
Finamex[®]

Casa de Bolsa

PCA

What is a PCA?:

Principal component analysis (PCA) is a technique used to emphasize variation and bring out strong patterns in a dataset. It's often used to make data easy to explore and visualize.

“How do I take all of the variables I’ve collected and focus on only a few of them?”

By reducing the dimension of your feature space, you have fewer relationships between variables to consider and you are less likely to overfit your model.

Reducing the dimension of the feature space is called “dimensionality reduction”.

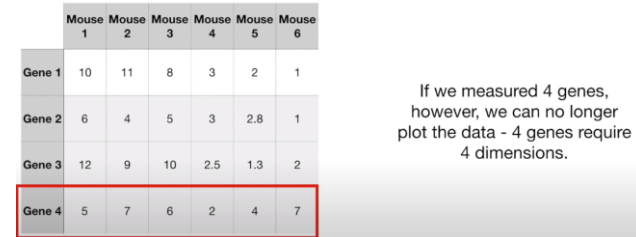
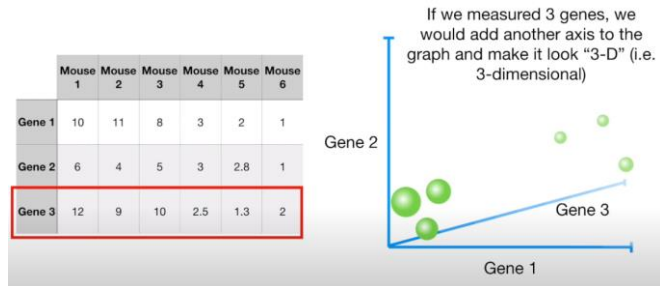
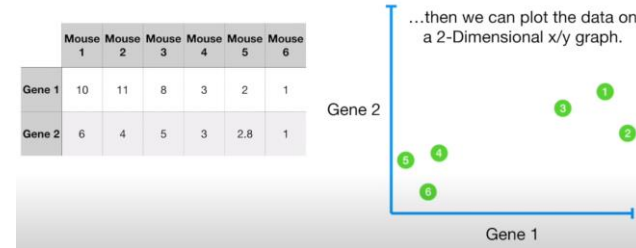
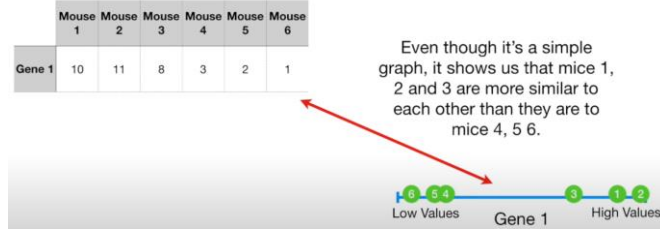
When should I use PCA?:

- Do you want to reduce the number of variables, but aren't able to identify variables to completely remove from consideration?
- Do you want to ensure your variables are independent of one another?
- Are you comfortable making your independent variables less interpretable?

If you answered "NO" to question 3, you should not use PCA.

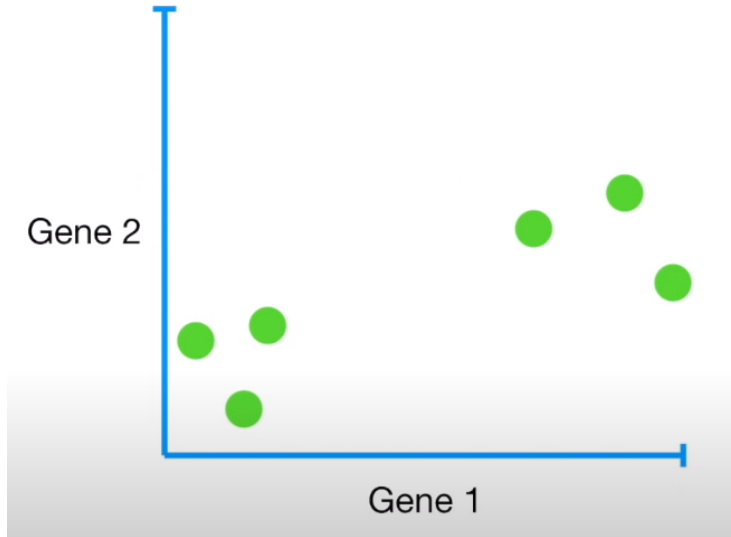
PCA

There are as many principal components as there are variables in the data.

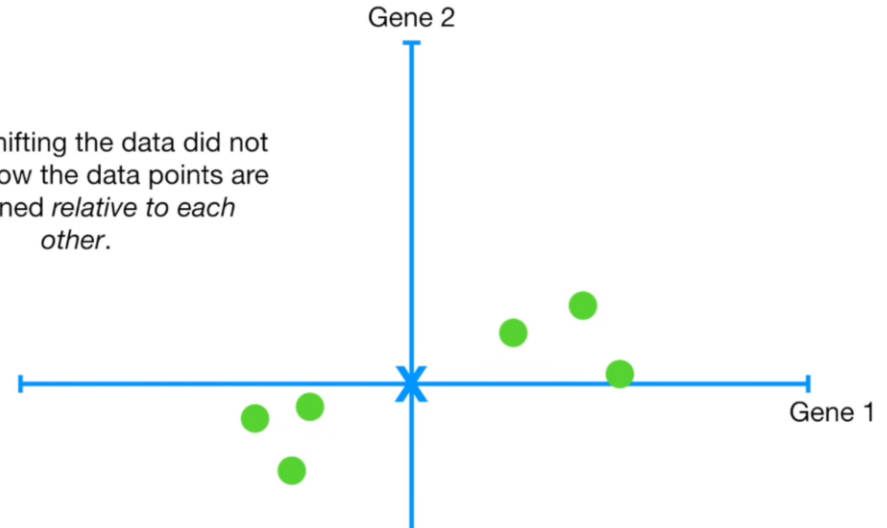


PCA

- **Step 1:** Centering the data (Normalization)

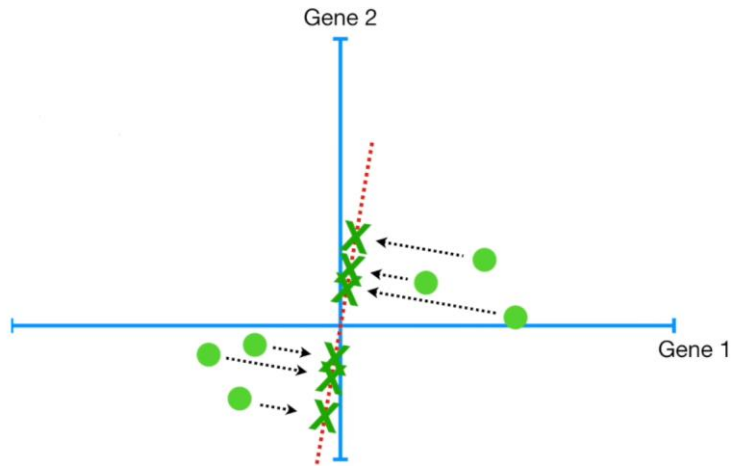


NOTE: Shifting the data did not change how the data points are positioned *relative to each other*.

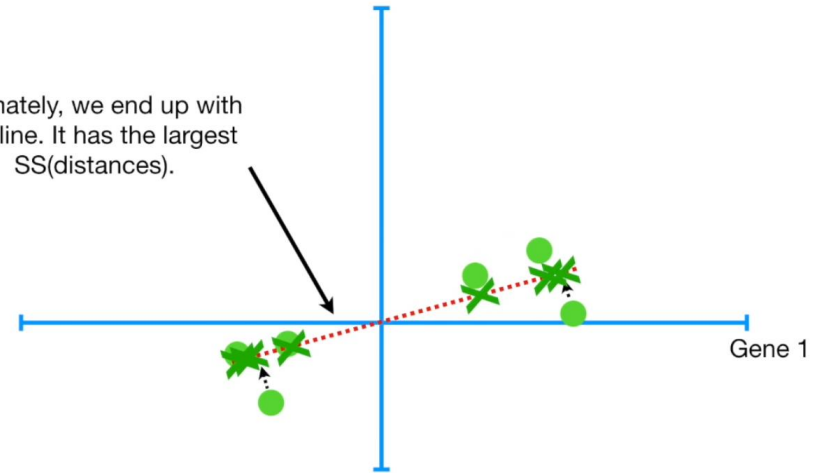


PCA

- **Step 2:** Make a line that maximizes the Sum of Squared distances.

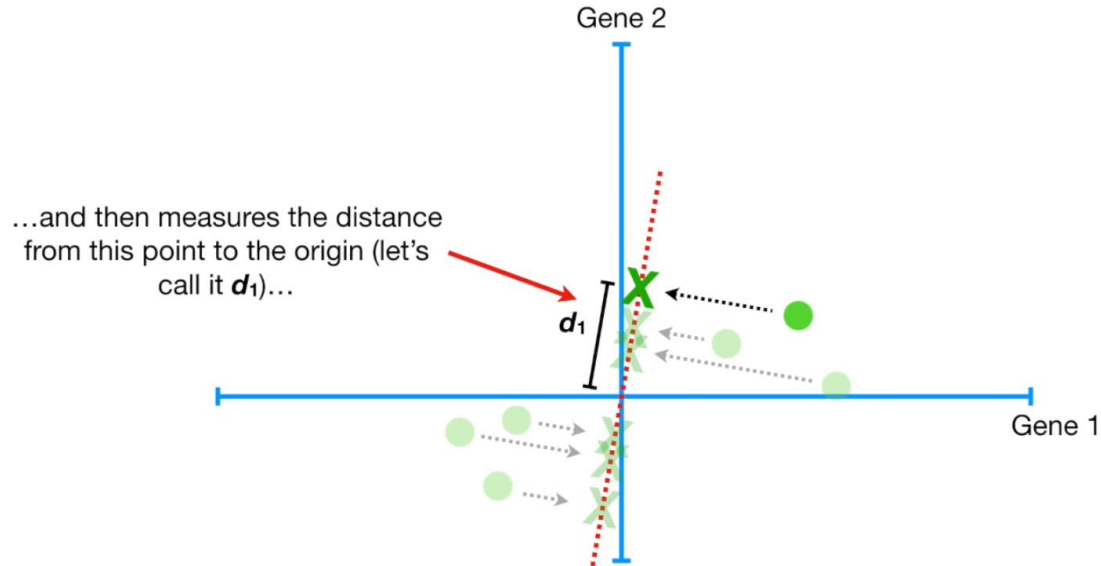


Ultimately, we end up with this line. It has the largest $SS(\text{distances})$.



PCA

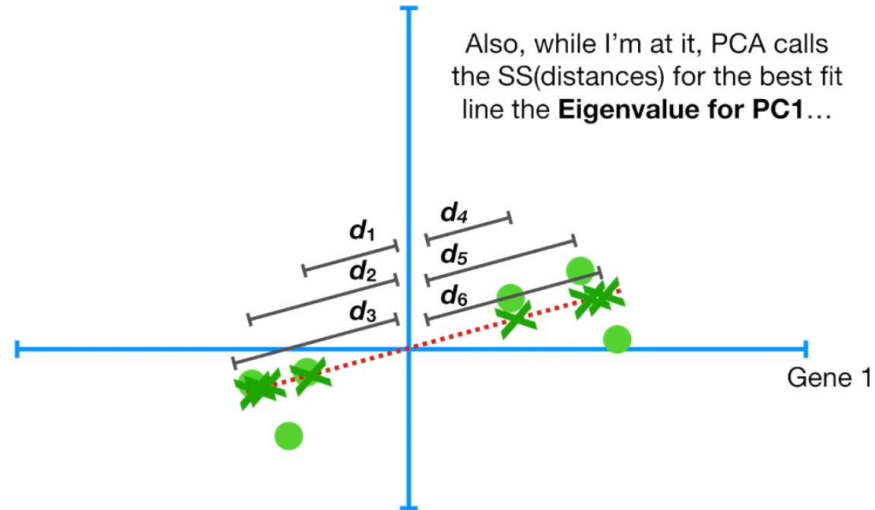
Distance (d_1) : The distance from the projected value to the center.



PCA

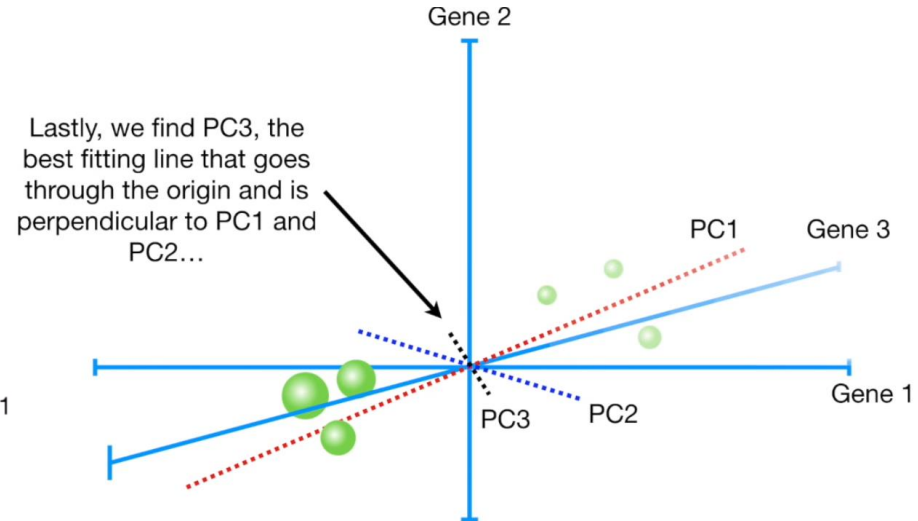
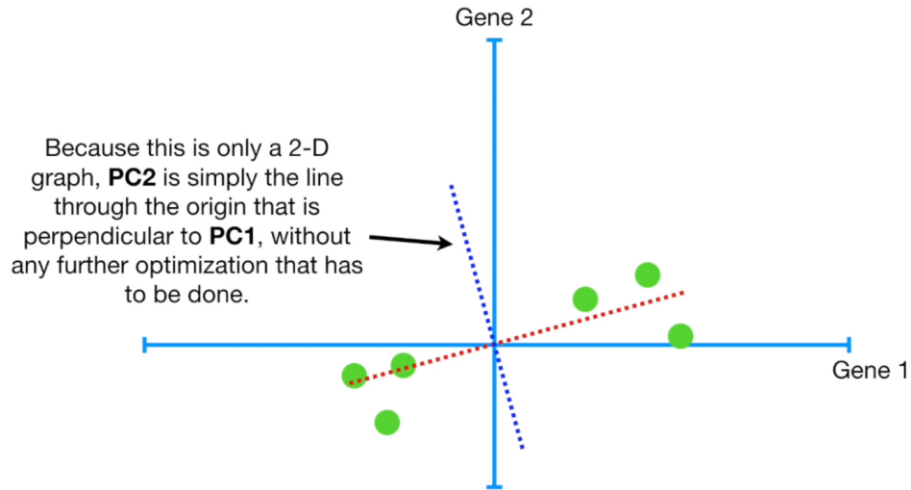
Sum of Squared distances: Called Eigenvalue.

$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2 = \text{sum of squared distances} = \text{SS}(\text{distances})$$



PCA

Second PCA: For any other PCA is simply a line through the origin that is perpendicular to the rest of PCAs.



PCA

VPC = Variation for PC1

SSPC1 = SS(distances for PC1)

n = Number of data points

PVE = Percentage of Variance Explained

$$VPC1 = \frac{SSPC1}{n - 1}$$

$$PVE = \frac{VPC1}{Sum\ of\ VPC}$$

PCA

Let's suppose that we have the next values of PVE

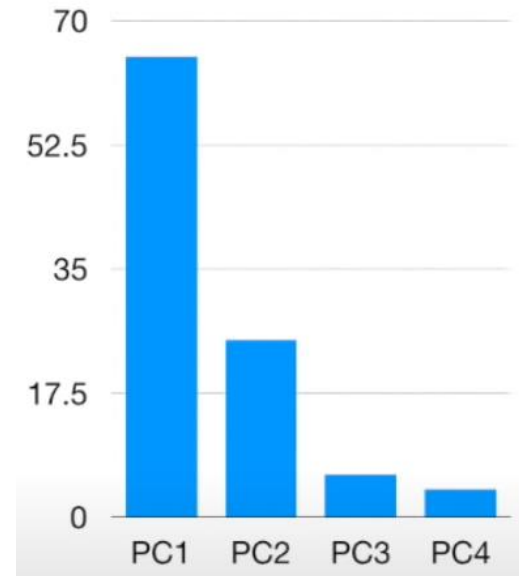
PC1 = 69%

PC2 = 21%

PC3 = 6%

PC4 = 4%

**With those values we can use just the PC1 and PC2
Instead of the 4 variables.**



References:

- [StatQuest: Principal Component Analysis \(PCA\), Step-by-Step](#)
- [A One-Stop Shop for Principal Component Analysis](#)
- [A STEP BY STEP EXPLANATION OF PRINCIPAL COMPONENT ANALYSIS](#)
- [Principal component analysis \(PCA\): Explained and implemented](#)

Disclaimers

Disclaimer: The content of this document is strictly informative and does not constitute an offer or recommendation of Casa de Bolsa Finamex S.A.B. de C.V. to buy, sell or subscribe any kind of securities, or to perform specific transactions. Casa de Bolsa Finamex S.A.B. de C.V. is not responsible for the interpretation given to the information and /or content of this document. Casa de Bolsa Finamex S.A.B. de C.V. does not accept or will accept any liability for losses or damages resulting from investment decisions that would have been based on this document. The persons responsible for the preparation of this report certify that the opinions stated reflect their own point of view and do not represent the view of Casa de Bolsa Finamex S.A.B. de C.V. nor of its officials. The persons responsible for the preparation of this report receive no compensation for expressing their opinion in any particular issues mentioned in this document. The persons responsible for the preparation of this report did not receive any compensation from entities other than Casa de Bolsa Finamex S.A.B. de C.V. This document is based on publicly available information, including BMV, Bloomberg, SI Bolsa, official government sites and other web sources which are considered reliable, however Casa de Bolsa Finamex S.A.B. de C.V. makes no warranty regarding its accuracy or completeness.