أكاديمية طويق
Tuwaiq Academy

Principal Component Analysis

PCA

Principal Component Analysis

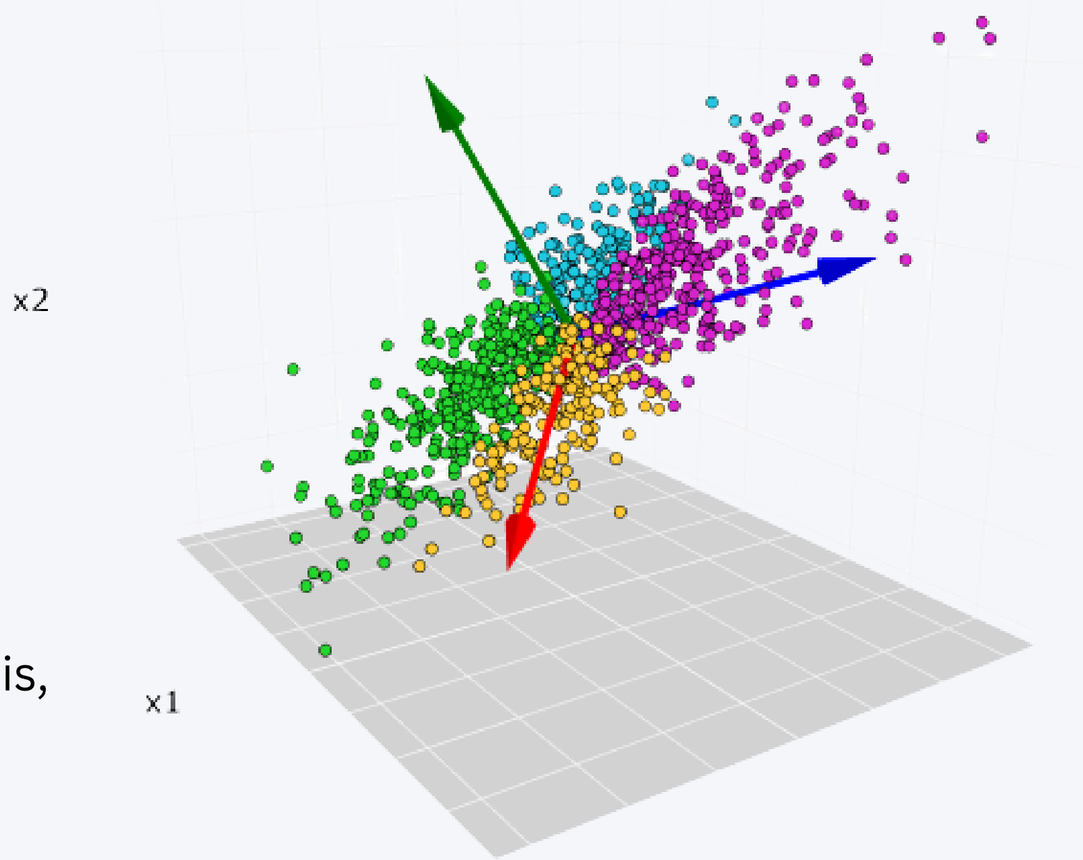
أكاديمية طويق
Tuwaiq Academy

PCA - USE CASES

A powerful dimensionality reduction technique in machine learning

Key Points:

- PCA is an unsupervised learning algorithm
- Used for dimensionality reduction, feature extraction, visualization, and noise filtering
- Common applications: Stock market prediction, gene analysis, image processing

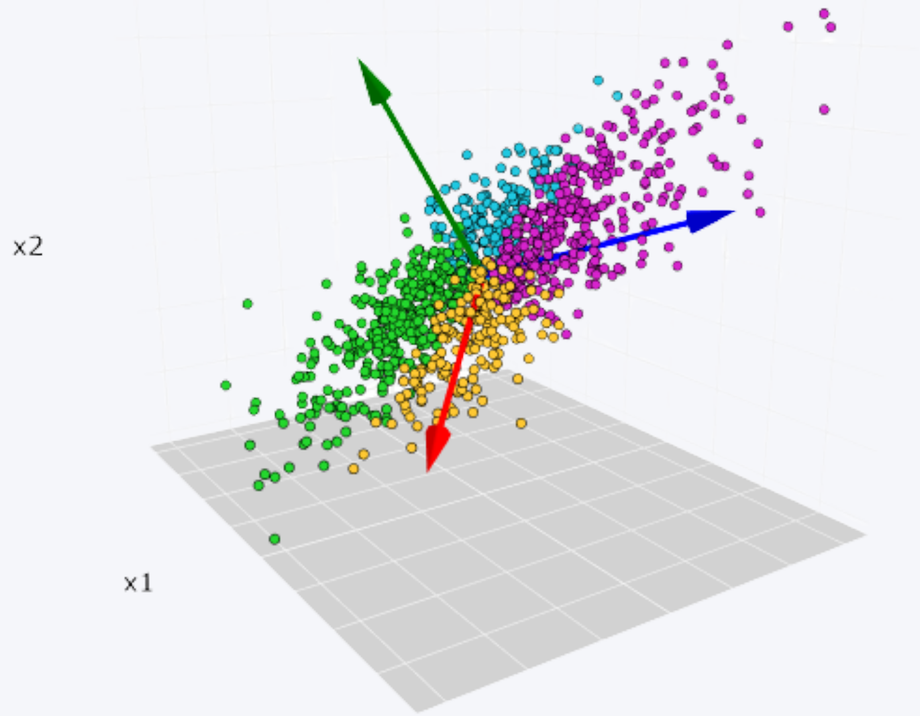


The original 3-dimensional data set. The red, blue, green arrows are the direction of the first, second, and third principal components, respectively.

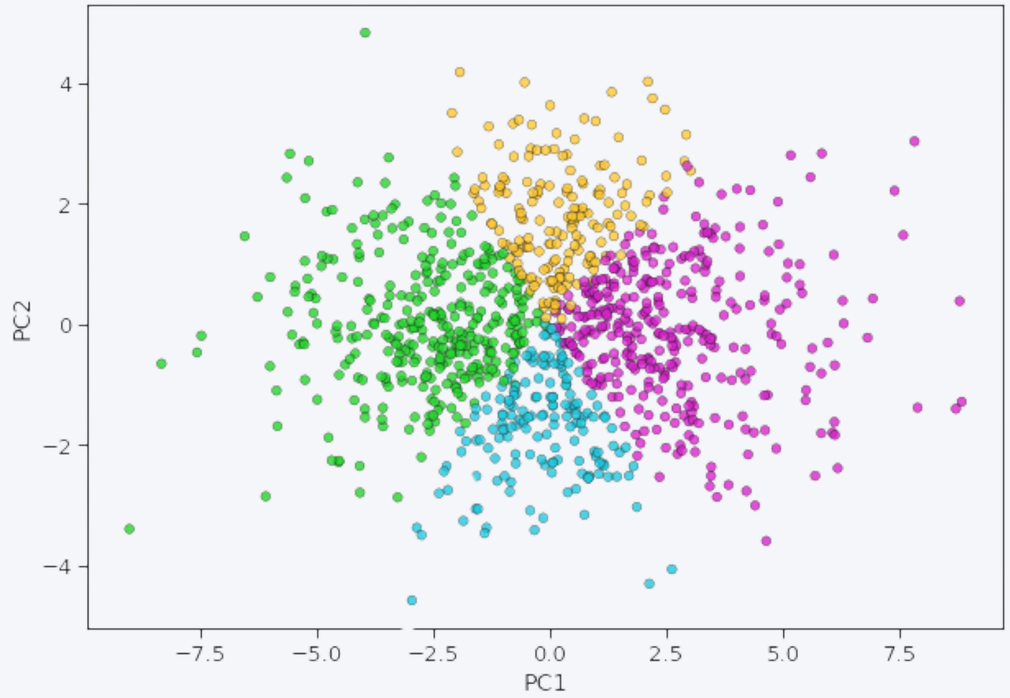
Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

PCA - USE CASES



The original 3-dimensional data set. The red, blue, green arrows are the direction of the first, second, and third principal components, respectively.



Scatterplot after PCA reduced from 3-dimensions to 2-dimensions.

Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

PCA - USE CASES

PCA is extremely useful when working with data sets that have **a lot of features.**

- Common applications such as image processing, genome research always have to deal with thousands-, **if not tens of thousands of columns.**

Sometimes, less is more.

Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

Simply,

Finding the time to read a 1000-pages book is a luxury that few can afford.

Wouldn’t it be nice if we can summarize the most important points in just 2 or 3 pages so that the information is easily received even by the busiest person?

We may lose some information in the process, but hey, at least we get the big picture.

Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

How does PCA work?

It’s a two-step process. We can’t write a book summary if we haven’t read or understood the content of the book.

PCA works the same way – understand, then summarize.

Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

Understanding data: The PCA Way

Human understands the meaning of a storybook through the use of expressive language. Unfortunately, PCA doesn't speak English. ***It has to find meaning within our data through its preferred language, mathematics.***

- Can PCA understand which part of our data is important?
- Can we mathematically quantify the amount of information embedded within the data?

Well, variance can.

The greater the variance, the more the information. Vice versa.

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

The formula of variance.

Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

So, where does this association comes from?

Our friends would cover their faces and we need to guess who's who based solely on their height.

A

B

C

Without a doubt, I am going to say that Person A is Chris, Person B is Alex, and Person C is Ben.

A

B

C

Can you guess who's who? It's tough when they are very similar in height.

In the same way, when our data has a higher variance, it holds more information. This is why we keep hearing **PCA** and maximum variance in the same sentence.

Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

Therefore,

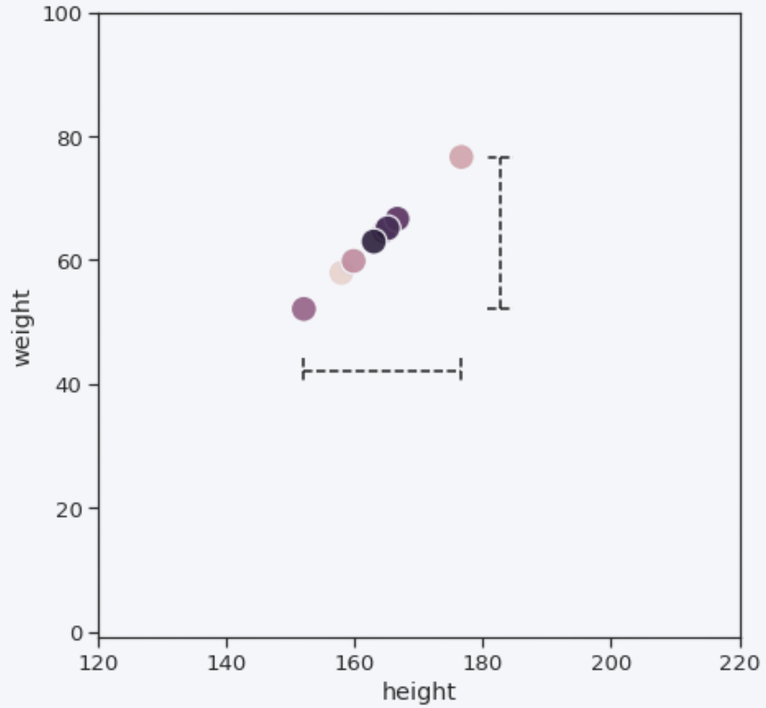
PCA is defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some scalar projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.

Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

Summarizing data with PCA



Feature	Variance
Height	1.11
Weight	1.11
TOTAL	**2.22**

All the features are standardized to the same scale for a fair comparison.

In this case, it's very difficult to choose the variables we want to delete. If I throw away either one of the variables, we are throwing away half of the information.

Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

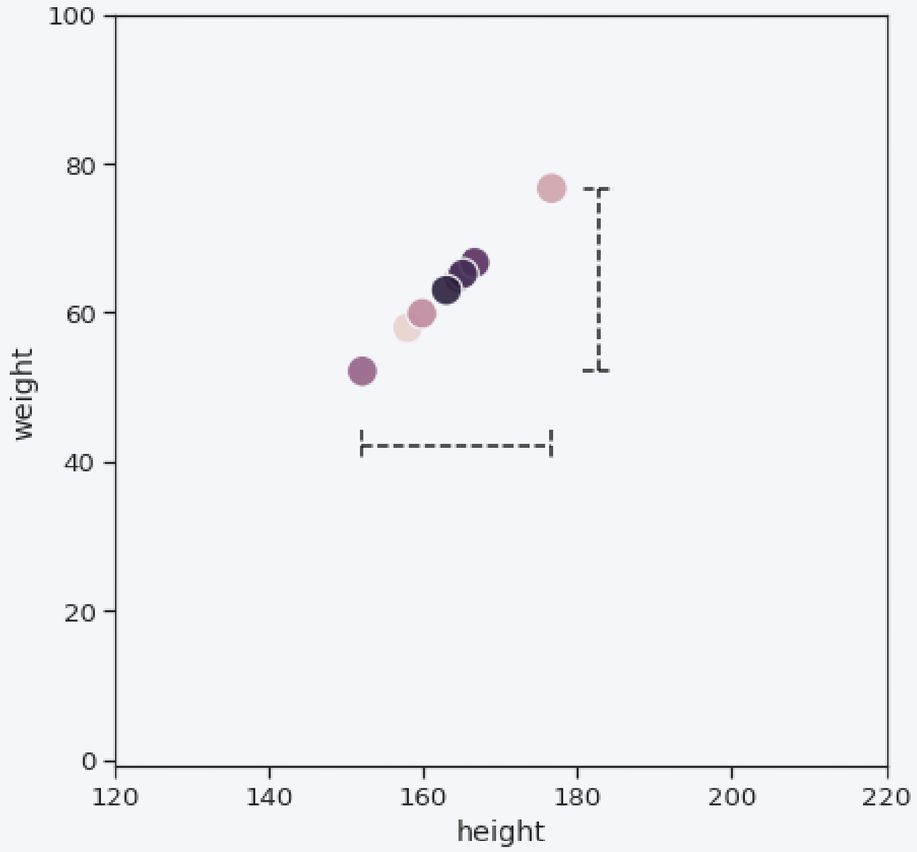
أكاديمية طويق
Tuwaiq Academy

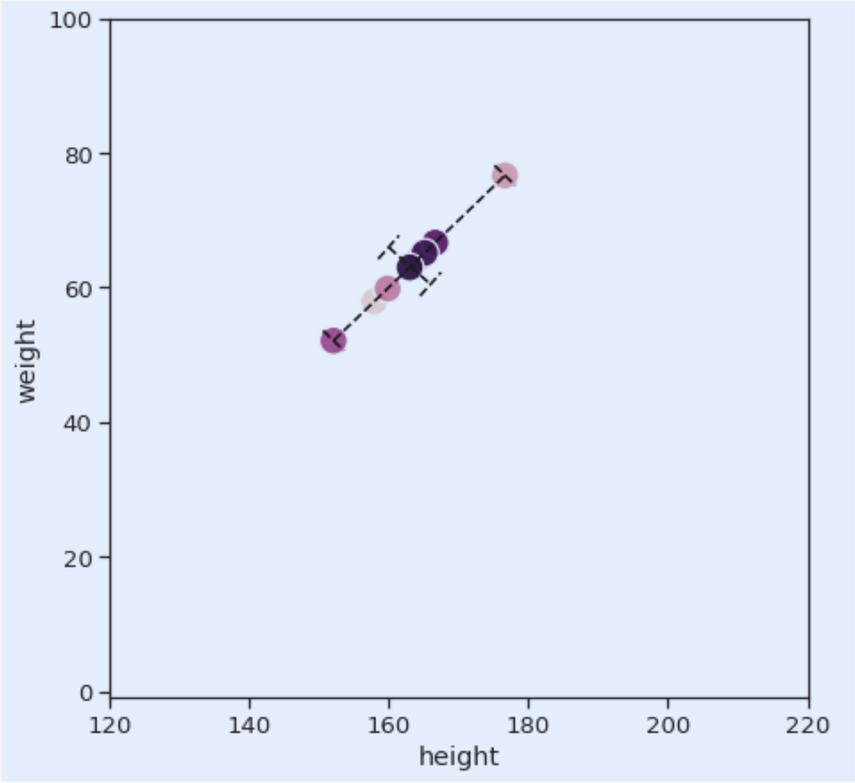
Can we keep both? -**Perhaps, with a different perspective.**

Instead of limiting ourselves to choose just one or the other, **why not combine them?**

When we look closer at our data, the maximum amount of variance lies not in the x-axis, not in the y-axis, but a diagonal line across.

- The second-largest variance would be a line 90 degrees that cuts through the first.





The dotted line shows the direction of maximum variance.

Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

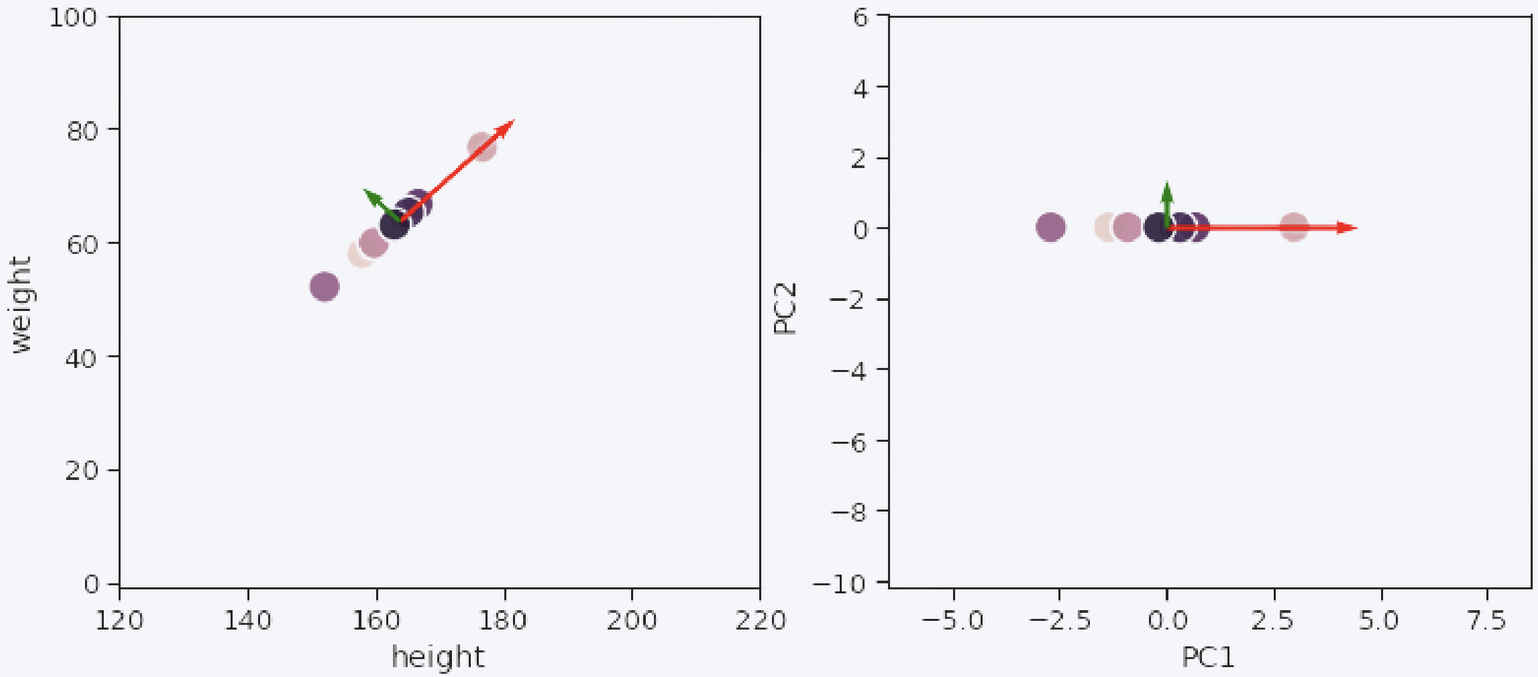
أكاديمية طويق
Tuwaiq Academy

Can we keep both? -**Perhaps, with a different perspective.**

To represent these 2 lines, PCA combines both height and weight to create two brand new variables.

It could be 30% height and 70% weight, or 87.2% height and 13.8% weight, or any other combinations depending on the data that we have.

These two new variables are called the first principal component (PC1) and the second principal component (PC2). Rather than using height and weight on the two axes, we can use PC1 and PC2 respectively.



(Left) The red and green arrows are the principal axes in the original data. Image by the author. | (Right) The direction of the principal axes have been rotated to become the new x- and y-axis.

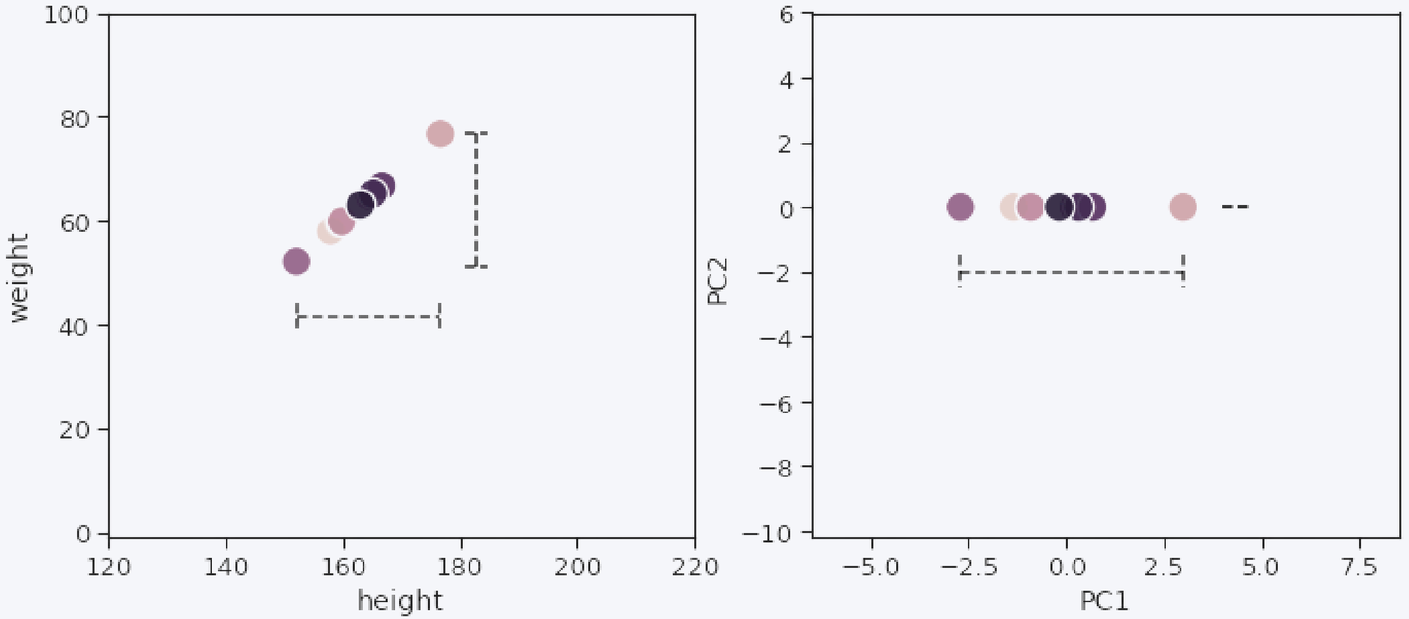
Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

let’s take a look at the variances again.

- **PC1** alone can capture the total variance of Height and Weight combined.
- Since PC1 has all the information – we can be very comfortable in **removing PC2** and know that our new data is still representative of the original data.



(Left) The variance of height and weight are similar in the original data. Image by the author. | (Right) After PCA transformation, all of the variances are shown in the PC1 axis.

Feature	Variance	Feature	Variance
Height	1.11	PC1	2.22
Weight	1.11	PC2	0.00
TOTAL	**2.22**	**TOTAL**	**2.22**

All the variables are standardized to the same scale for a fair comparison.

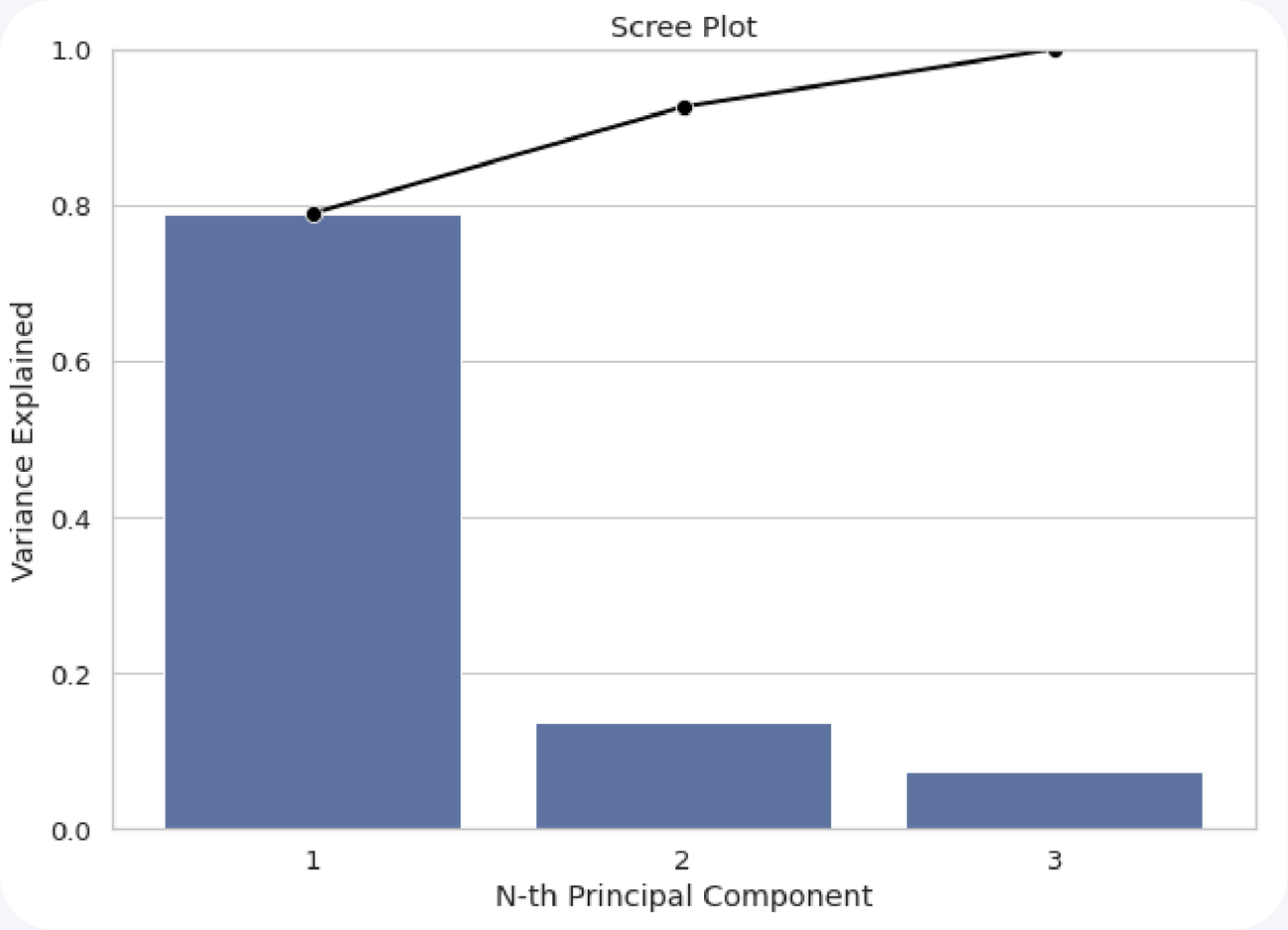
Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

let’s take a look at the variances again.

- When it comes to real data, more often than not, we won’t get a principal component that captures 100% of the variances.
  - Performing a PCA will give us N number of principal components, where N is equal to the dimensionality of our original data.
  - From this list of principal components, we generally choose the least number of principal components that would explain the most amount of our original data.
- The bar chart tells us the proportion of variance explained by each of the principal components.
- On the other hand, the line chart gives us the cumulative sum of explained variance up until N-th principal component. Ideally, we want to get at least 90% variance with just 2- to 3-components so that enough information is retained while we can still visualize our data on a chart.



Looking at the chart, I would feel comfortable using 2 principal components.

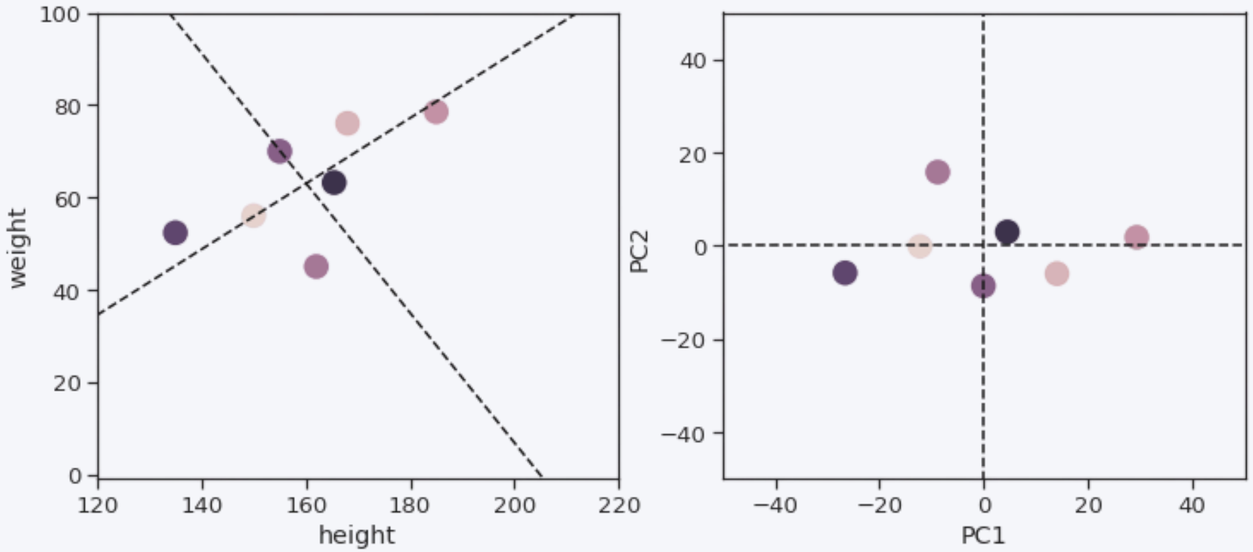
Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

The PC That Got Away

Since we're **not** choosing all the principal components, we inevitably **lose** some information. But we haven't exactly described what we are losing.



If we feed our data through the PCA model, it would start by drawing the First Principal Component followed by the Second Principal Component.

- When we transform our original data from 2-dimensions to 2-dimensions, everything stays the same except the orientation.
- We just rotated our data so that the maximum variance is in PC1. Nothing new here.

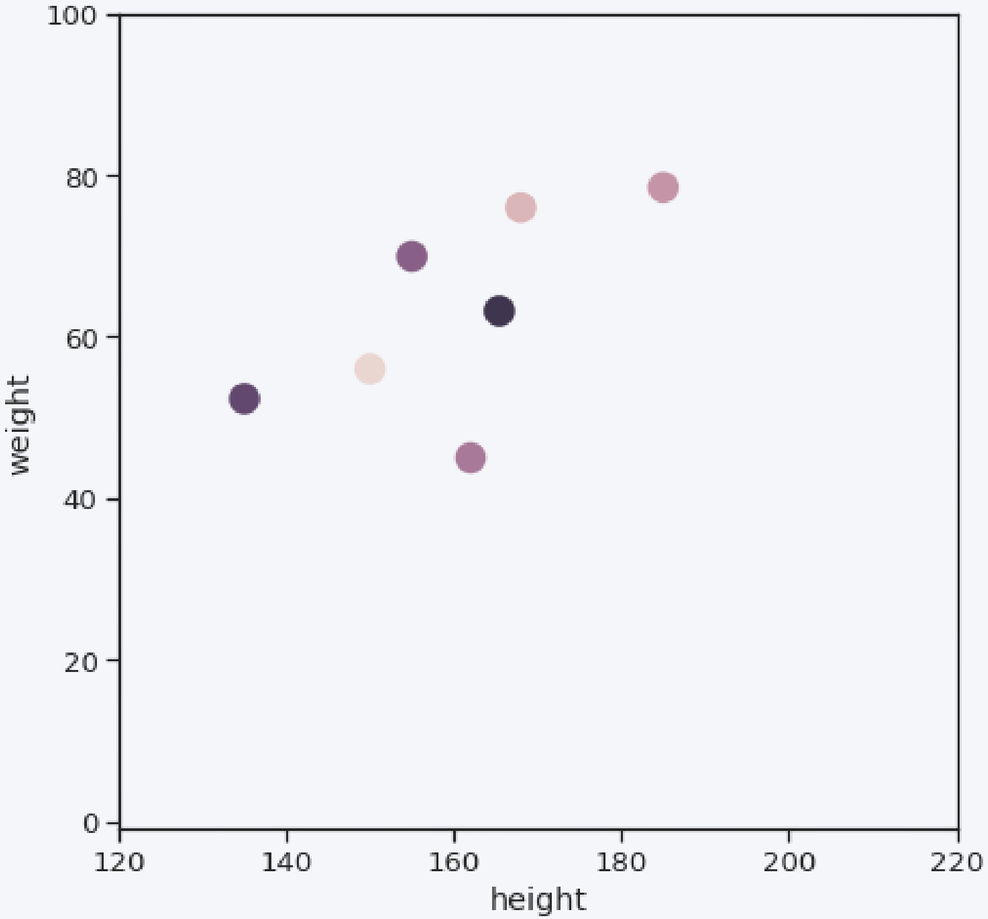


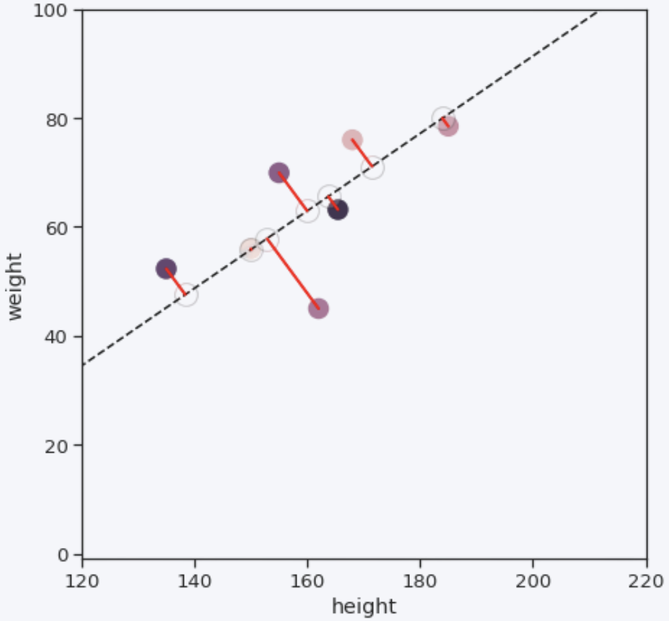
Image source: mdpi.com/2073-8994/10/12/73

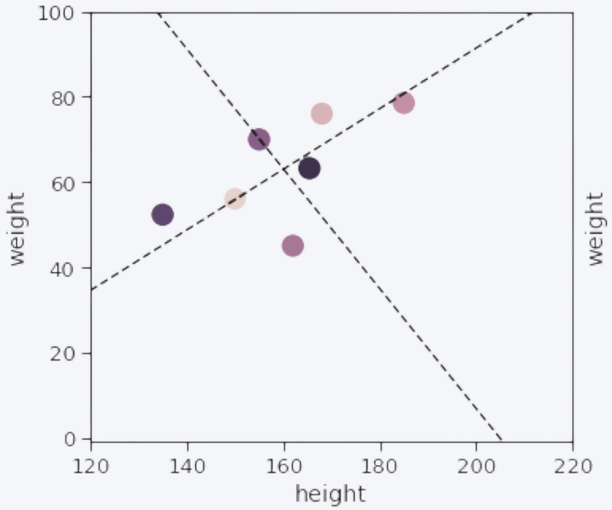
Principal Component Analysis

The PC That Got Away

However, suppose that we have decided to keep only the First Principal Component, we would have to project all our data points onto the First Principal Component because we no longer have the y-axis.

What we would lose is the distance in the Second Principal Component, highlighted with the red color line below.





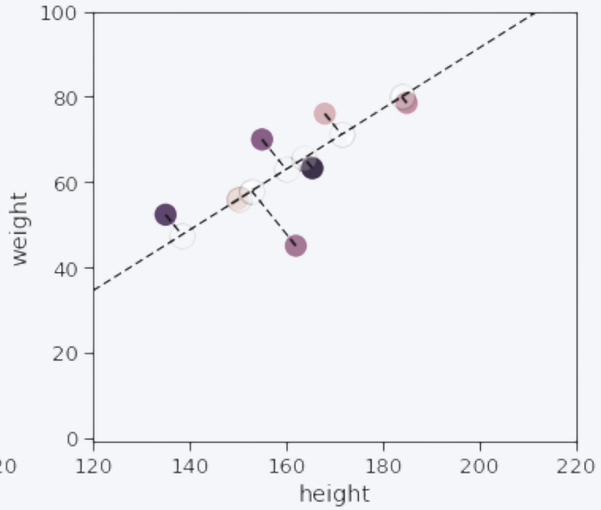


Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

The PC That Got Away

This has implications on the perceived distance of each data point. If we look at the Euclidean distance between two specific points (a.k.a pairwise distance), you will notice that some points are much farther in the original data than in the transformed data.

The PCA is a linear transformation so in and of itself does not alter distances, but when we start removing dimensions, the distances get distorted.

It gets trickier— not all pairwise distance gets affected equally.

If we take the two furthest points, you will see that they are almost parallel to the principal axes. Although their Euclidean distance is still distorted, it is to a much lesser degree.

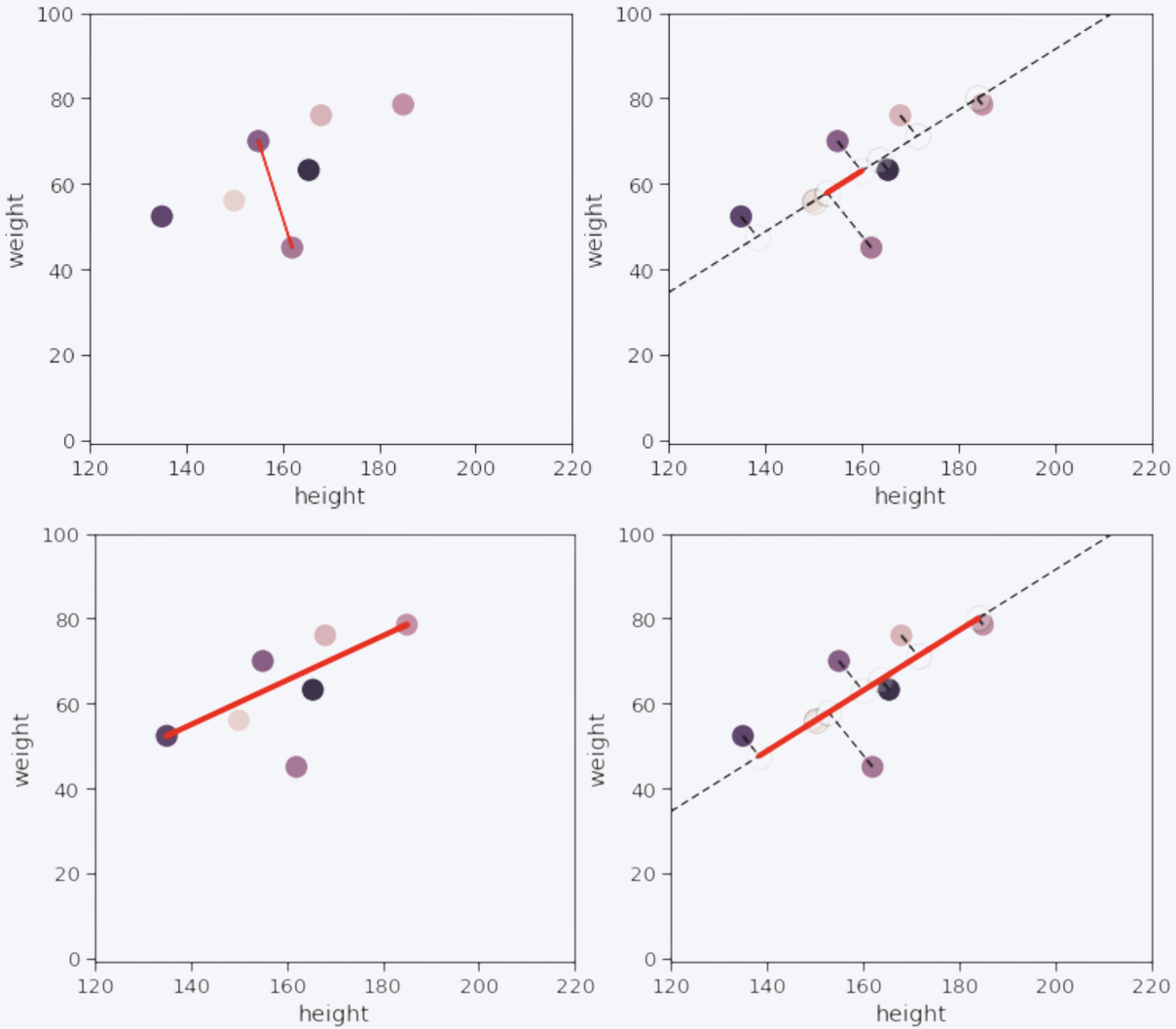


Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

The PC That Got Away

The reason is that principal component axes are drawn in the direction where we have the largest variance.

By definition, variance increases when the data points are further apart. So naturally, the points furthest apart would align themselves better with the principal axes.

To sum it all up, reducing dimensions with PCA changes the distances of our data.

It does so in a way that preserves large pairwise distance better than small pairwise distance.

This is one of the few drawbacks of reducing dimensions with PCA and we need to be aware of that, especially when working with Euclidean distance-based algorithm.

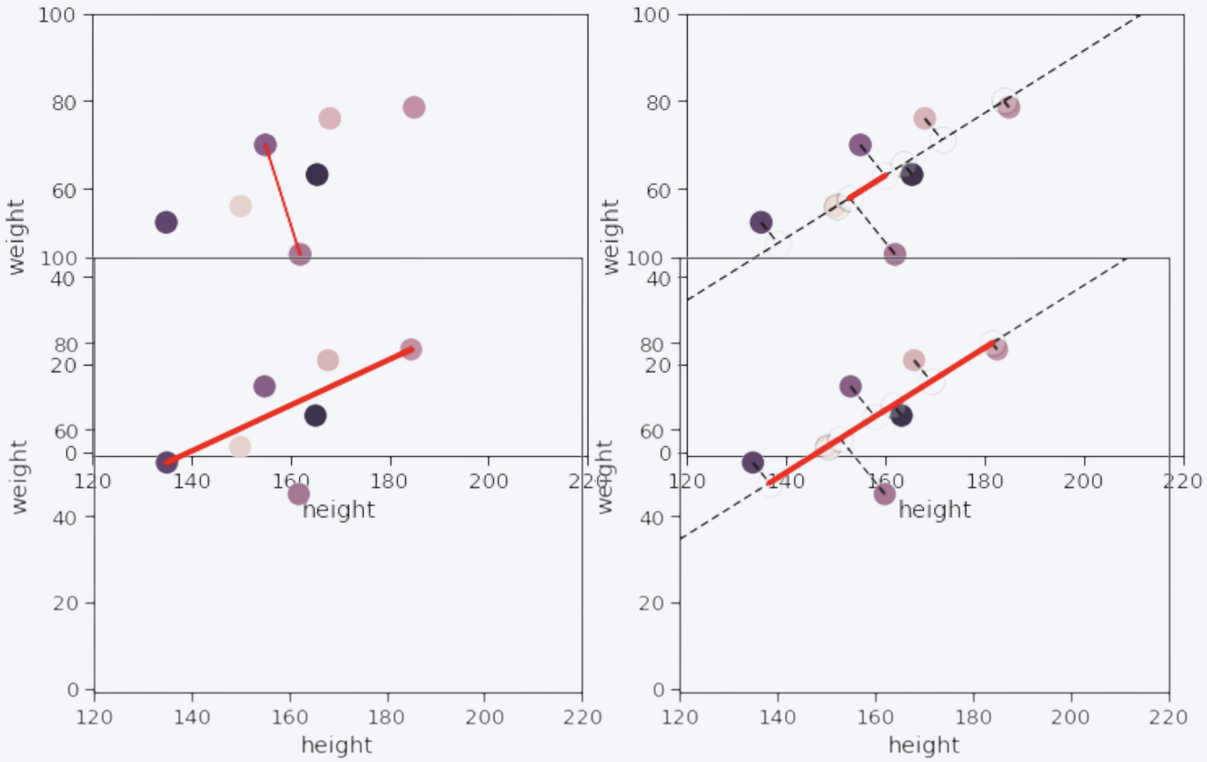


Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

Steps in PCA Algorithm

- **Standardize** the data (Mean = 0, Variance = 1)

- Compute **eigenvectors & eigenvalues** from the covariance matrix
  - Eigenvectors represent the directions (axes) of maximum variance.
  - Eigenvalues tell us the magnitude of variance in those directions.

- Sort eigenvalues in **descending order** to determine importance

- Construct **projection matrix (W)** from selected **K** eigenvectors

- Transform original dataset into a **new feature space**

Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

PCA vs. Linear Regression

- **PCA** is NOT the same as linear regression
- **Linear Regression**: Predicts Y values from X values
- **PCA**: Finds principal axes that best describe the data (not predicting values)

Image source: mdpi.com/2073-8994/10/12/73

Principal Component Analysis

أكاديمية طويق
Tuwaiq Academy

Weakness of PCA

- Sensitive to Outliers
- Large deviations can skew principal components
- Preprocessing techniques like removing outliers or scaling data help improve performance

Image source: mdpi.com/2073-8994/10/12/73

أكاديمية طويق
Tuwaiq Academy

Hands-On Code

PCA Implementation