

machine learning

07 - pca

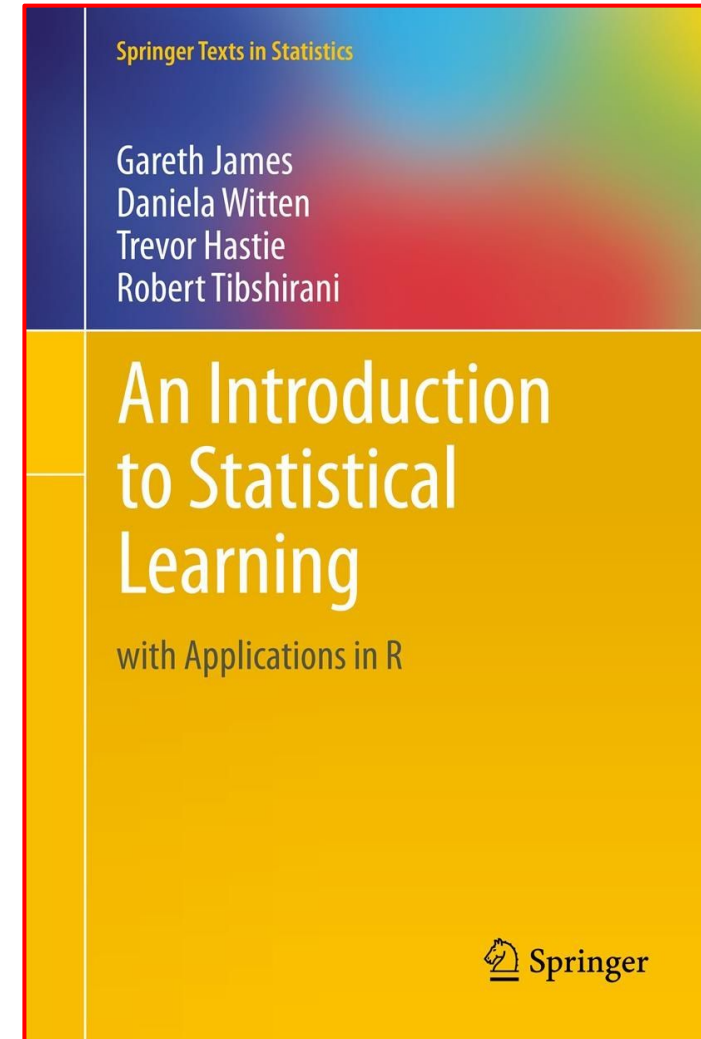
francisco josé diego acosta

Section 10.2

An Introduction to Statistical Learning

by Gareth James, et al.

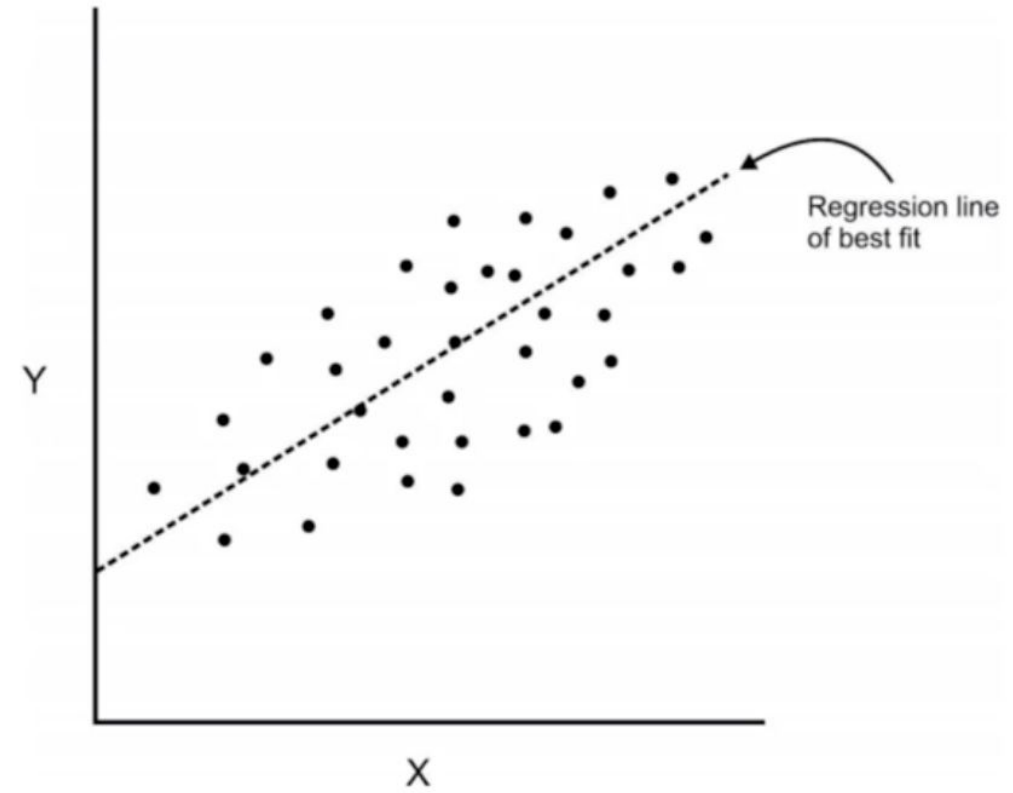
[https://www-bcf.usc.edu/~gar
eth/ISL/ISLR%20Seventh%20P
rinting.pdf](https://www-bcf.usc.edu/~gar
eth/ISL/ISLR%20Seventh%20P
rinting.pdf)



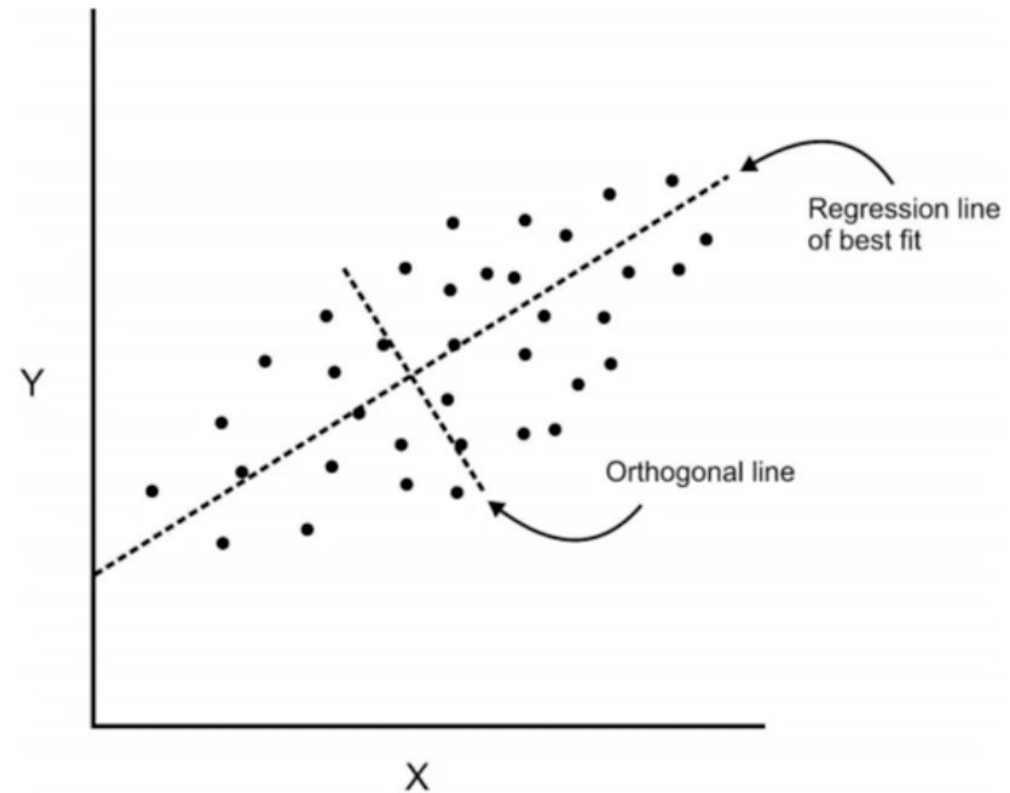
- **It is not a machine learning algorithm!**
- It is an unsupervised statistical technique used to examine the interrelation among a set of variables in order to identify the underlying structure of those variables
- It is also known in general terms as **factor analysis**

- where regression determines a line of best fit to a data set, factor analysis determines several orthogonal lines of best fit to the data set.
- Orthogonal means “at right angles”.
 - Actually the lines are perpendicular to each other in n-dimensional space
- n-Dimensional Space is the variable sample space
 - There are as many dimensions as there are variable, so in a dataset with 4 variables the sample space has 4 dimensions

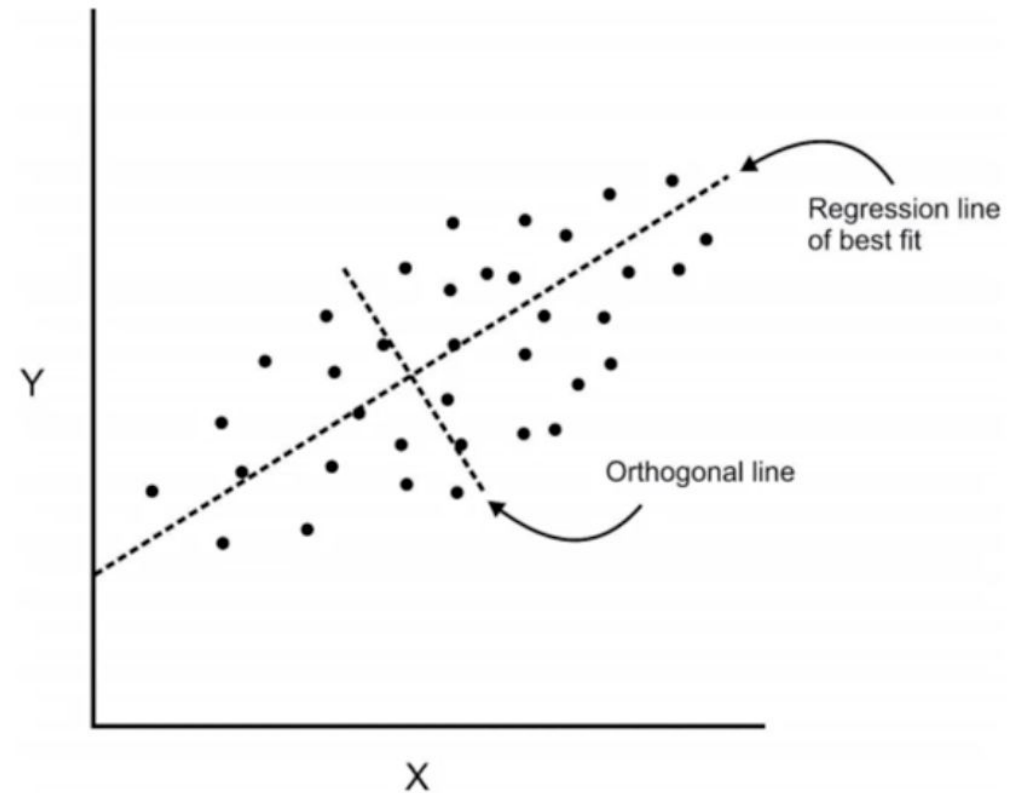
- here we have some data plotted along two features, x and y



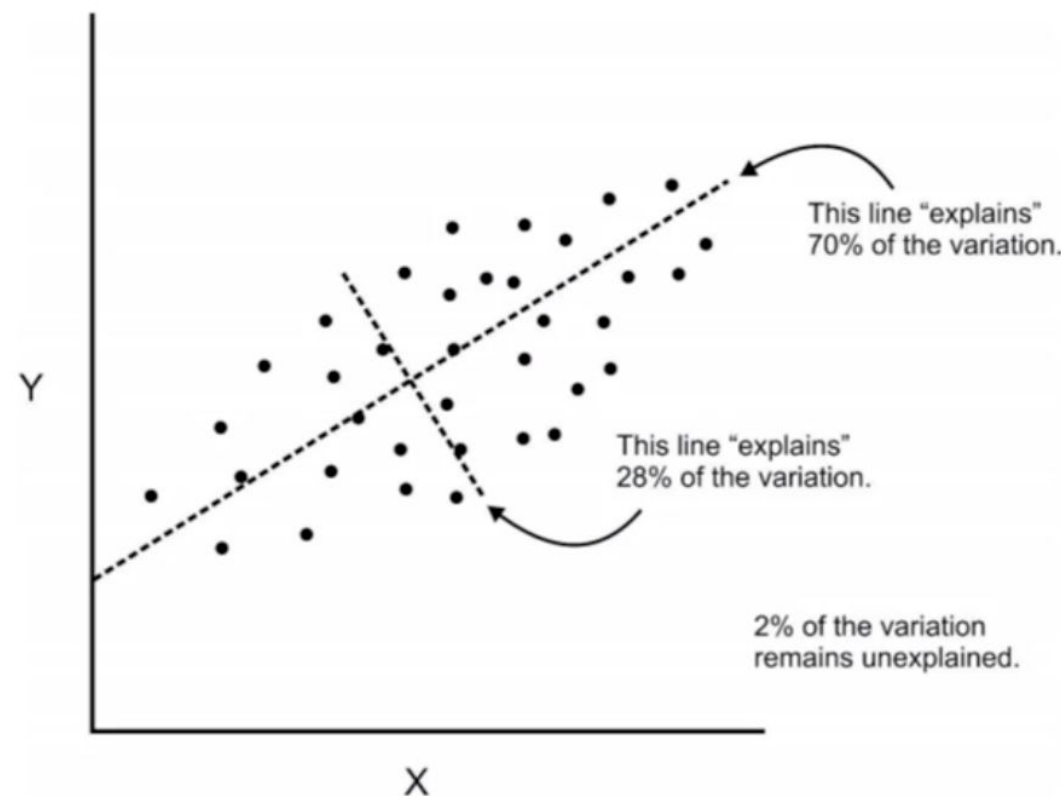
- here we have some data plotted along two features, x and y
- we can add an orthogonal line
- Now we can begin to understand the components!



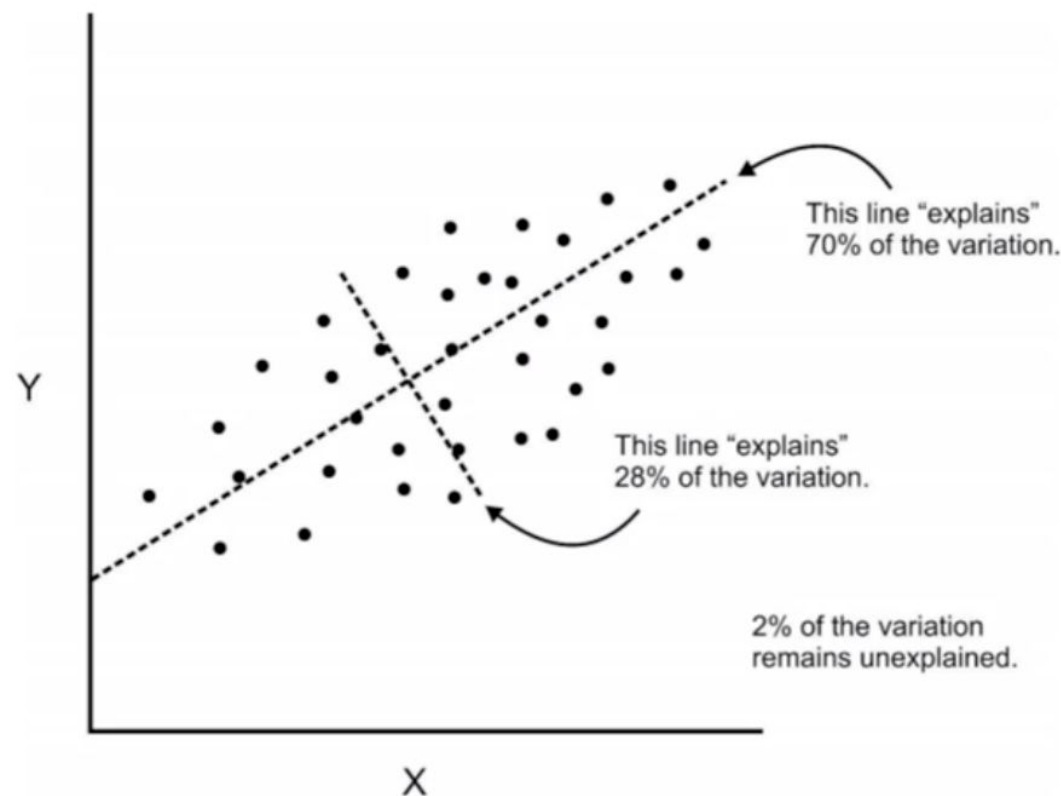
- Components are a linear transformation that choose a variable system for the dataset such that the greatest variance of the dataset comes to lie on the first axis



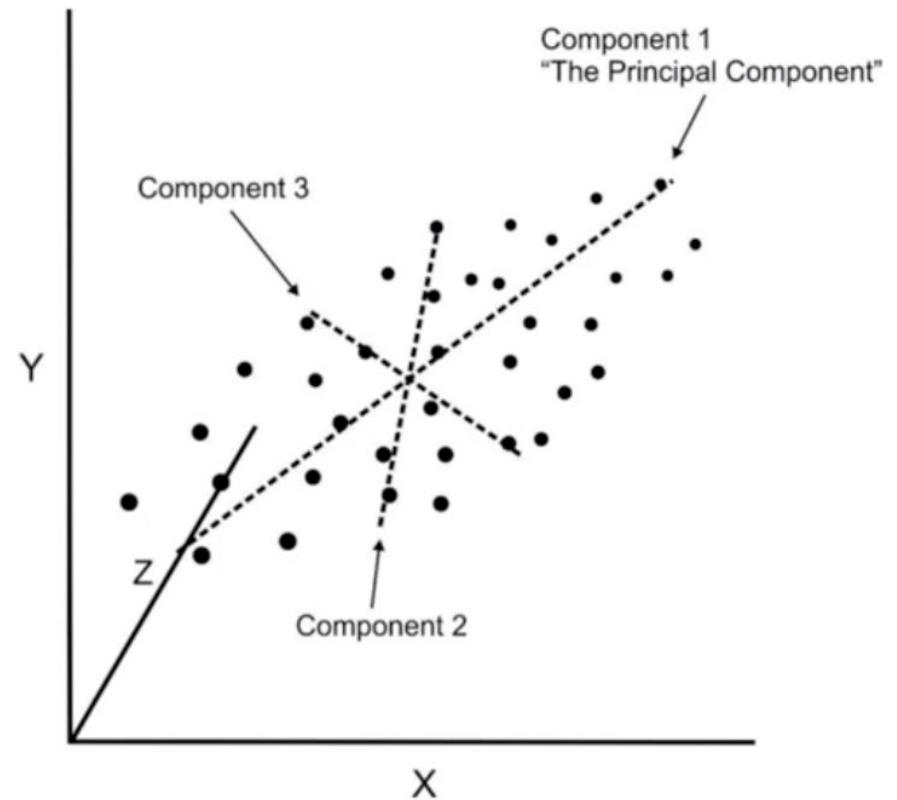
- The second greatest variance on the second axis and so on..
- this process allows us to reduce the number of variables used in a analysis.



- note that components are uncorrelated since in the same space they are orthogonal to each other

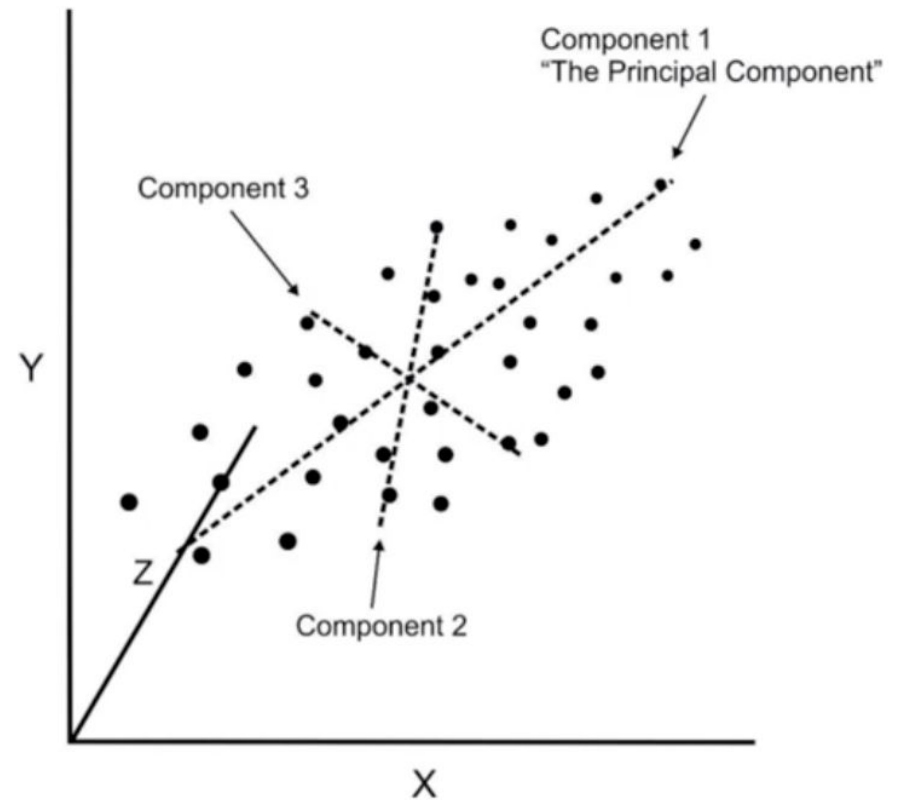


- we can continue this analysis into higher dimensions



- we can continue this analysis into higher dimensions

- we can continue this analysis into higher dimensions



pca

practice

