

This website uses cookies. If you continue to use the website, we assume that you agree.

[OK](#) [Privacy Policy](#)

# PCA vs Linear Regression – Therefore you should know the differences

JANUARY 16, 2021 / RAINERGEWALT / 2 COMMENTS

**PCA vs Linear Regression** – Two statistical methods that run very similarly.

This website uses cookies. If you continue to use the website, we assume that you agree.

[OK](#) [Privacy Policy](#)

## Table of Contents

### 1. What is a PCA?

1.1. What are linear combinations?

1.2. How does the algorithm work?

### 2. PCA vs Linear Regression – How do they Differ?

## What is a PCA?

Principal Component Analysis (PCA) is a multivariate statistical method for structuring or simplifying a large data set. The main goal here is the discovery of relationships in 2 or 3 dimensional domain.

This method enjoys great popularity in almost all scientific disciplines and is mostly used when variables are highly correlated.

However, PCA is only a reliable method if the data are at least interval scaled and approximately normally distributed.

Although the variables are adjusted to avoid redundant effects, the error and residual variance of the data are not taken into account.

The following figure shows the basic principle of a PCA. High dimensional data relationships should be represented in a low dimensional way, with as little loss of information as possible.

This website uses cookies. If you continue to use the website, we assume that you agree.

[OK](#) [Privacy Policy](#)

### *PCA vs Linear Regression – Basic principle of a PCA*

The key point of PCA is dimensional reduction. It is to extract the most important features of a data set by reducing the total number of measured variables with a large proportion of the variance of all variables.

This reduction is done mathematically using linear combinations.

## **What are linear combinations?**

PCA works in a purely exploratory way, searching the data for a linear pattern that

.....

This website uses cookies. If you continue to use the website, we assume that you agree.

OK Privacy Policy

values.

In the figure below, the **linear combinations** have been applied to a data set.

*Linear combinations*

## How does the algorithm work?

In the principal component analysis procedure, a set of fully uncorrelated principal components are first generated.

These contain the main changes in the data and are also known as latent variables,

This website uses cookies. If you continue to use the website, we assume that you agree.

[OK](#) [Privacy Policy](#)

The first principal component is formed by minimizing the sum of squared variances of all variables.

During extraction, the variance component is maximized over all variables.

Then, the remaining variance is gradually resolved by the second component until the total variance of all data is explained by the principal components.

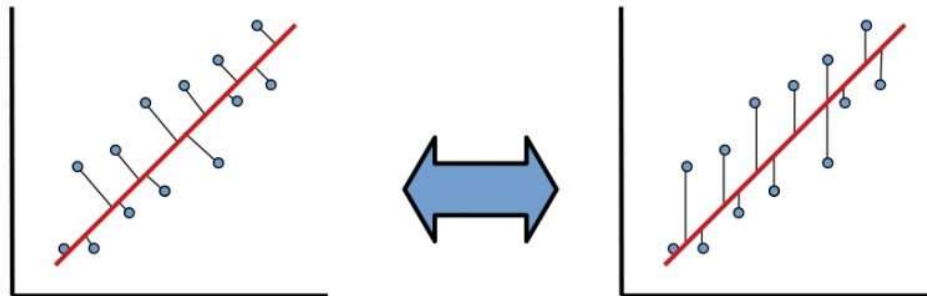
The first factor always points in the direction of the maximum variance in the data.

The second factor must be perpendicular to it and explain the next largest variance

## PCA vs Linear Regression – How do they Differ?

We have studied the PCA and how it works in great detail. But what are the differences to linear regression?

In the following illustration the main difference is set up against each other.



This website uses cookies. If you continue to use the website, we assume that you agree.

OK Privacy Policy

the y-direction.

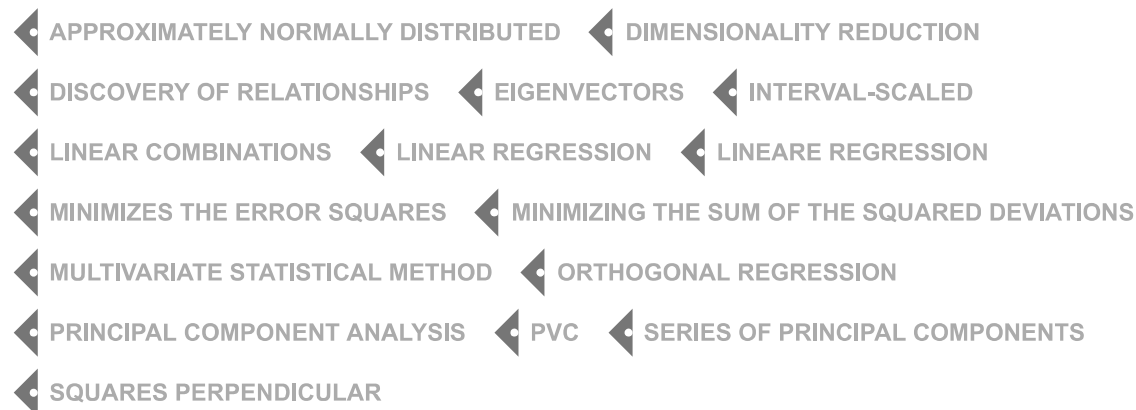
Thus, linear regression is more about finding a straight line that best fits the data, depending on the internal data relationships.

Principal component analysis uses an orthogonal transformation to form the principal components, or linear combinations of the variables.

So this difference between the two techniques only becomes apparent when the data are not completely independent, but there is a correlation.

If you want to know more about machine learning methods and how they work, check out our [article on the t-SNE algorithm](#).

Big Data, datascience, Machine learning, Python, Uncategorized



#### PREVIOUS POST

t-SNE – Great Machine Learning Algorithm for Visualization of High-Dimensional Datasets

#### NEXT POST

Messaging Patterns – It is not enough to decide to use a message

## 1 Comment

**cola**

11/24/21, 6:16 PM

This website uses cookies. If you continue to use the website, we assume that you agree.

OK Privacy Policy

**REPLY**

## 1 Pingback

1. [Supervised Vs Unsupervised Vs Reinforcement Learning - Simplified. What Are The Essential Differences? – Fly Spaceships With Your Mind](#)

## Leave a Reply

Your email address will not be published. Required fields are marked \*

Comment

Name \*

Email \*

Website

This website uses cookies. If you continue to use the website, we assume that you agree.

[OK](#) [Privacy Policy](#)



Post Comment

© 2021 FLY SPACESHIPS WITH YOUR MIND — UP ↑

This website uses cookies. If you continue to use the website, we assume that you agree.

[OK](#) [Privacy Policy](#)