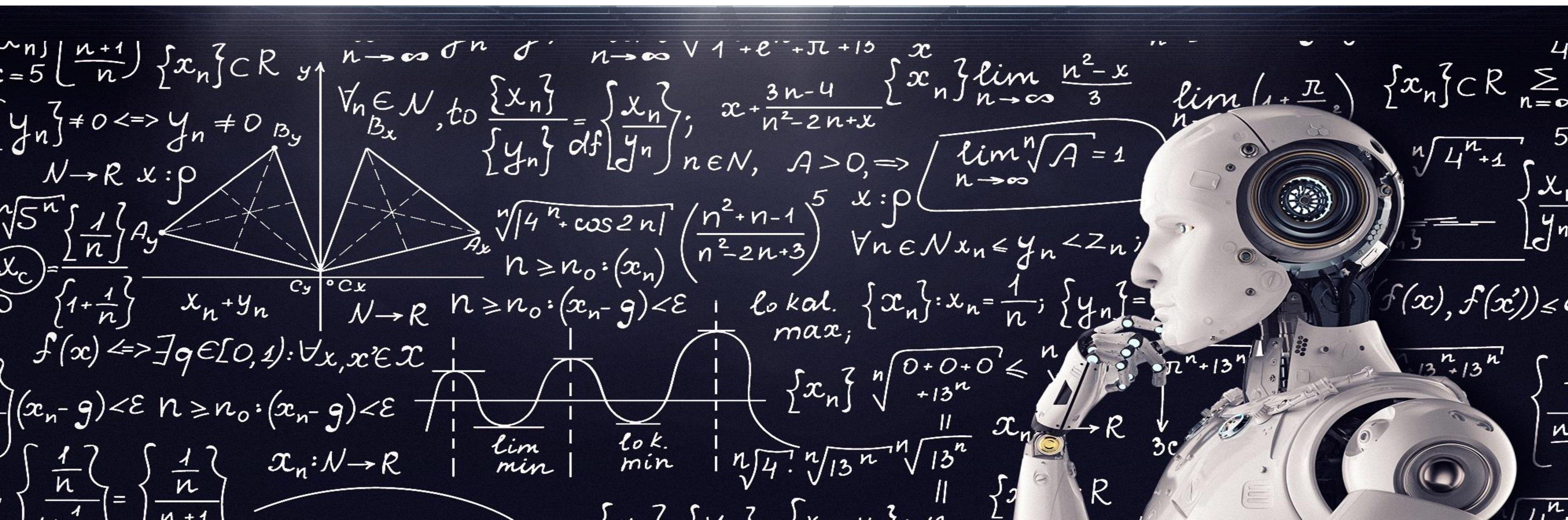




Análisis de Componentes Principales



PCA

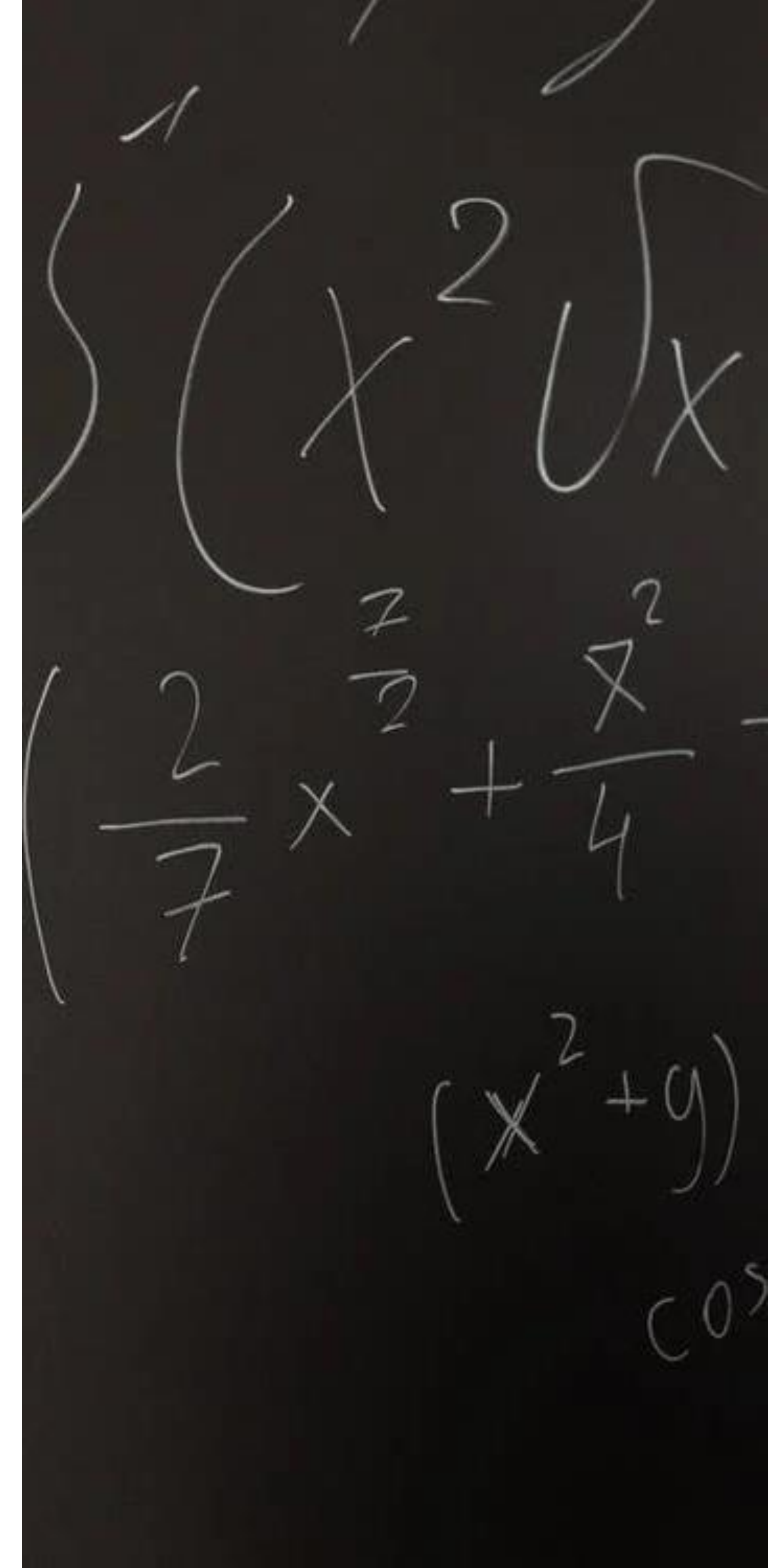




Desarrollo de conceptos

Qué preguntas nos puede ayudar a resolver?

- Se pueden resumir los datos con una menor cantidad de variables?
 - PCA - Un análisis de la canasta básica
 - Análisis factorial – Calidad de vida
- Qué grupos homogéneos podemos distinguir?
 - Conglomerados
- Si ya clasificamos los datos en 3 grupos. Para los nuevos datos cómo podemos clasificarlo en alguno de los 3 grupos y cuáles son las variables que más pesan?
 - Análisis discriminante.

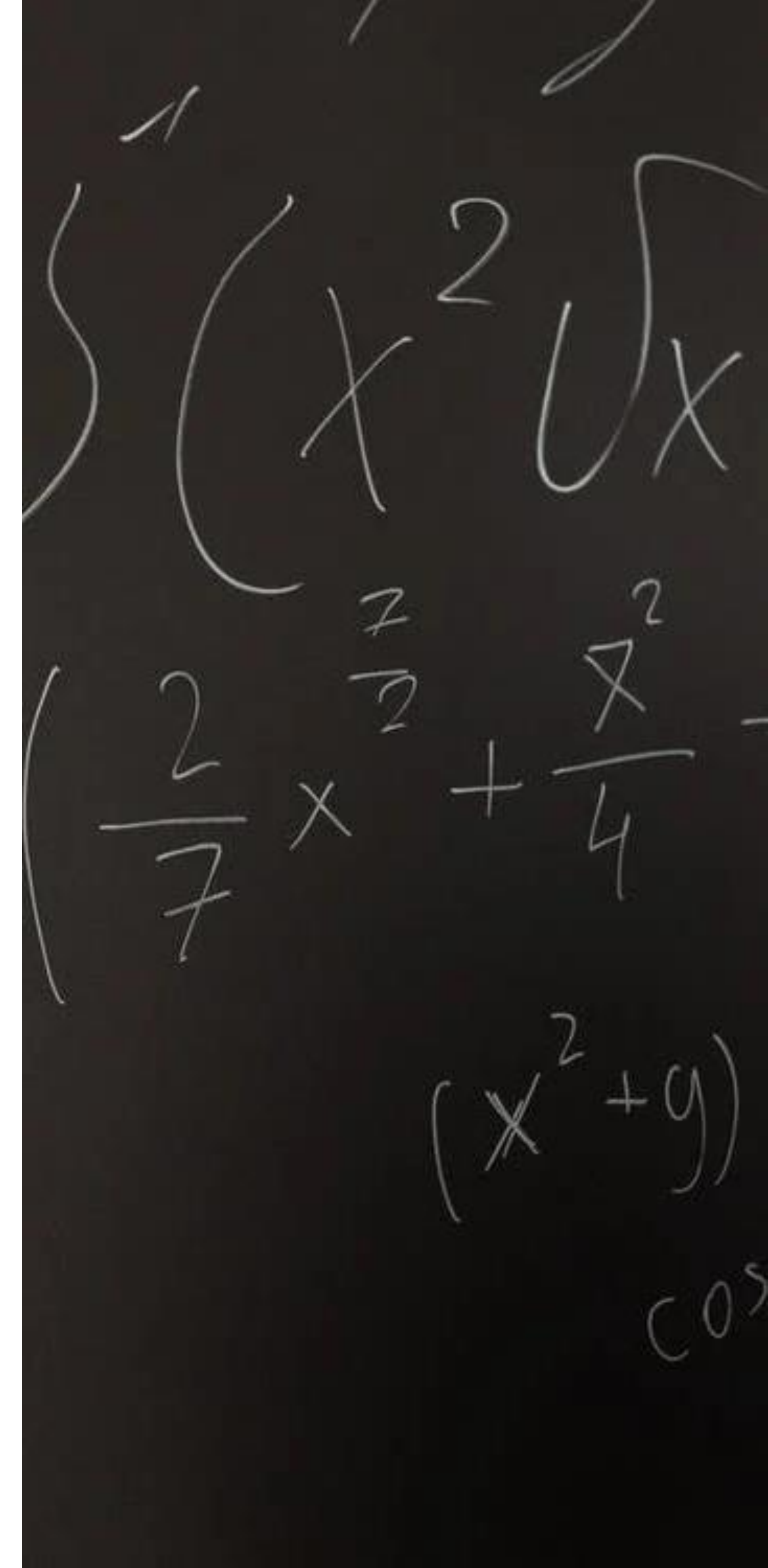




Desarrollo de conceptos

Características de PCA

- PCA – Principal Component Analysis
- Técnica de reducción de dimensiones
 - 1. Decorrelacionar
 - 2 . Reducir dimensión
- Hacer más eficiente el almacenamiento y procesamiento
- Remover las características (features) , que menos información aportan (ruido)
- Cuando tengamos problemas en modelos predictivos: Clasificación y regresión

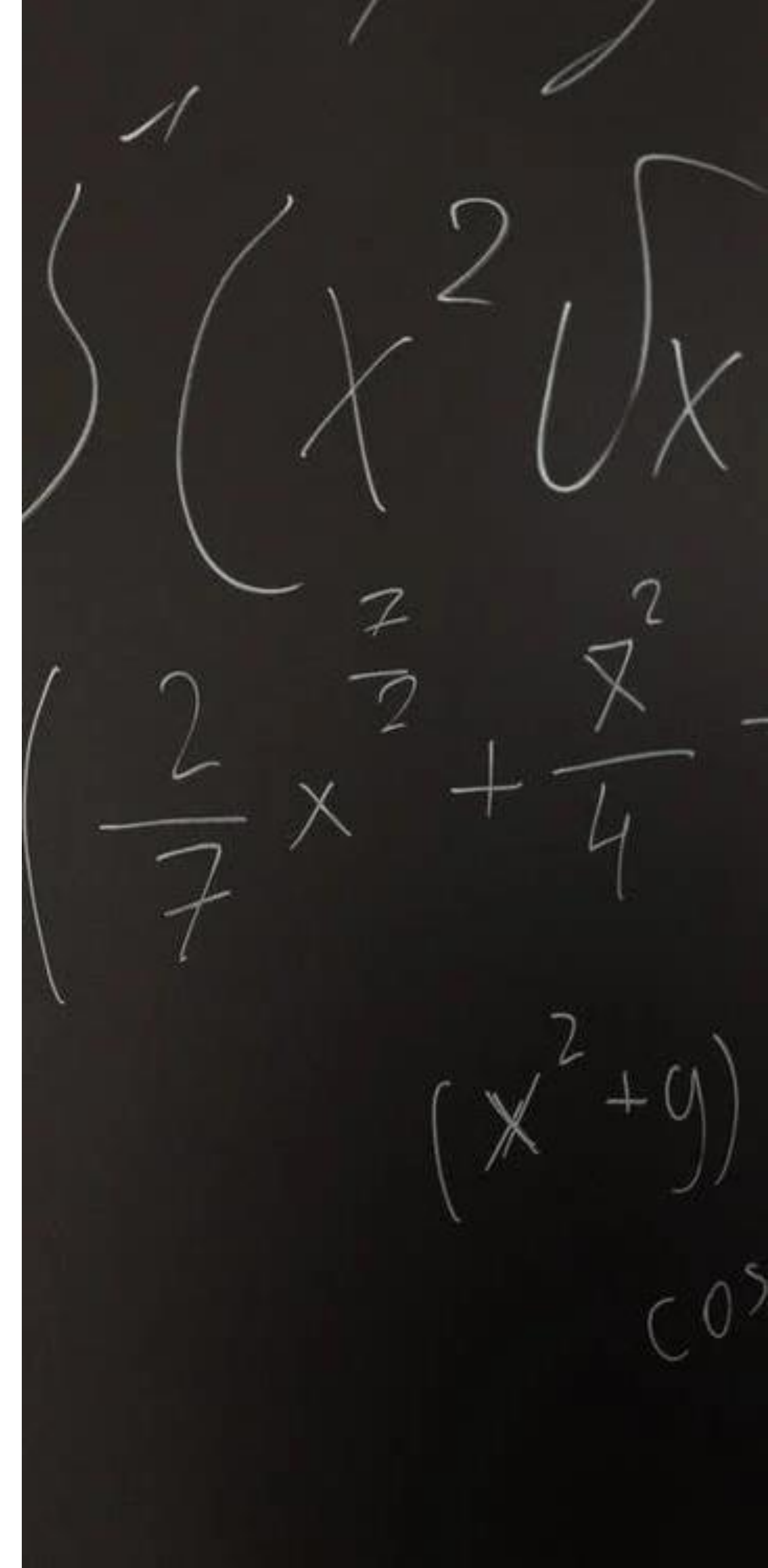




Desarrollo de conceptos

Características de PCA

- Matemáticamente, el PCA se basa en una descomposición de la matriz de covarianza de las variables a lo largo de las direcciones que mejor explican la principales causas de variabilidad de la información analizada.

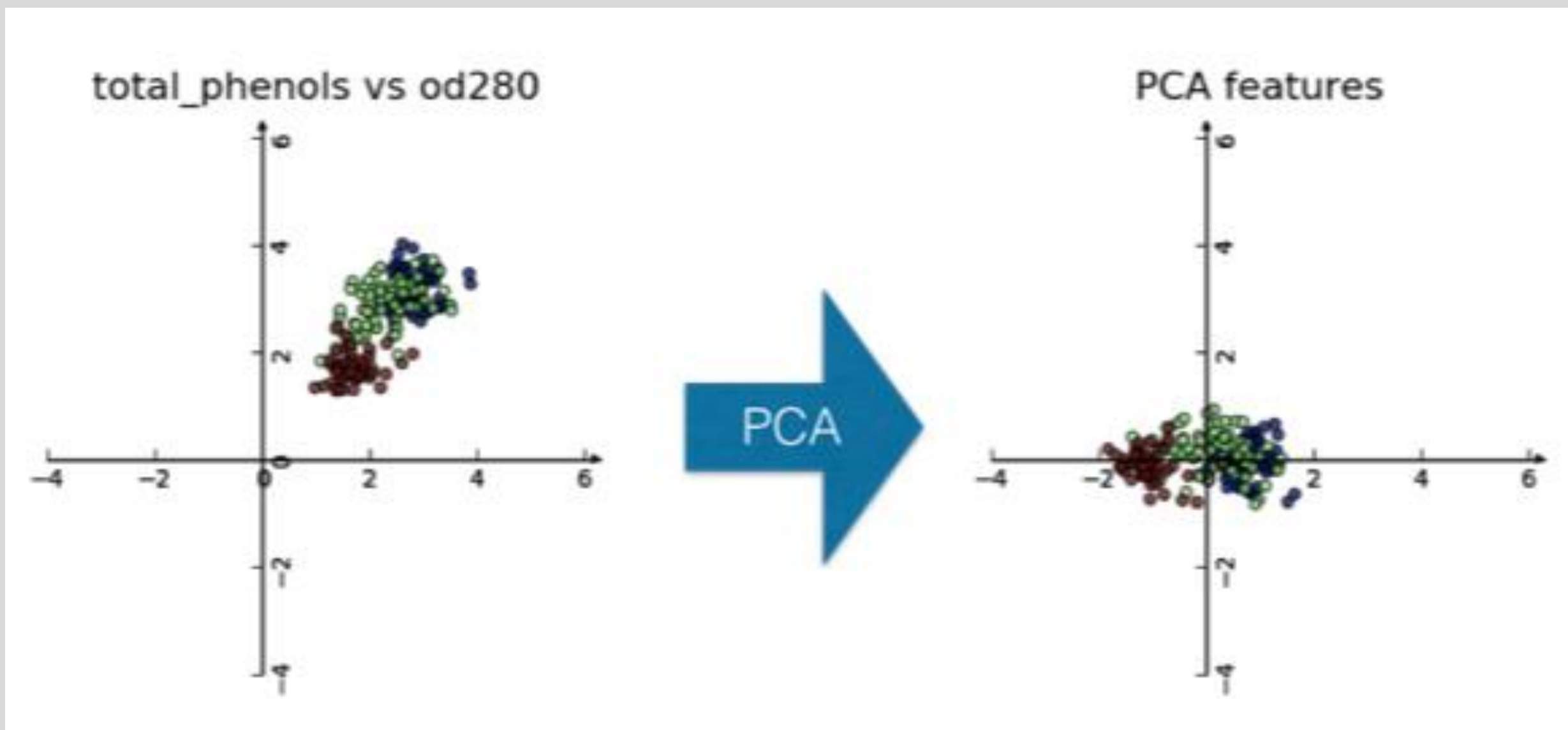




Desarrollo de conceptos

Características de PCA

- Rota los datos para alinearlos
- Cambia los datos para que tengan media 0



$$\int (x^2 \sqrt{x})$$

$$\left(\frac{2}{7} x^{\frac{7}{2}} + \frac{x^2}{4} \right)$$

$$(x^2 + y)$$

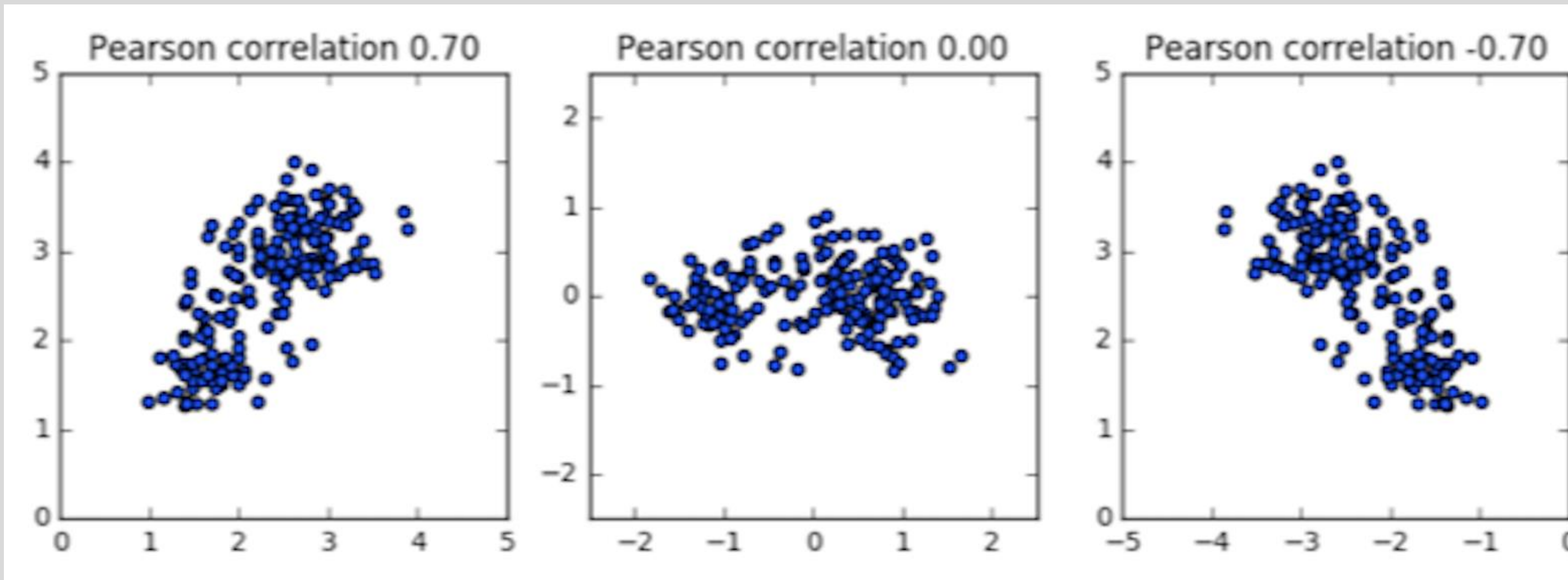
cos



Desarrollo de conceptos

Correlación de Pearson

- Correlación lineal de las características
- Valor entre -1 y 1
- Valor 0 no tiene una correlación lineal



$$\left(x^2 + \frac{2}{7}x + \frac{7}{4} \right)$$
$$\left(x^2 + 9 \right)$$
$$\cos$$



Desarrollo de conceptos

Qué se necesita?

Multidimensionalidad

Correlacionados

Matriz de correlaciones – Pearson o Spearman

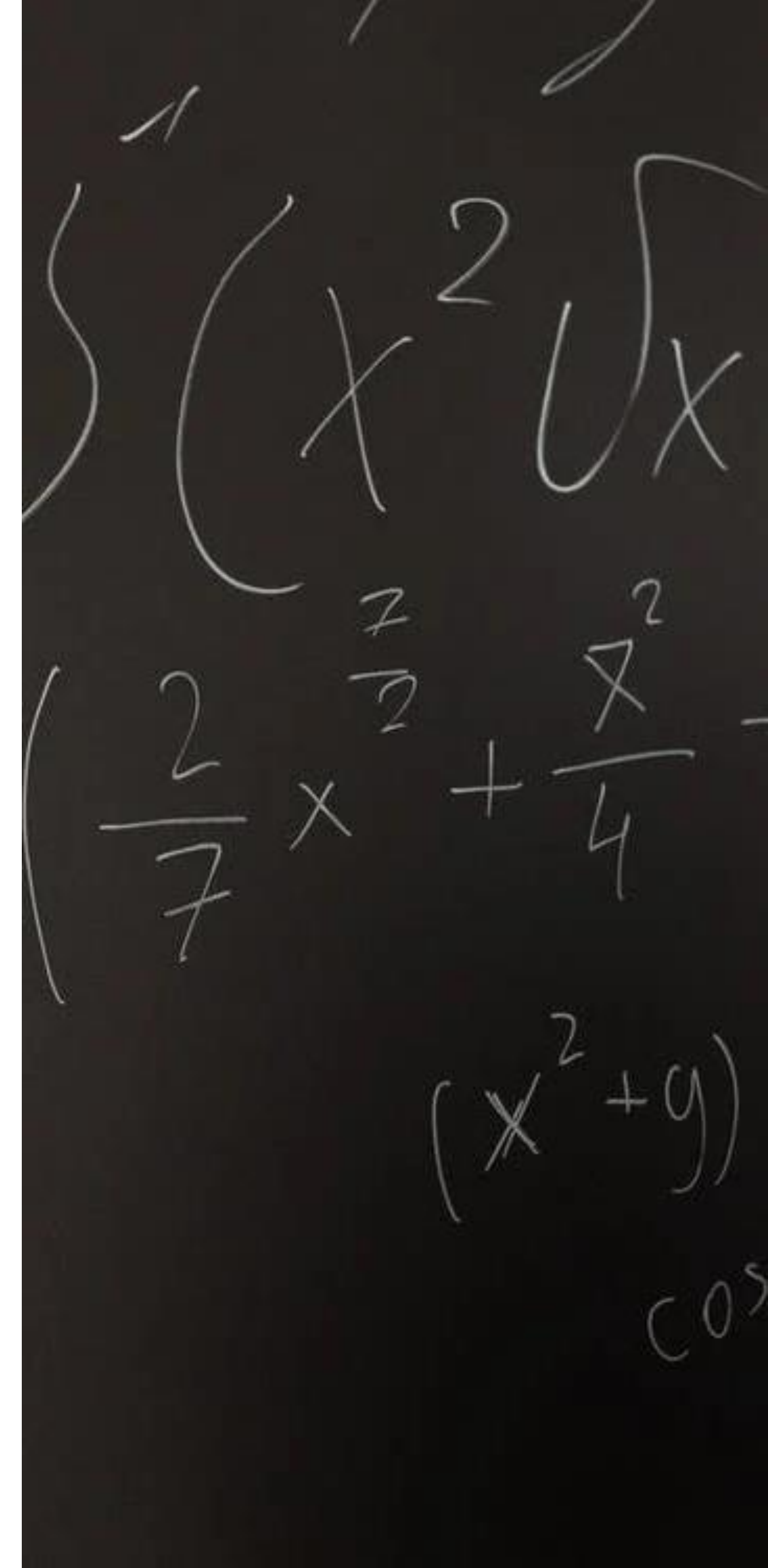
Qué se hace?

Combinar variables, cada combinación (Componente Principal CP)

CP1 – Explica el 60% de la varianza

CP2 – Explica un 30%

La relación entre los CP es ortogonal

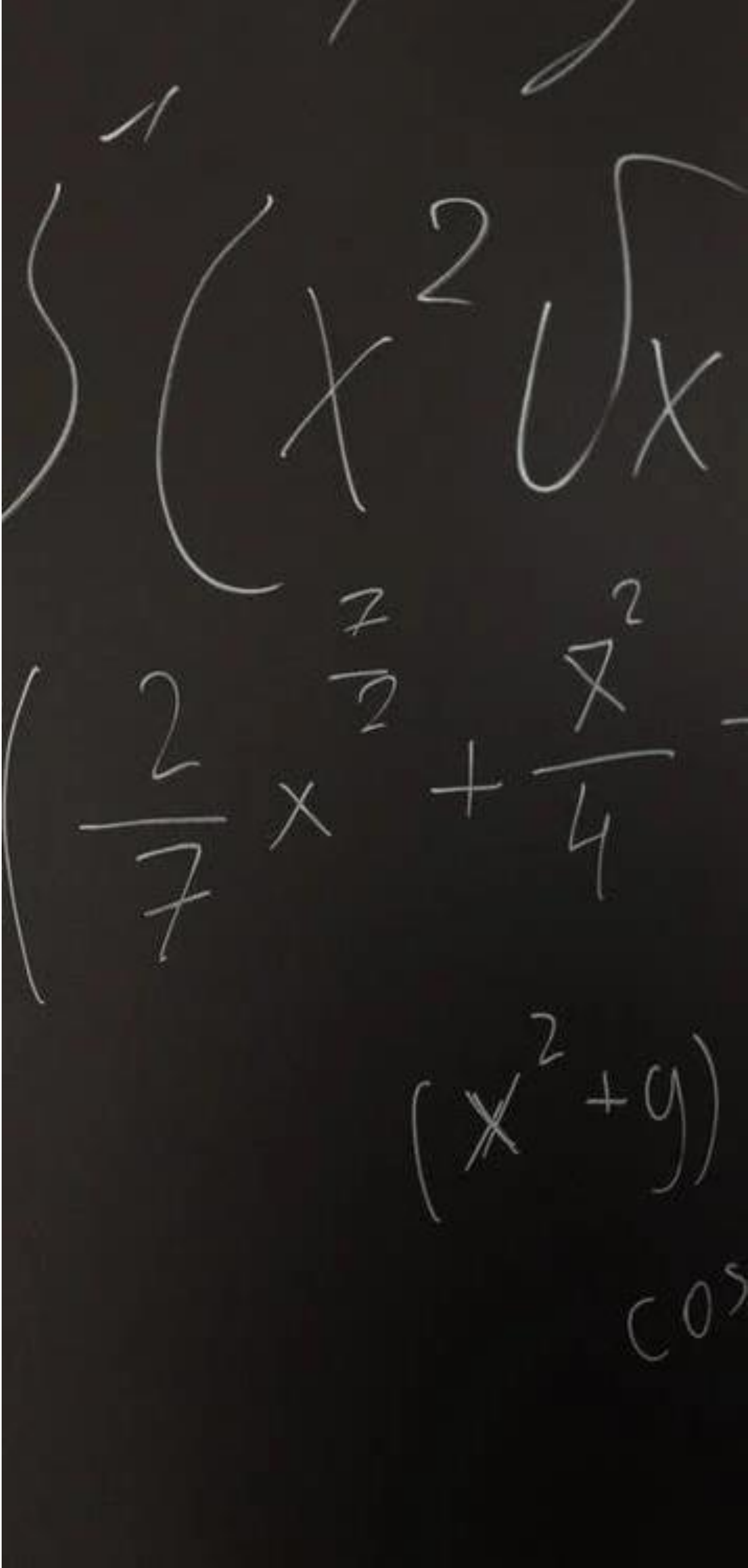




Desarrollo de conceptos

PCA

Filosofía	Ética	Física	Cómputo
6	7	9	10
7	7	10	9
10	8	7	8
9	10	6	7
8	9	7	6



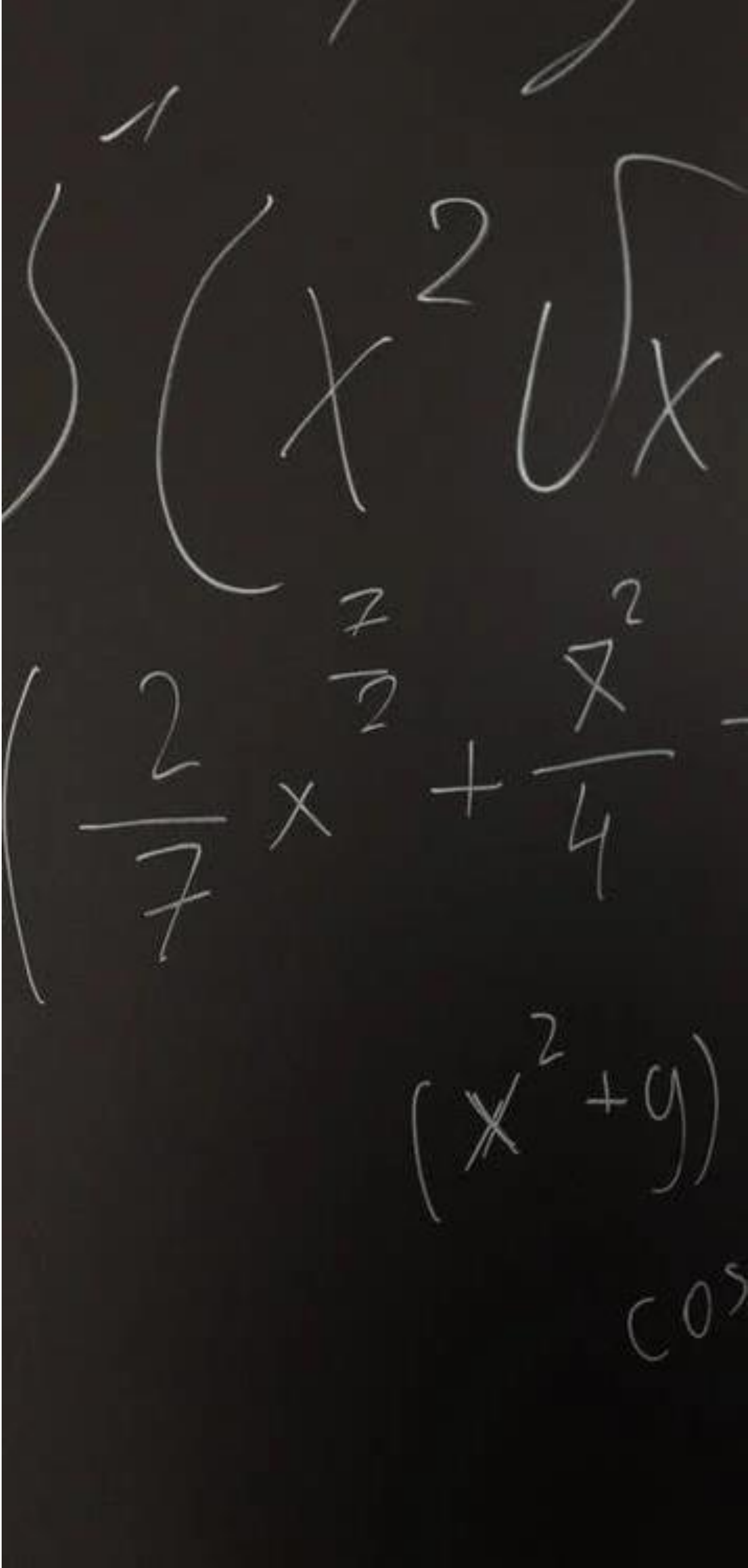


Desarrollo de conceptos

PCA

Filosofía	Ética		CP1
6	7		6.5
7	7		7
10	8		9
9	10		8.5
8	9		8.5

Física	Cómputo		CP2
9	10		9.5
10	9		9.5
7	8		7.5
6	7		6.5
7	6		6.5



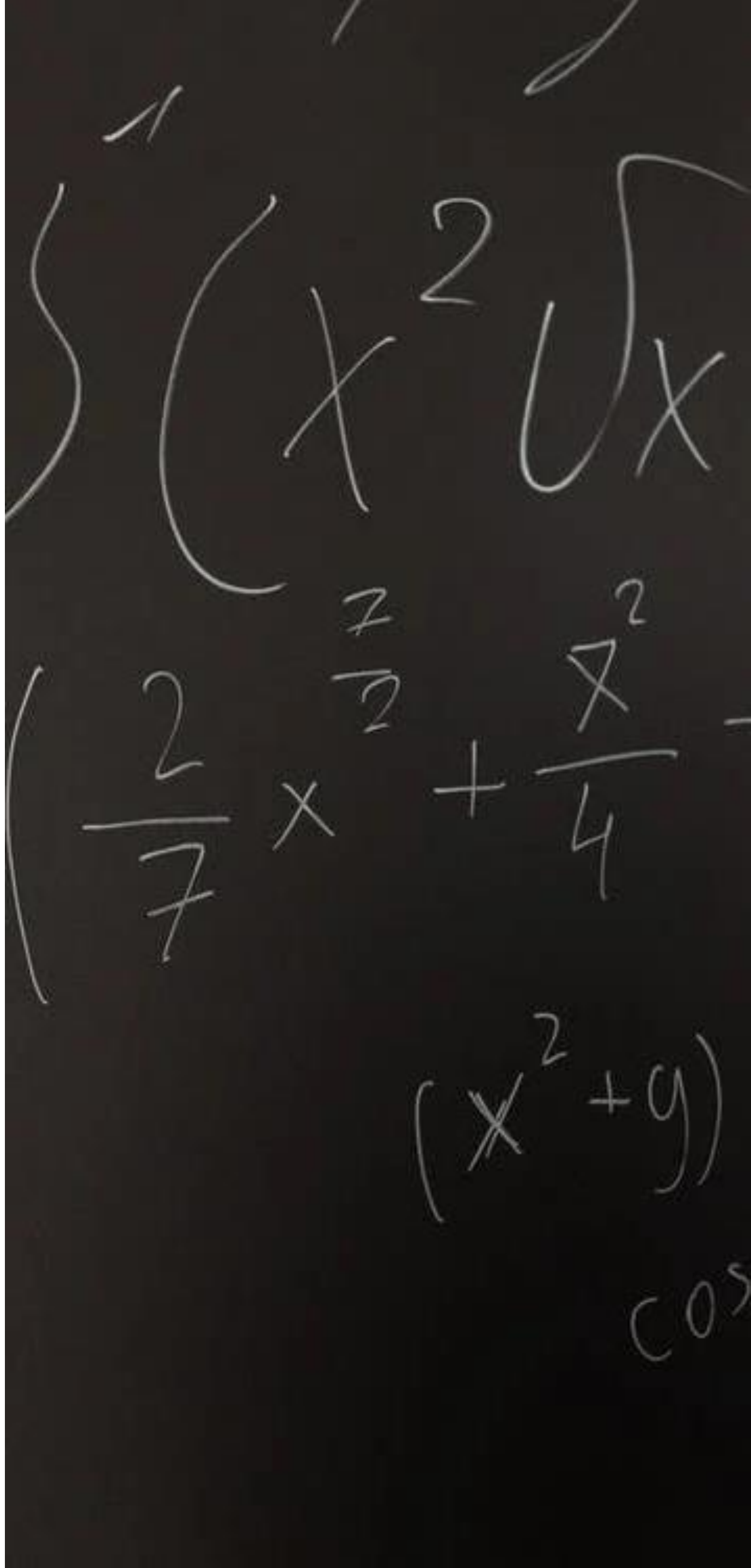
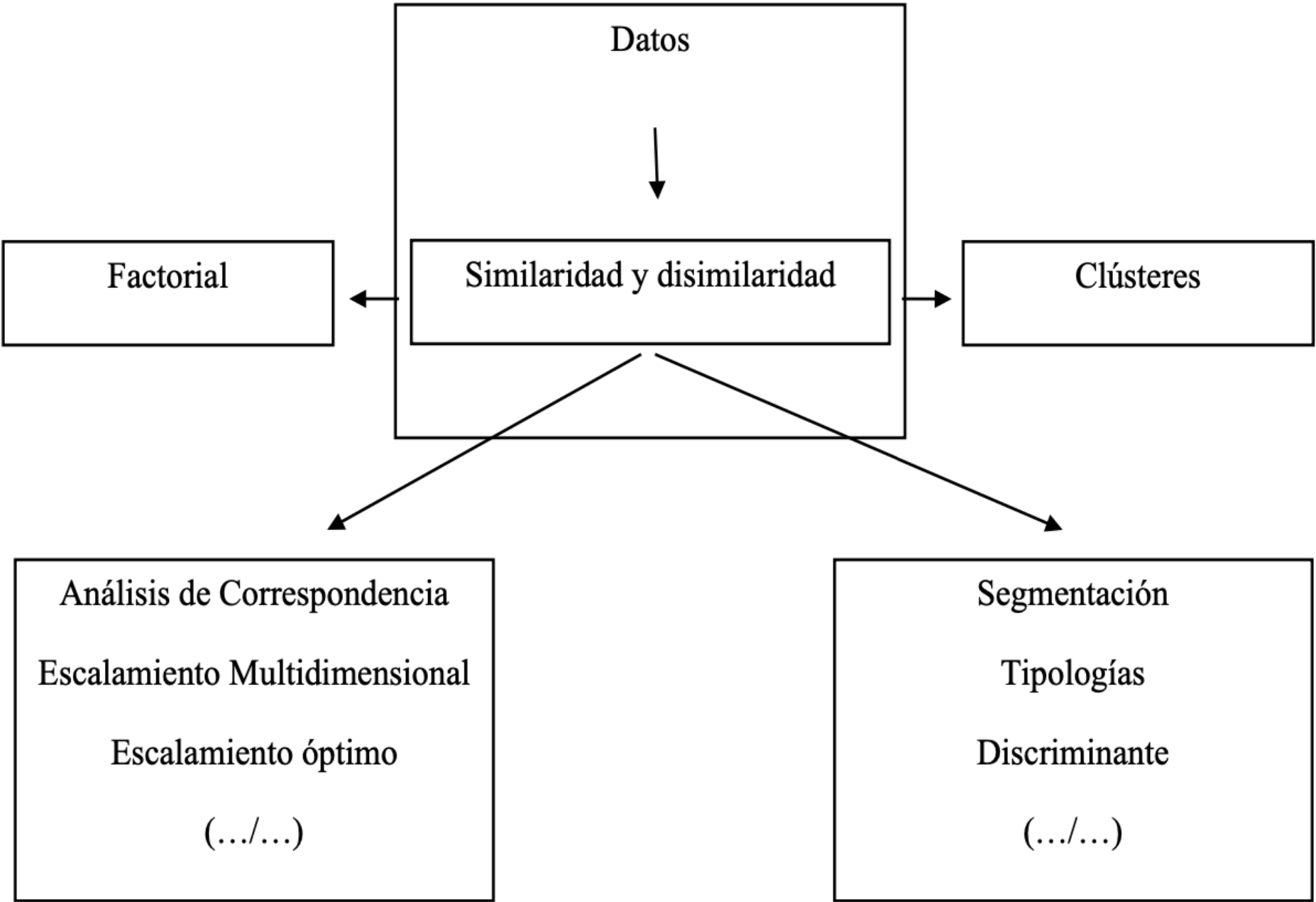


Desarrollo de conceptos

Modelos basados en la distancia

[Alaminos 2015]

Cuadro 1. Ejemplos de modelos basados en la determinación de distancias





Solución

- Crear Cuenta en Colab
 - <https://colab.research.google.com>
- Descargar Orange 3
 - <https://orangedatamining.com>





Bibliográficos

- Introduction to machine learning - Ethem Alpaydın
- [Galeano-Arias 2020] - <https://scielo.conicyt.cl/pdf/infotec/v32n1/0718-0764-infotec-32-01-17.pdf>
- An Introduction to Multivariate Statistical Analysis (Wiley Series in Probability and Statistics) (T. W. Anderson)
- [Alaminos 2015] Análisis multivariante para las Ciencias Sociales, Antonio Alaminos Chica
- D. Garcia-Alvarez, Estudio comparativo de técnicas de detección de fallos basadas en el Análisis de Componentes Principales (PCA), <https://www.sciencedirect.com/science/article/pii/S1697791211000070>
- <https://www.python.org/doc/>
- <https://realpython.com/world-class-companies-using-python/>
- <https://www.anaconda.com>
- <https://jupyter.org>
- <https://numpy.org>
- <https://pandas.pydata.org>

Recursos



Dataset

<https://archive.ics.uci.edu/ml/datasets/wine>

https://jdvelasq.github.io/courses/notebooks/sklearn_dataset_utilities/2-11_load_wine.html

Notas Python

<https://github.com/erlinares/mlengineer/tree/main/cheatsheets>

Software

- Python
- Colab
- Orange



Bibliográficos

- <https://www.python.org/doc/>
- <https://realpython.com/world-class-companies-using-python/>
- <https://www.anaconda.com>
- <https://jupyter.org>
- <https://numpy.org>
- <https://pandas.pydata.org>