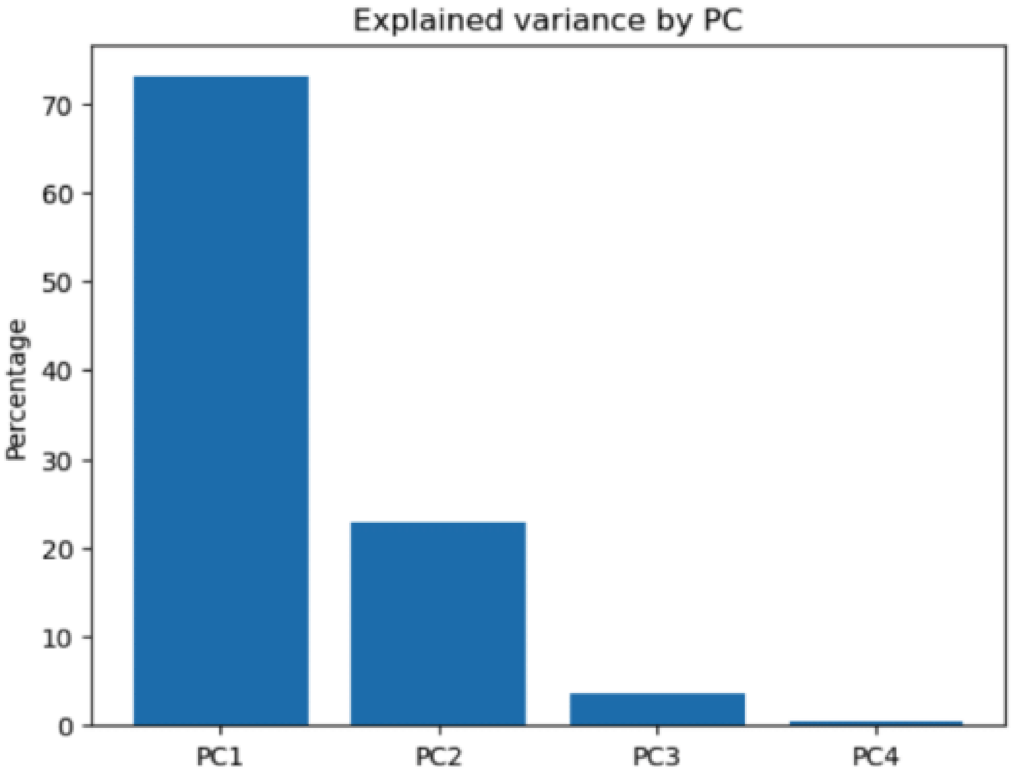


PCA with more Features

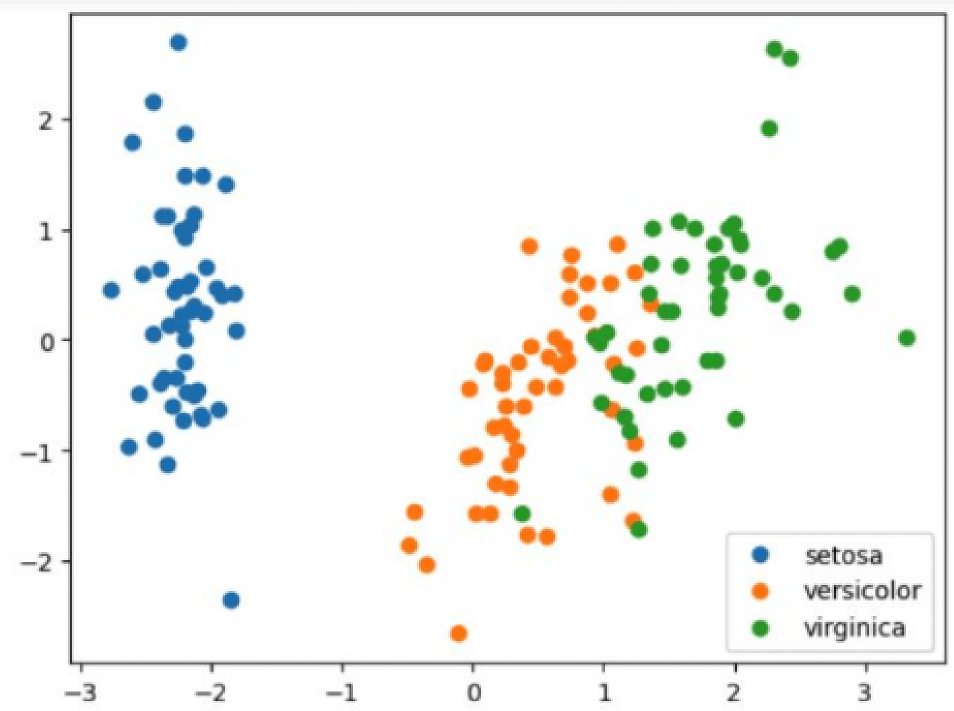


NOTE: If the scree plot looked like this, where PC3 and PC4 account for a substantial amount of variation, then just using the first 2 PCs would not create a very accurate representation of the data.

PCA example with the iris dataset



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.**

PCA Terminology

- **Principal Component:** Linear combinations of original variables that capture the maximum variance in the data.
- **Eigenvector:** Direction in the feature space that defines the principal components.
- **Eigenvalue:** Scalar indicating the amount of variance explained by its corresponding eigenvector.
- **Explained Variance:** Proportion of total variance in the data explained by each principal component.
- **Loading Score:** Weight or coefficient assigned to each original variable in the construction of principal components.

PCA Final Considerations

- **Non-Linear Relationships:** PCA assumes linear relationships between variables, limiting its effectiveness in capturing complex non-linear patterns; for such cases, consider non-linear techniques like t-SNE or UMAP.

- **Interpretability:** PCA creates new combinations of features that may be hard to interpret, making it less suitable when maintaining interpretability of individual features is crucial.

- **Sparse Data:** PCA's performance can suffer with sparse data where most values are zero or missing, leading to distortions in the data's structure.

- **Outliers:** PCA is sensitive to outliers, which can skew results; for datasets with outliers, robust PCA techniques may be necessary to mitigate their impact.

- **Data Scale:** PCA is sensitive to the scale of the features, so it's important to standardize or normalize the data before applying PCA to ensure that all features contribute equally to the analysis.

Other Dimesionality Reduction Techniques

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

Multi-Dimensional Scaling (MDS)

- **Objective:** Reduce high-dimensional data to a lower-dimensional space while preserving pairwise relationships, enabling visualization and analysis.

- **Approaches:**
  - **Metric MDS:** Seeks to preserve actual distances between data points in the lower-dimensional space, often utilizing optimization algorithms such as gradient descent.
  - **Non-Metric MDS:** Focuses on preserving the rank order of distances rather than their absolute values, making it suitable for ordinal or non-quantitative data.

- Computationally intensive for large datasets and sensituve to input.

Other Dimesionality Reduction Techniques

UNIVERSIDADE CATÓLICA PORTUGUESA BRAGA

t-Distributed Stochastic Neighbor Embedding (t-SNE)

- **Objective:** Reduce high-dimensional data to a lower-dimensional space, emphasizing **local relationships between data points**.

- **Approach:**
  - t-SNE uses a **non-linear** mapping approach that aims to preserve **local similarities** in the high-dimensional space by modeling them as conditional probabilities.
  - It minimizes the divergence between conditional probabilities in the high-dimensional and lower-dimensional spaces using gradient descent optimization.

- **Key Features:**
  - Emphasizes preservation of local structures, making it effective for visualizing clusters and manifold structures in the data.
  - Particularly useful for exploring complex and non-linear relationships in high-dimensional datasets.

Other Dimesionality Reduction Techniques

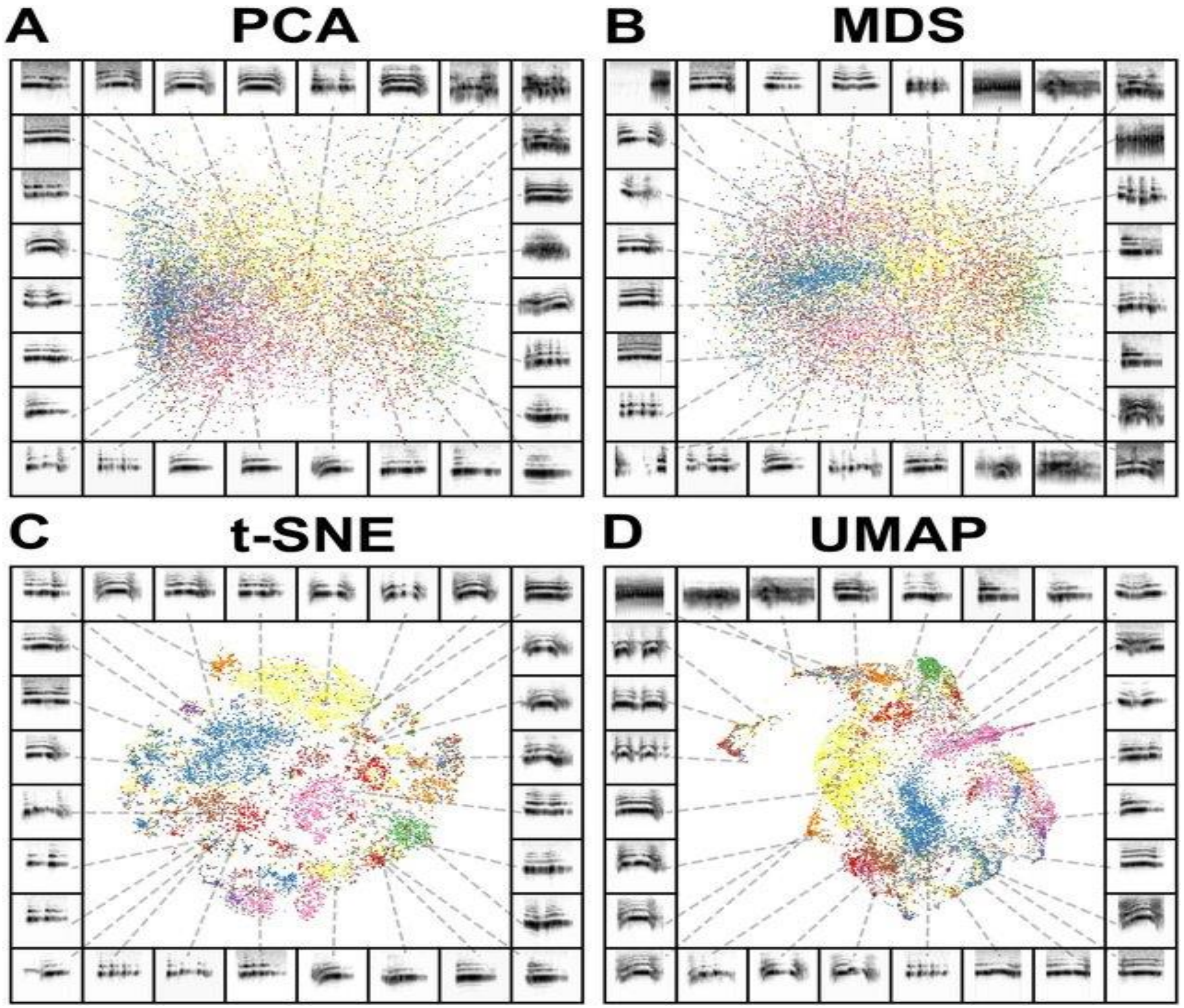
Uniform Manifold Approximation and Projection (UMAP)

- **Objective:** Reduce the dimensionality of high-dimensional data while preserving both local and global structure, offering a balance between preserving **local details and capturing global patterns**.

- **Approach:**
  - UMAP constructs a high-dimensional graph representing local relationships between data points and optimizes the embedding in a lower-dimensional space to match the graph topology.
  - It employs a combination of fuzzy set theory and Riemannian geometry to model the manifold structure of the data.

- **Key Features:**
  - Preserves both local and global structure, allowing for a more comprehensive representation of the data.
  - Offers flexibility in balancing preservation of local details and capturing global patterns through parameter tuning.
  - Known for its scalability and efficiency, making it suitable for large datasets.

PCA vs MDS vs tSNE vs UMAP



Sainburg, T., Thielk, M., & Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. In F. E. Theunissen (Ed.), PLOS Computational Biology (Vol. 16, Issue 10, p. e1008228). Public Library of Science (PLoS). https://doi.org/10.1371/journal.pcbi.1008228

Resources

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

- Dimension Reduction: A Guided Tour:

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

- Oskolkov, N. (2022). Dimensionality Reduction. In Applied Data Science in Tourism (pp. 151–167). Springer International Publishing. https://doi.org/10.1007/978-3-030-88389-8_9