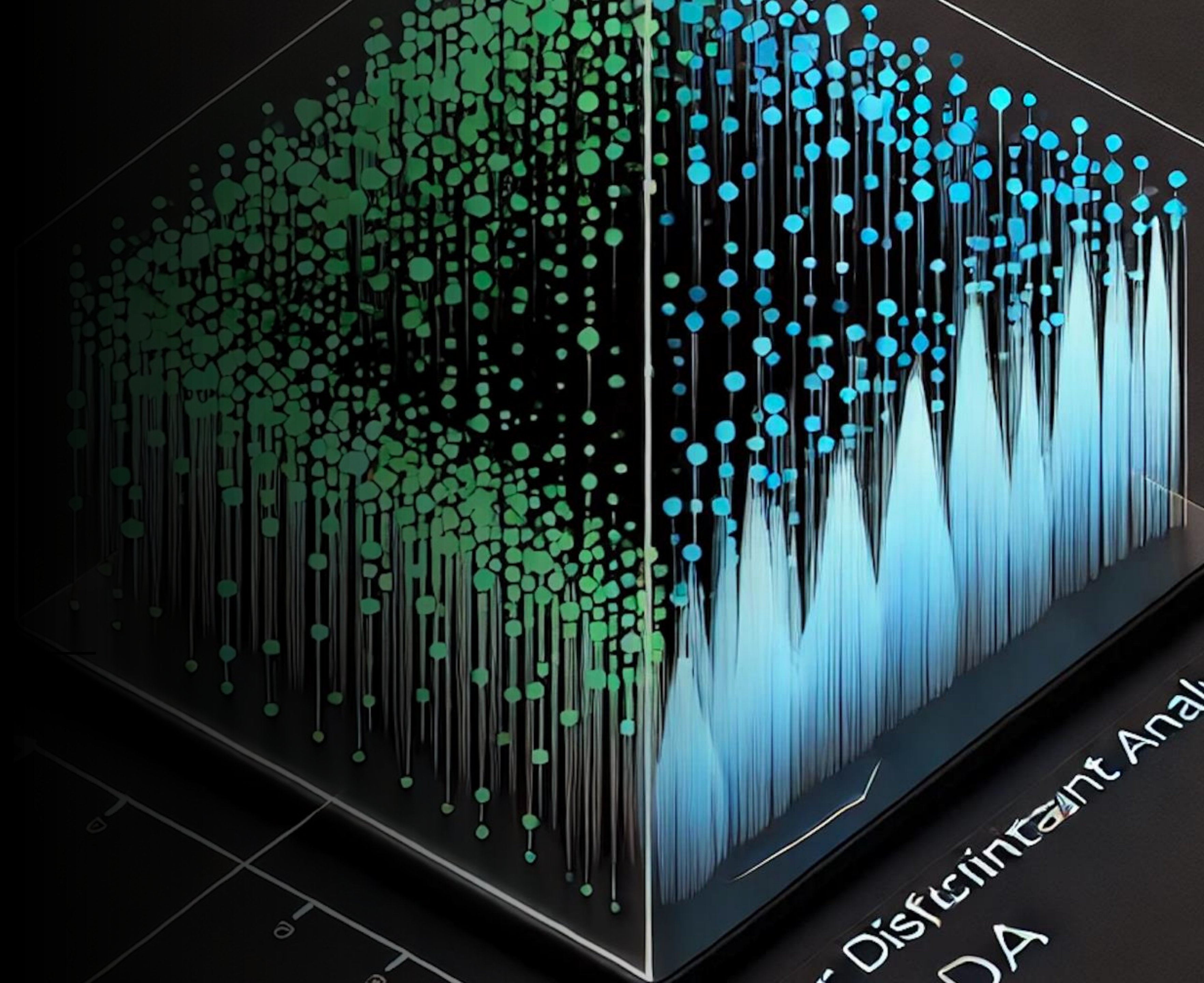




# Análisis discriminante Lineal



Discriminant Analysis  
DA





# Objetivo

## Análisis discriminante

- Comprender el análisis discriminante lineal LDA.
- Resolver un problema utilizando python y un conjunto de datos de UCI Machine Learning Repository.

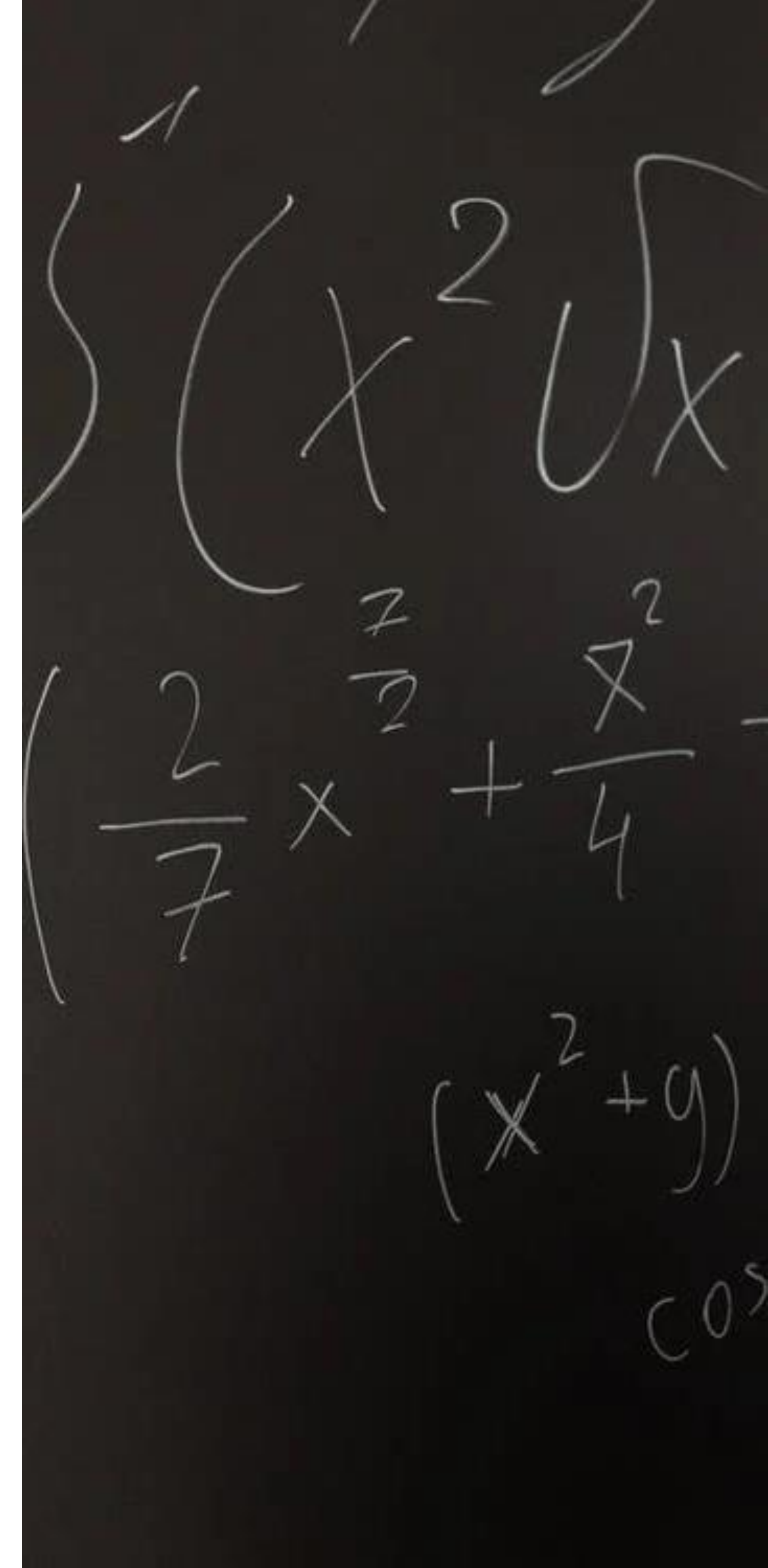




# Desarrollo de conceptos

## Que és?

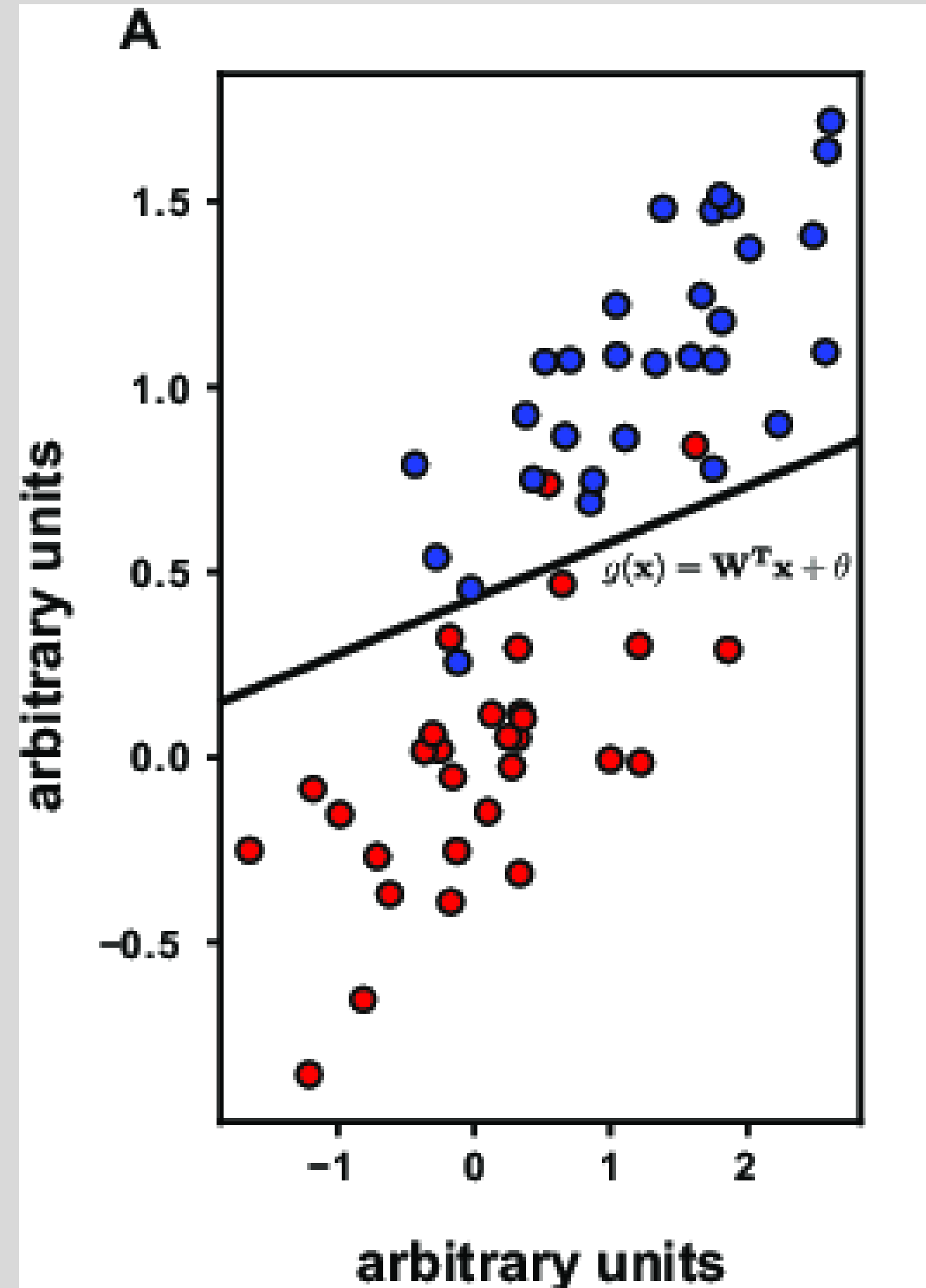
- Es un método estadístico de reconocimiento de patrones
- Aprendizaje de máquina supervisado
- Generalización del discriminante lineal de Fisher
- Análisis de la varianza (ANOVA)
- Busca una combinación lineal de ramos que separen las clases
- Explicar la variable dependiente como la combinación lineal de las variables independientes





# Desarrollo de conceptos

Que és?



$$\left( x^2 + \frac{2}{7}x + \frac{1}{4} \right)$$
$$\left( x^2 + g \right)$$
$$\cos$$

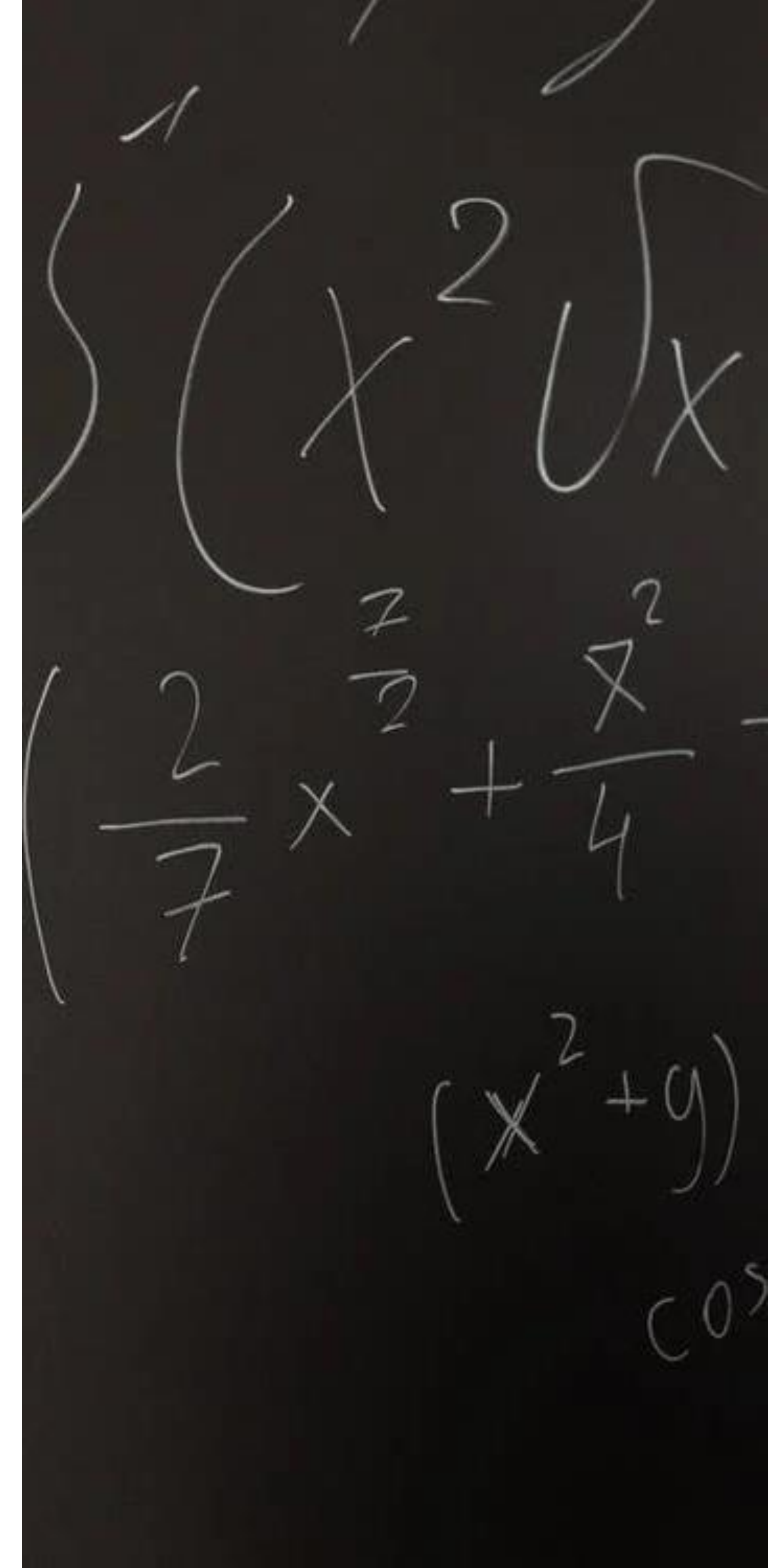




# Desarrollo de conceptos

## Utilidad

- Reducir las dimensiones de los datos
- Crear un modelo predictivo supervisado

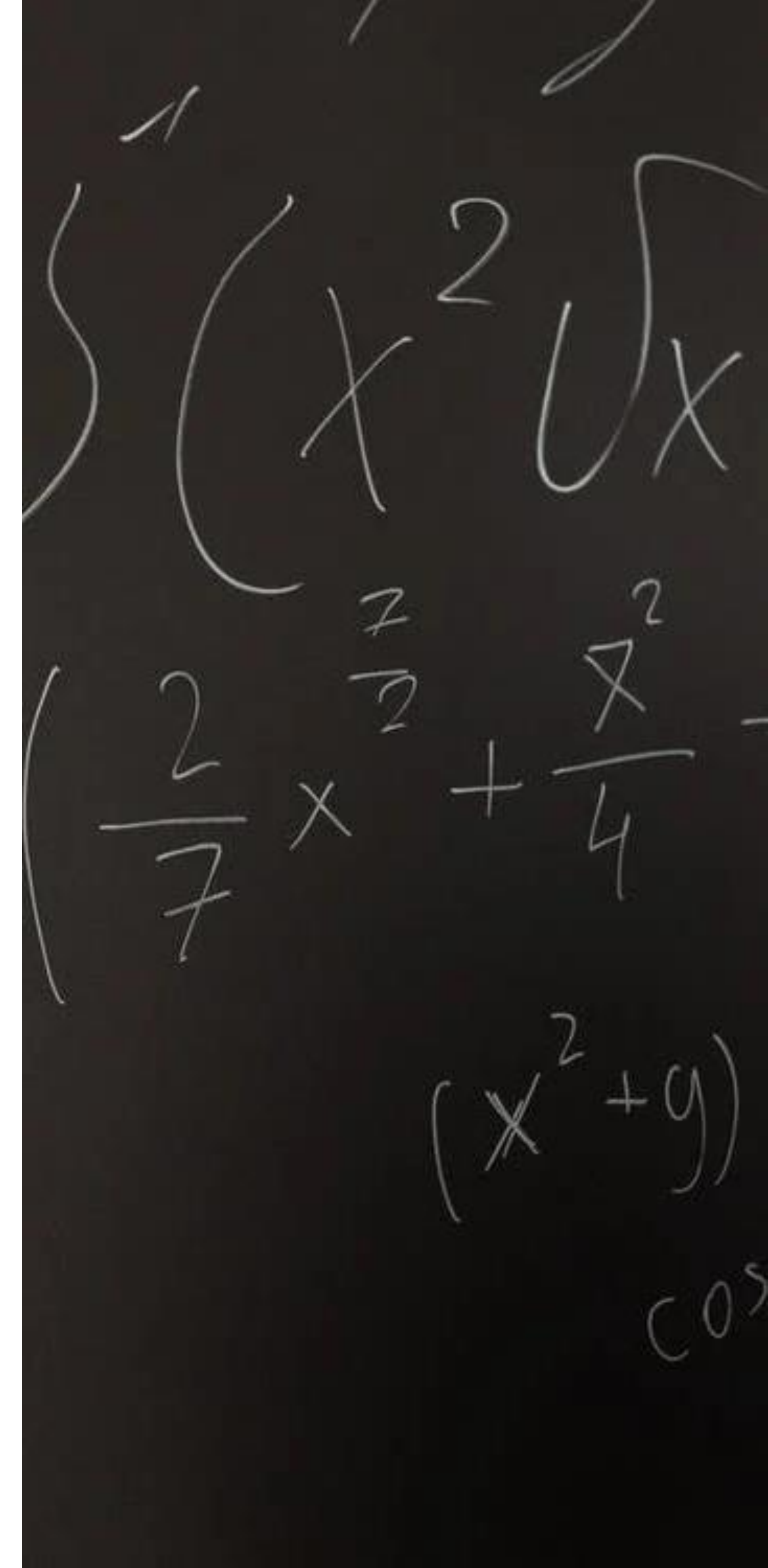




# Desarrollo de conceptos

## Casos reales

- Predicción de quiebra de empresas - Edward Altman
- Reconocimiento facial - "caras de Fisher"
- Marketing
  - Tipos de clientes
  - Tipos de productos
- Estudios biomédicos
  - Ligera, moderada, severa

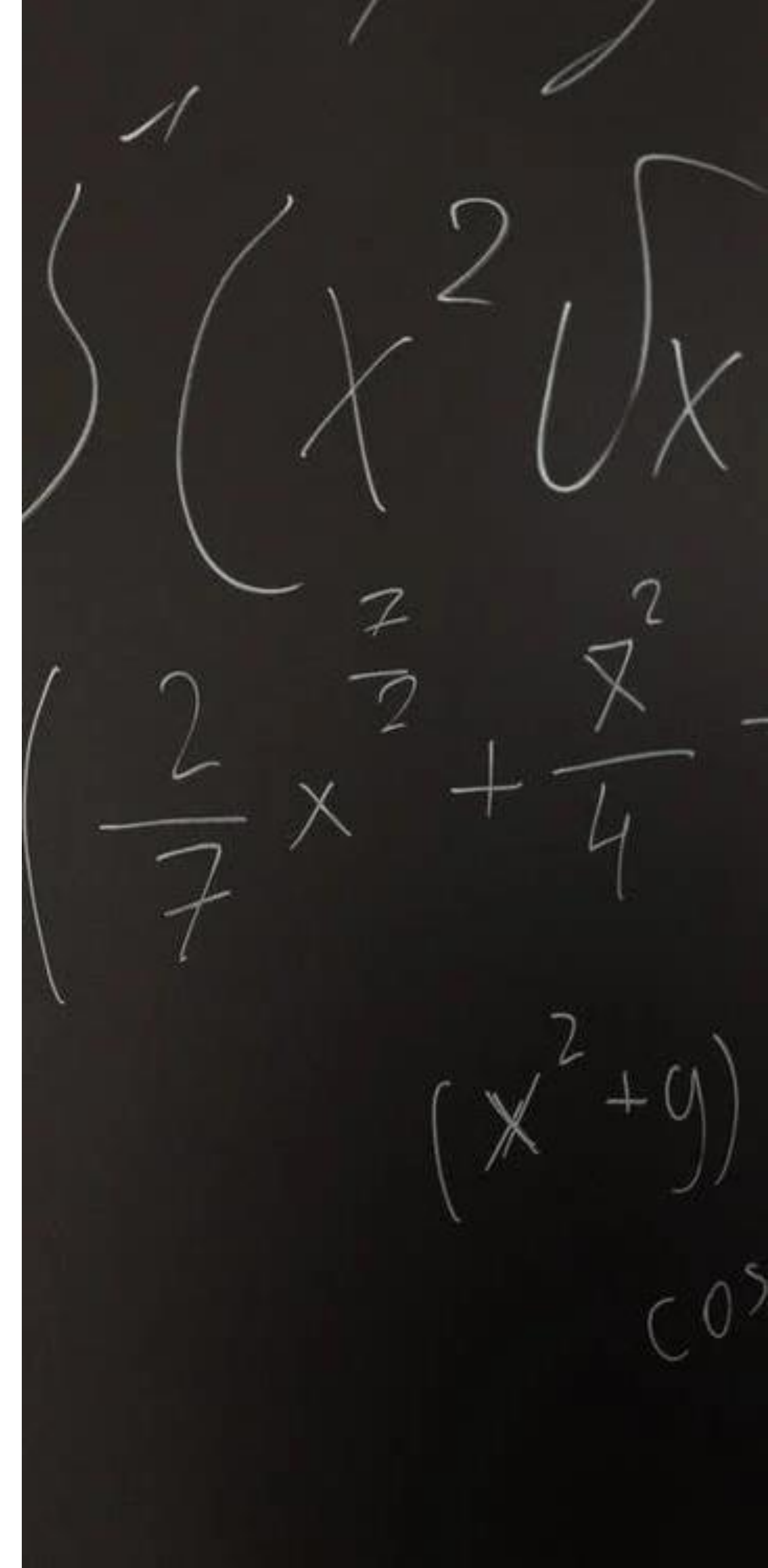




# Desarrollo de conceptos

## Utilidad

- Predecir a que categoría pertenece un nuevo elemento
- Identificar las características que diferencian o discriminan uno o más grupos. Función discriminante.
- Determinar la significancia de cada variable independiente sobre la variable dependiente.
- Que conjunto de variables diferencian los grupos





# Desarrollo de conceptos

## Procedimiento

- Se toma una muestra de observaciones de cada grupo
- Cada observación tiene un conjunto de variables
- Se contrasta si las medias y/o matrices de covarianzas entre los grupos son iguales
- Se estima la probabilidad de pertenencia a un grupo para futuras observaciones usando las variables discriminantes

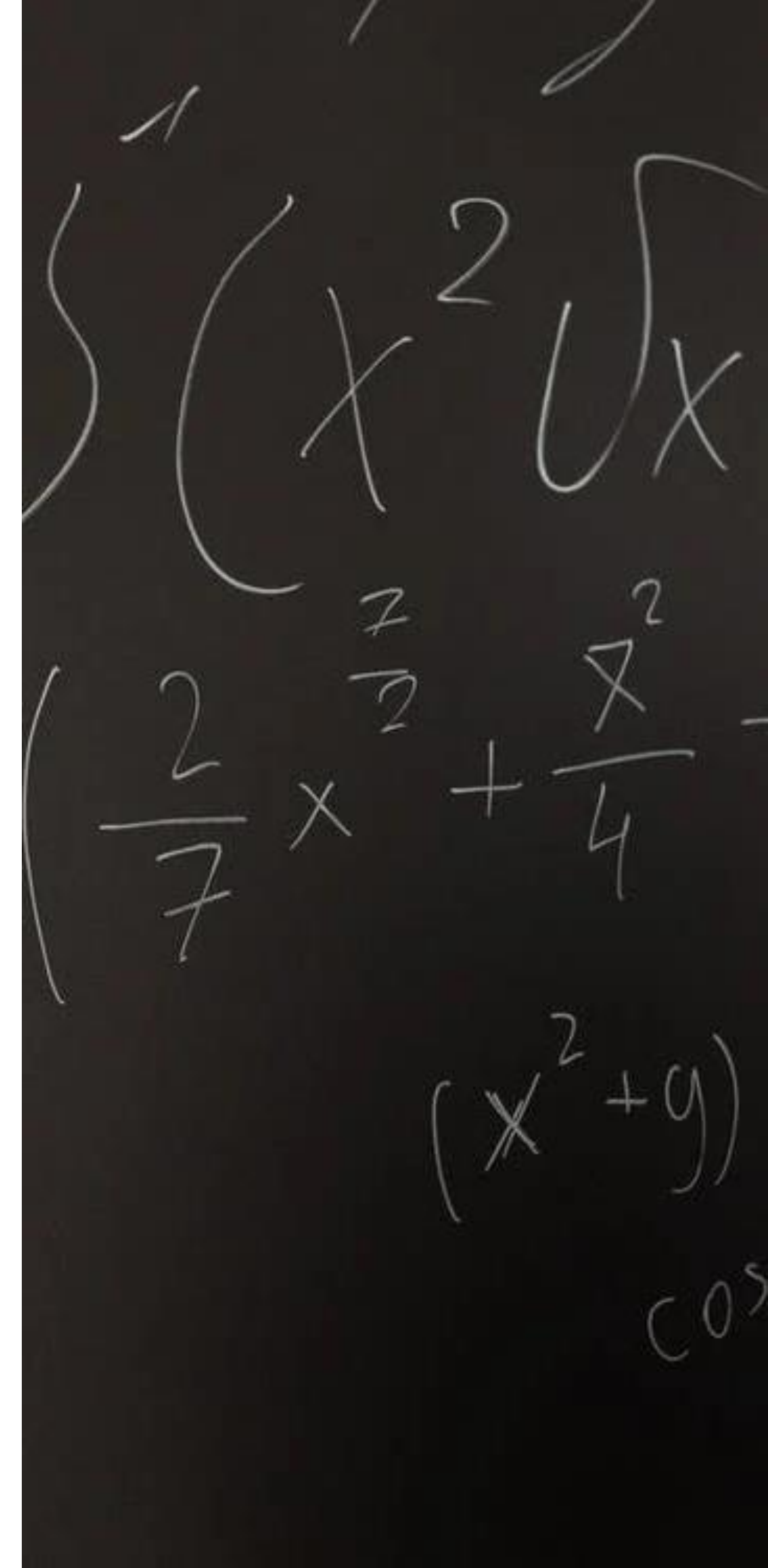
Asume grupos conocidos

Modelo lineal

distribuciones normales

Matrices de covarianzas iguales

- Técnica de análisis exploratorio



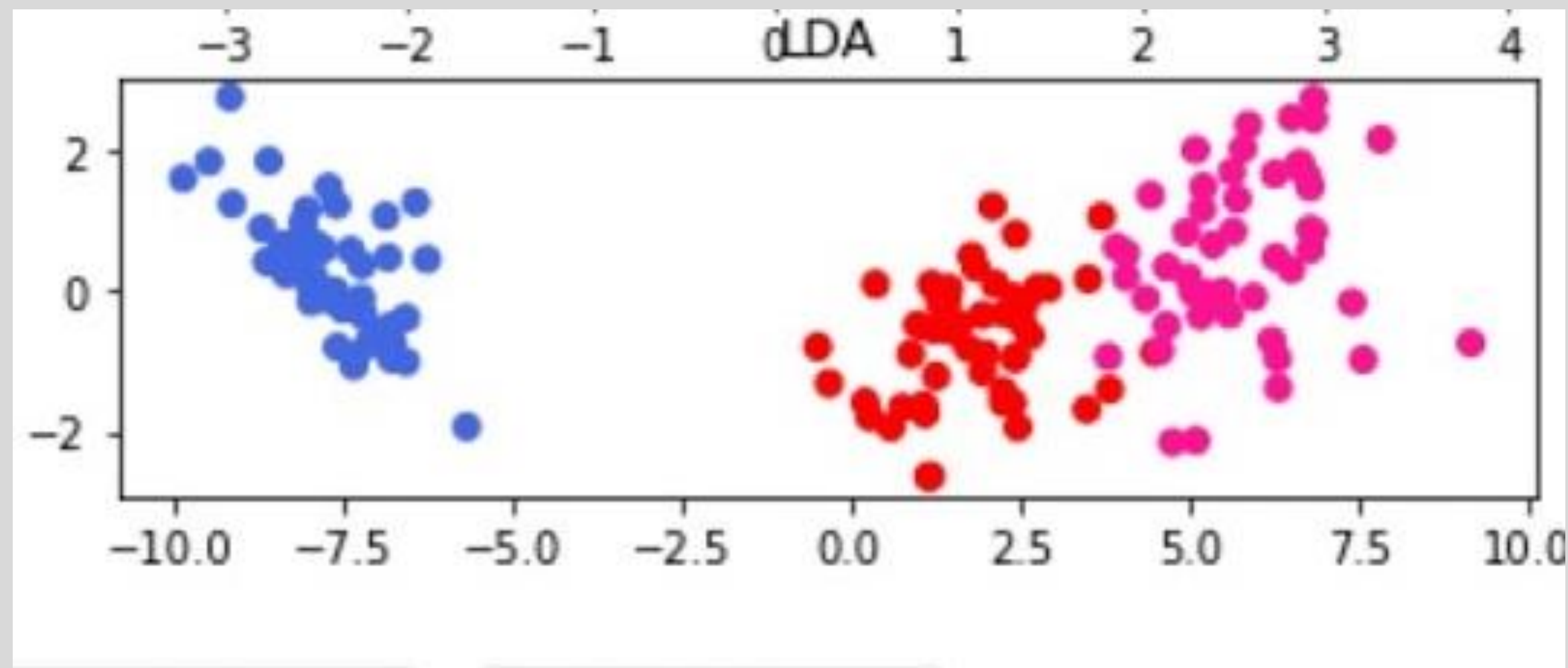




# Desarrollo de conceptos

# Base matemática

- **n** variables independientes
- **p** nuevas variables
- **$p < n$**



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

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

$$(x^2 + y)$$

cos



# Desarrollo de conceptos

## Base matemática

- **Variables discriminantes**, son las variables que nos ayudan a diferenciar a los grupos
- **Funciones discriminantes**, predicción de pertenencia utilizando ecuaciones

Handwritten mathematical equations on a chalkboard:

- $$\sqrt{x^2 + 1}$$
- $$\left( \frac{2}{7} x^{\frac{7}{2}} + \frac{7^2}{4} \right)$$
- $$(x^2 + 9)$$
- $$\cos$$





# Desarrollo de conceptos

## Variables

- Útiles
- Redundantes

Handwritten mathematical expressions on a chalkboard:

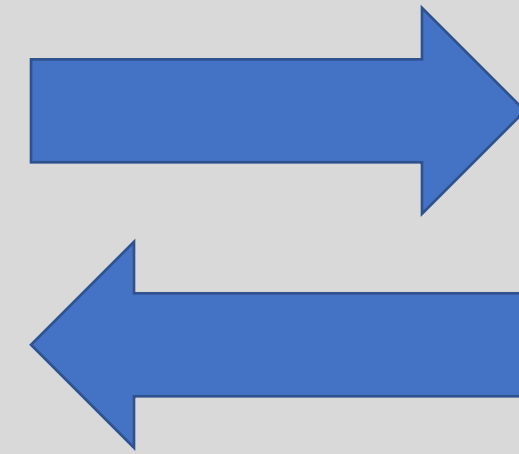
- $$\sqrt{x^2}$$
- $$\left( \frac{2}{7} x^{\frac{7}{2}} + \frac{7^2}{4} \right)$$
- $$(x^2 + 9)$$
- $$\cos$$



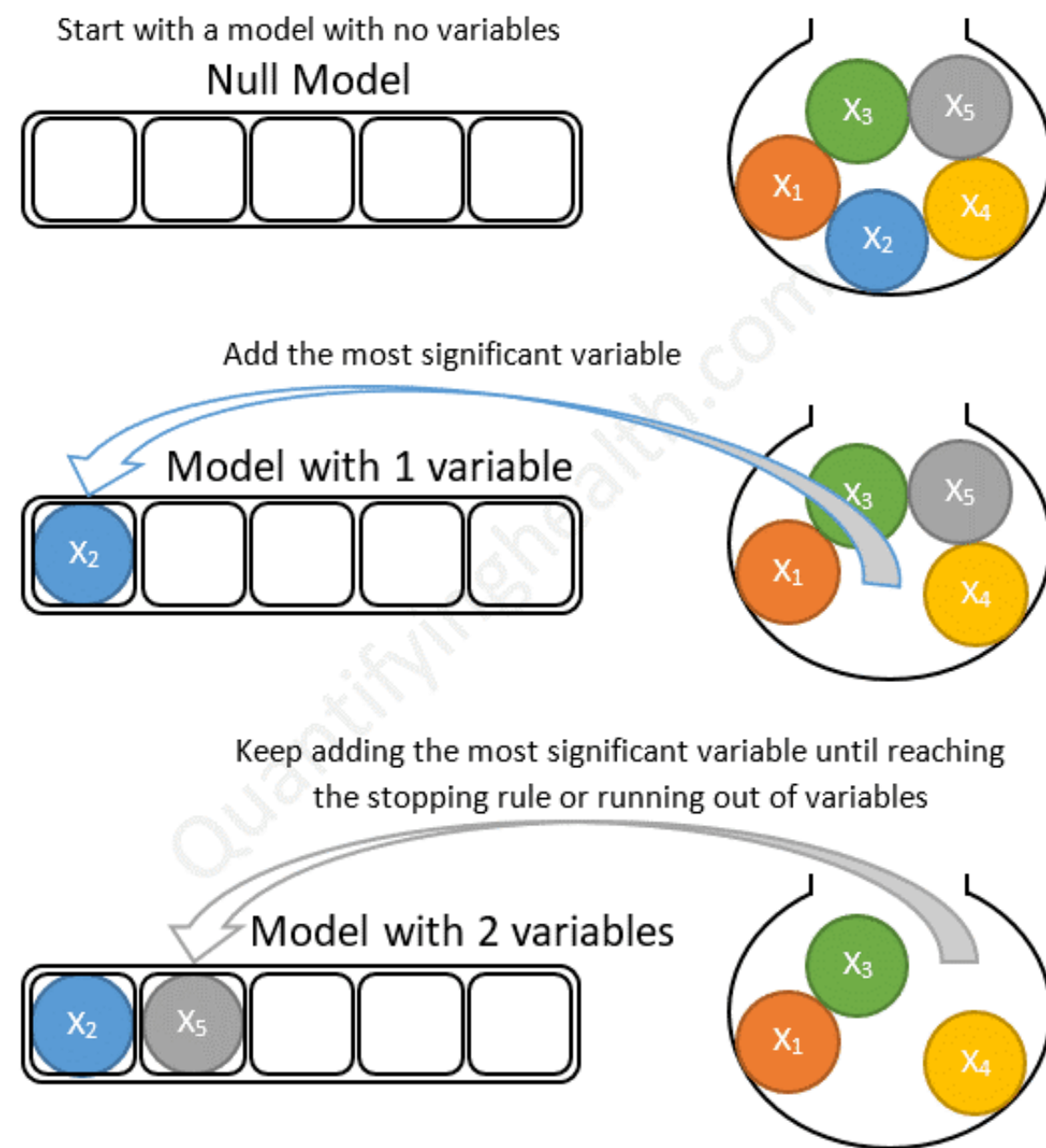
# Desarrollo de conceptos

## Métodos

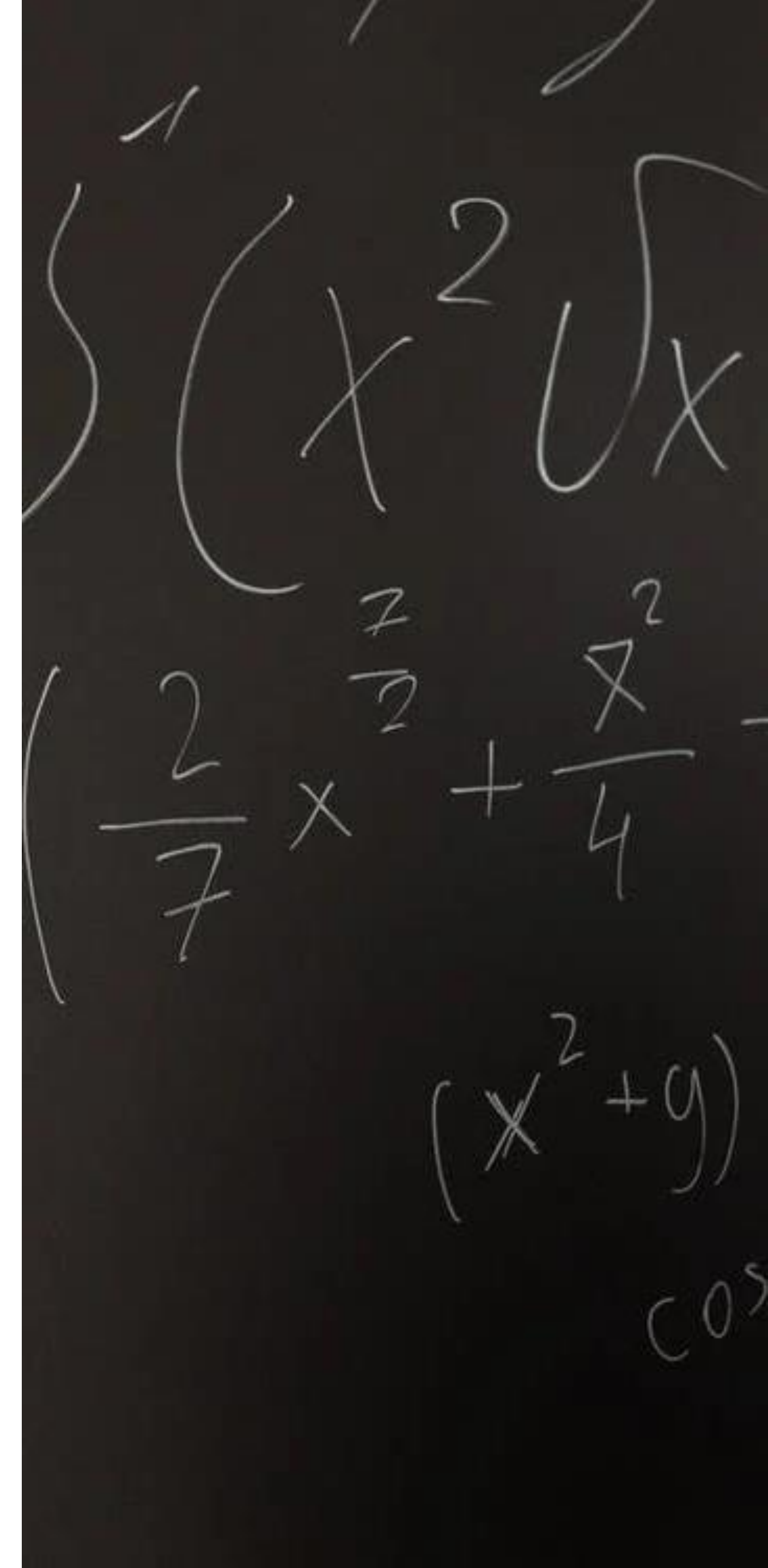
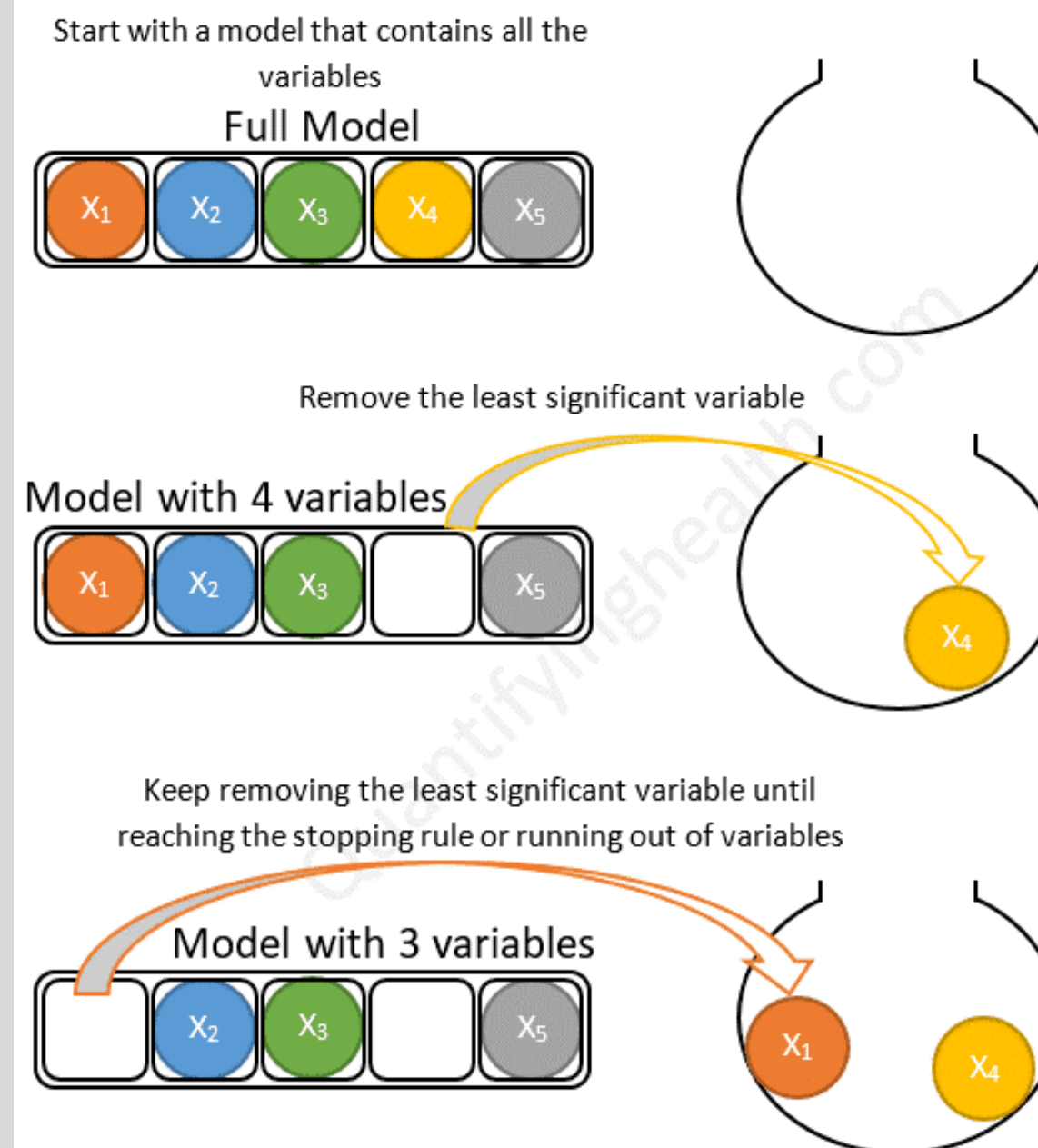
- Stepwise Forward
- Stepwise backward



Forward stepwise selection example with 5 variables:



Backward stepwise selection example with 5 variables:



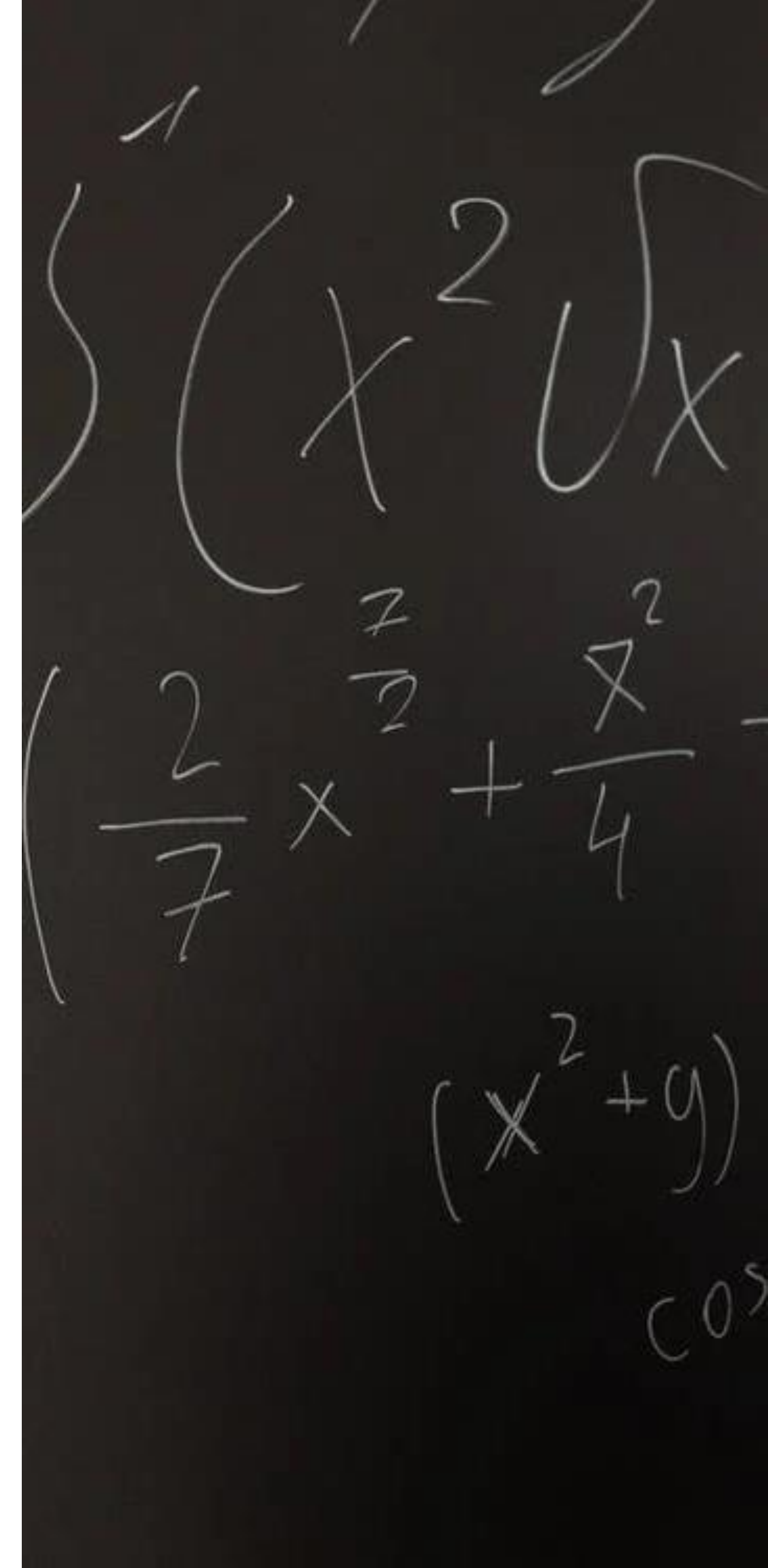
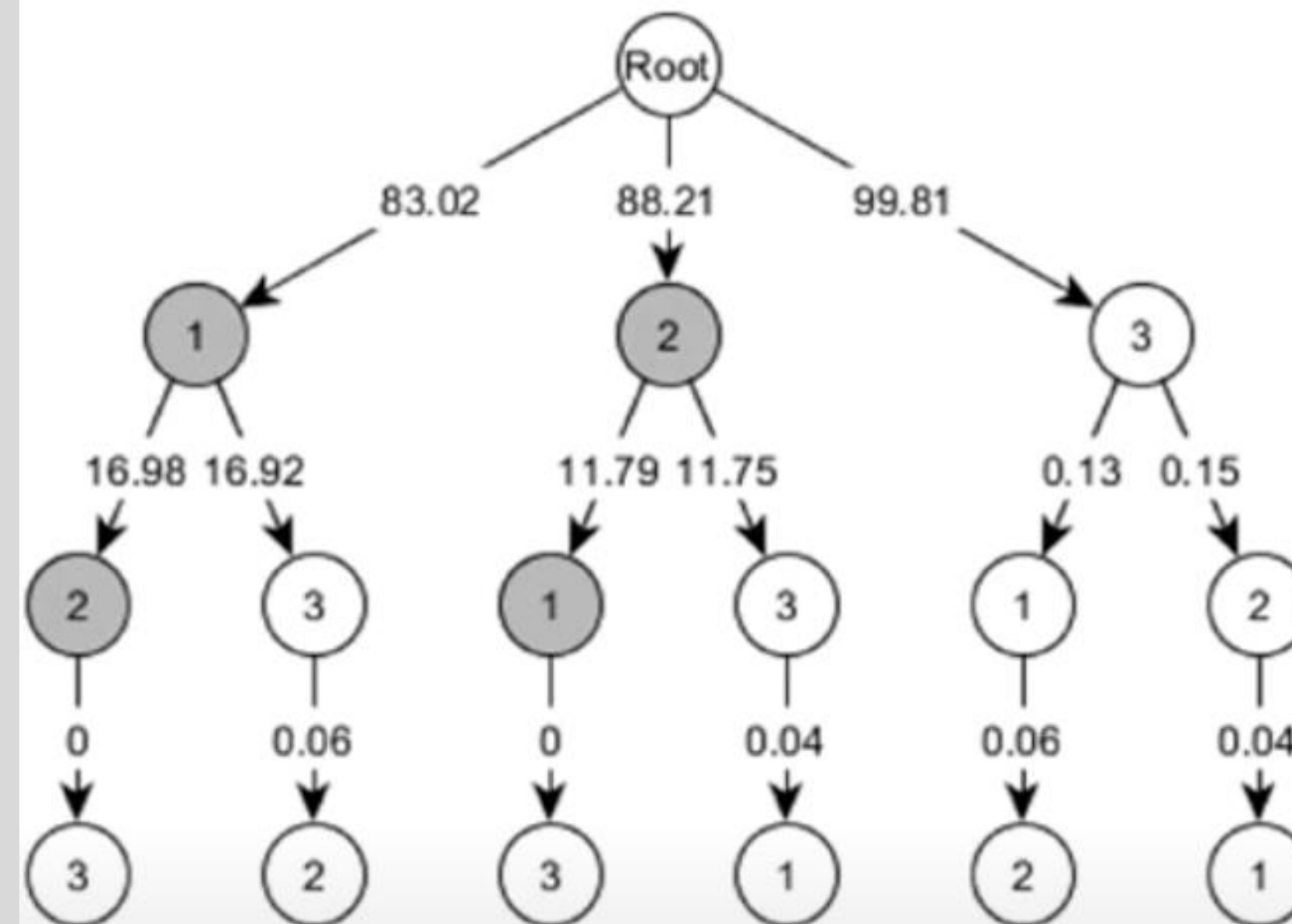




# Desarrollo de conceptos

## Stepwise Forward

- Identificar variable que maximiza la separación
- Formas parejas con las restantes
- Encontrar la pareja con la mayor discriminación
- Con los pares se forman triadas para encontrarla combinación más discriminante

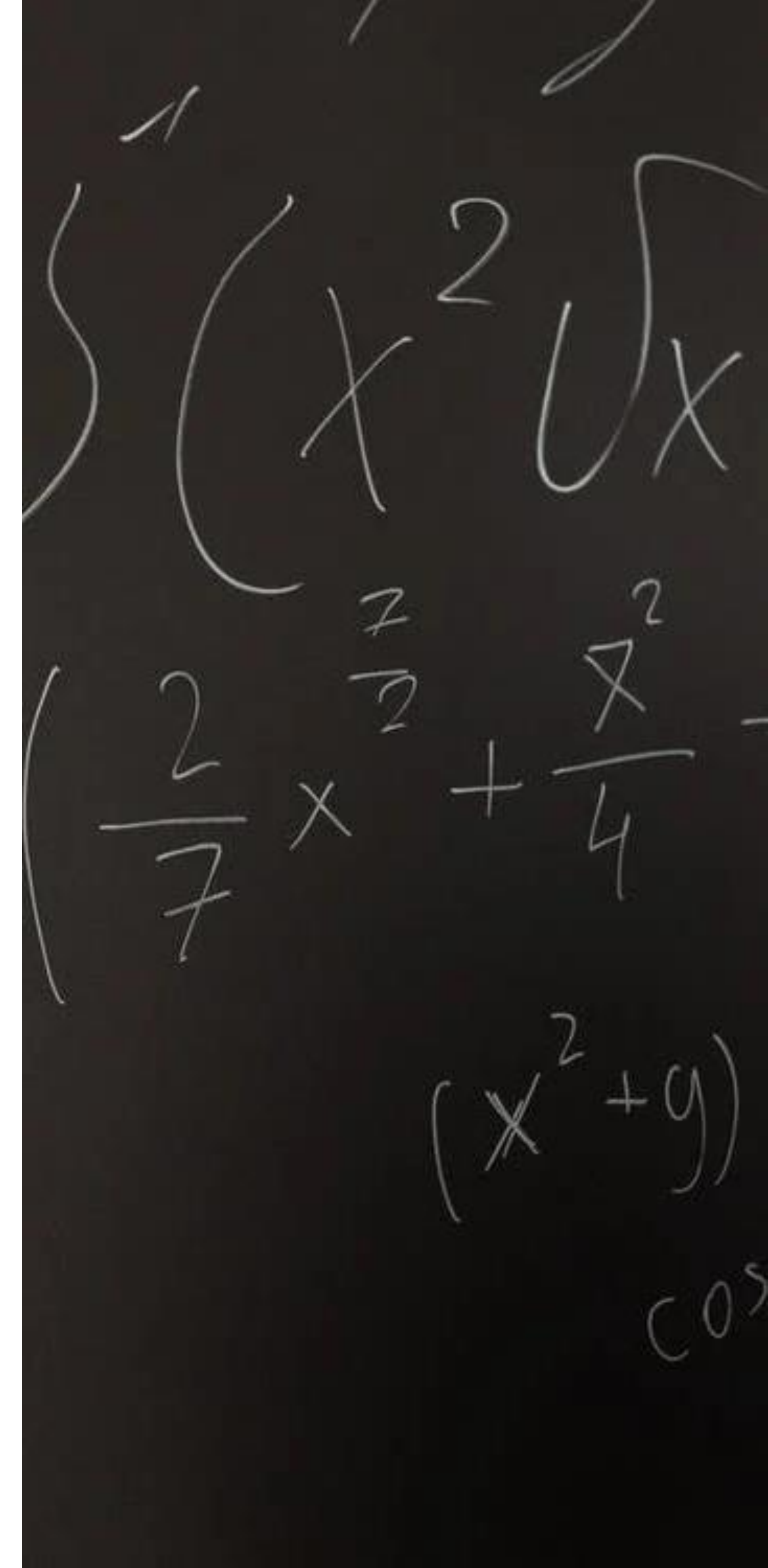




# Desarrollo de conceptos

## Stepwise Backward

- Se consideran todas las variables
- Se van eliminando las variables que no aportan discriminación en los datos

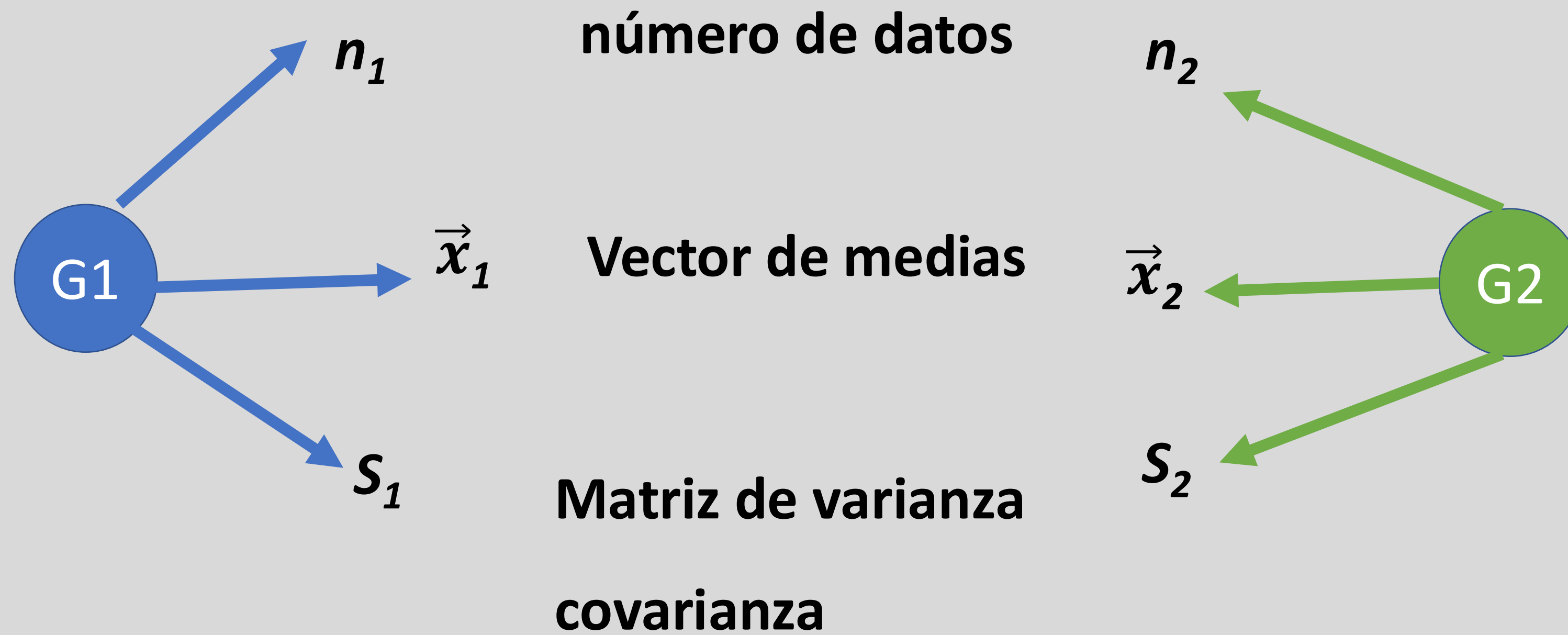






# Desarrollo de conceptos

## LDA para dos clases



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



# Desarrollo de conceptos

## LDA para dos clases

Función lineal sobre las dos variables

$$f(x) = a_1 x_1 + \dots + a_p x_p$$

Que nos ayude a

-Maximizar la variabilidad entre grupos

$$(a' \bar{x}_2 - a' \bar{x}_1)^2$$

-Minimizar la variabilidad dentro del grupo

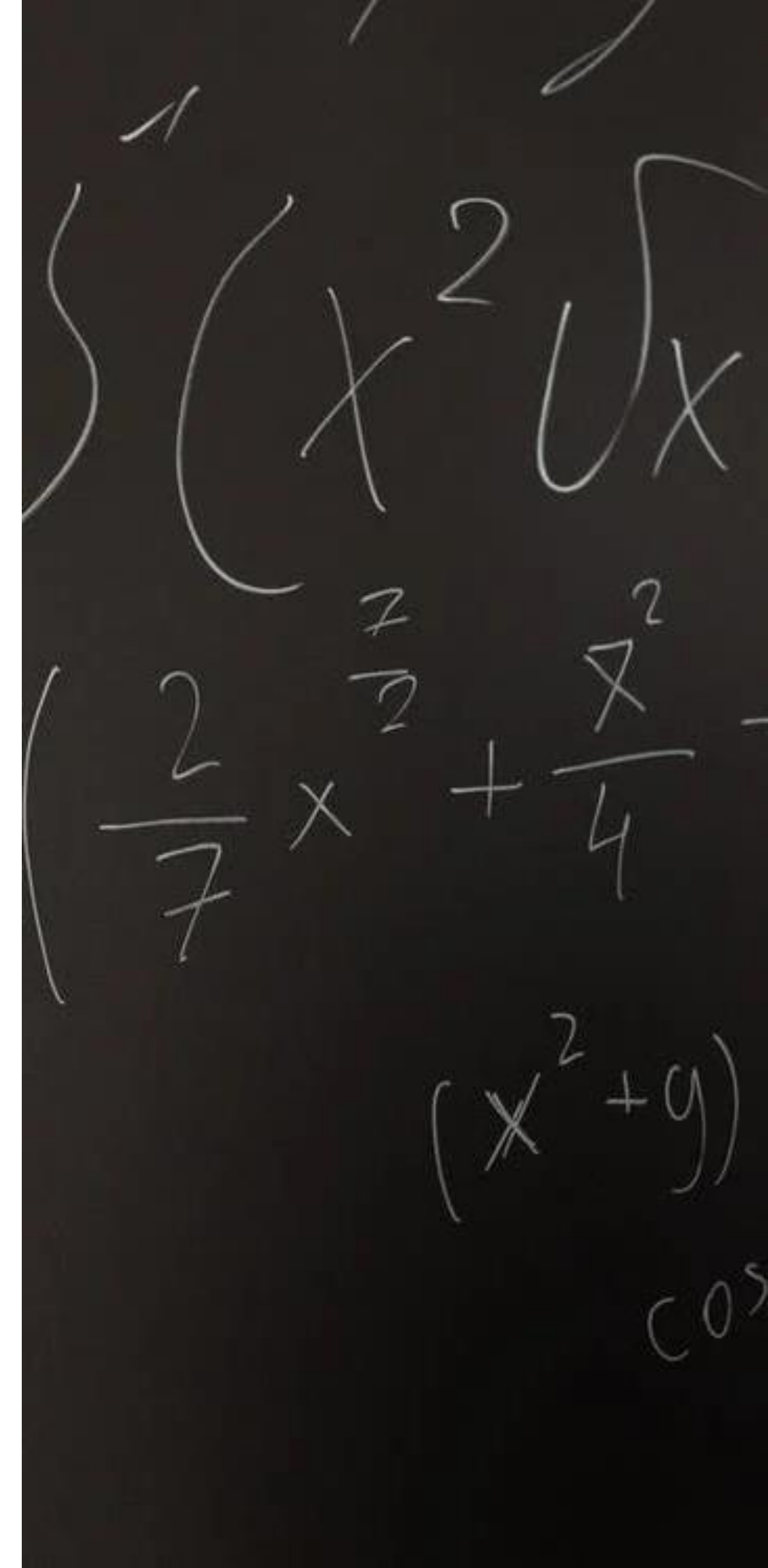
$$a' S_w a$$

Resolver

$$\max \frac{(a' \bar{x}_2 - a' \bar{x}_1)^2}{a' S_w a}$$

Elegir la dirección del vector (eigenvector)

$$\hat{w} = S_w^{-1} (\bar{x}_2 - \bar{x}_1)$$

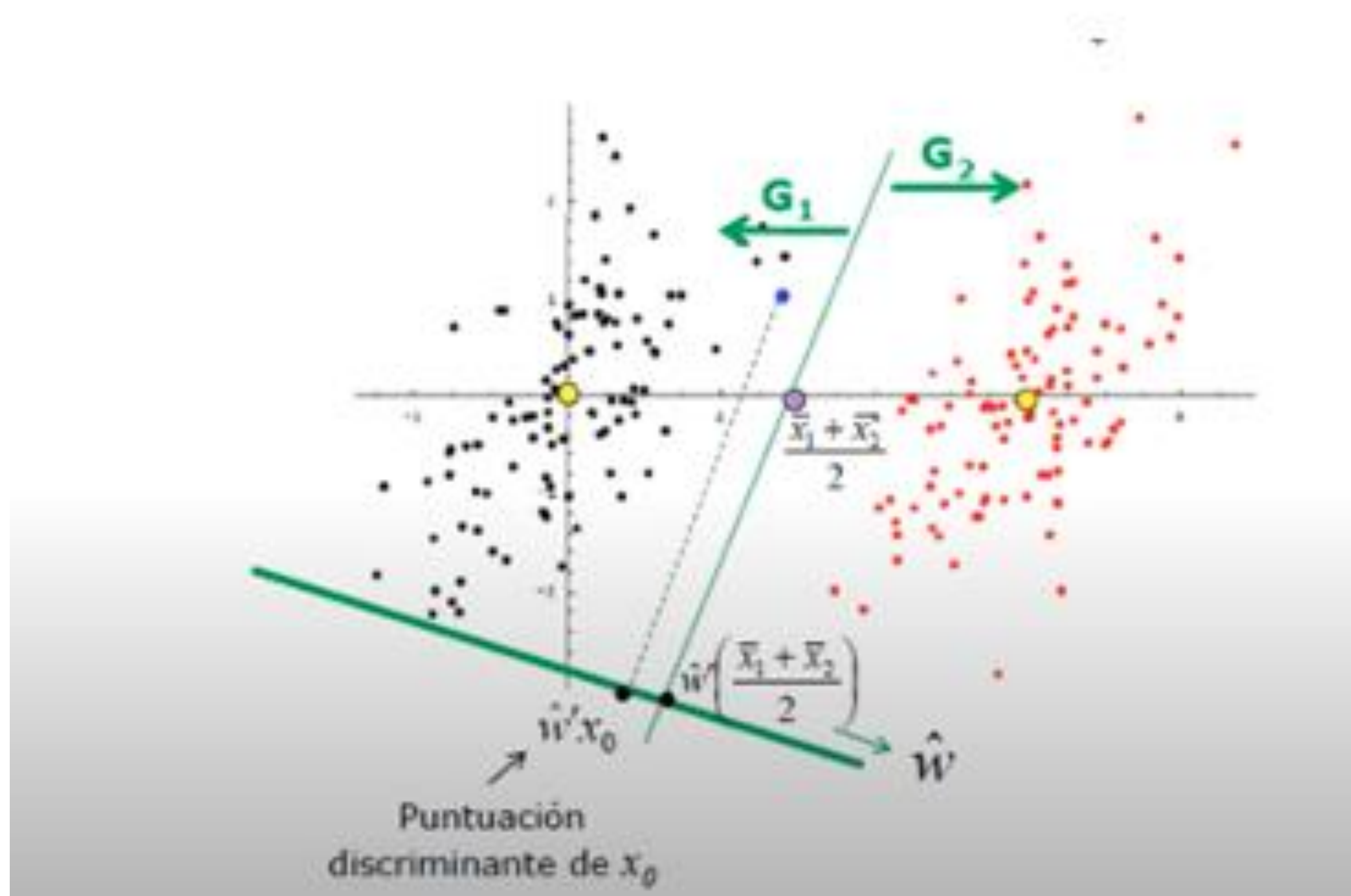






# Desarrollo de conceptos

## Puntuación discriminante



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



# Desarrollo de conceptos

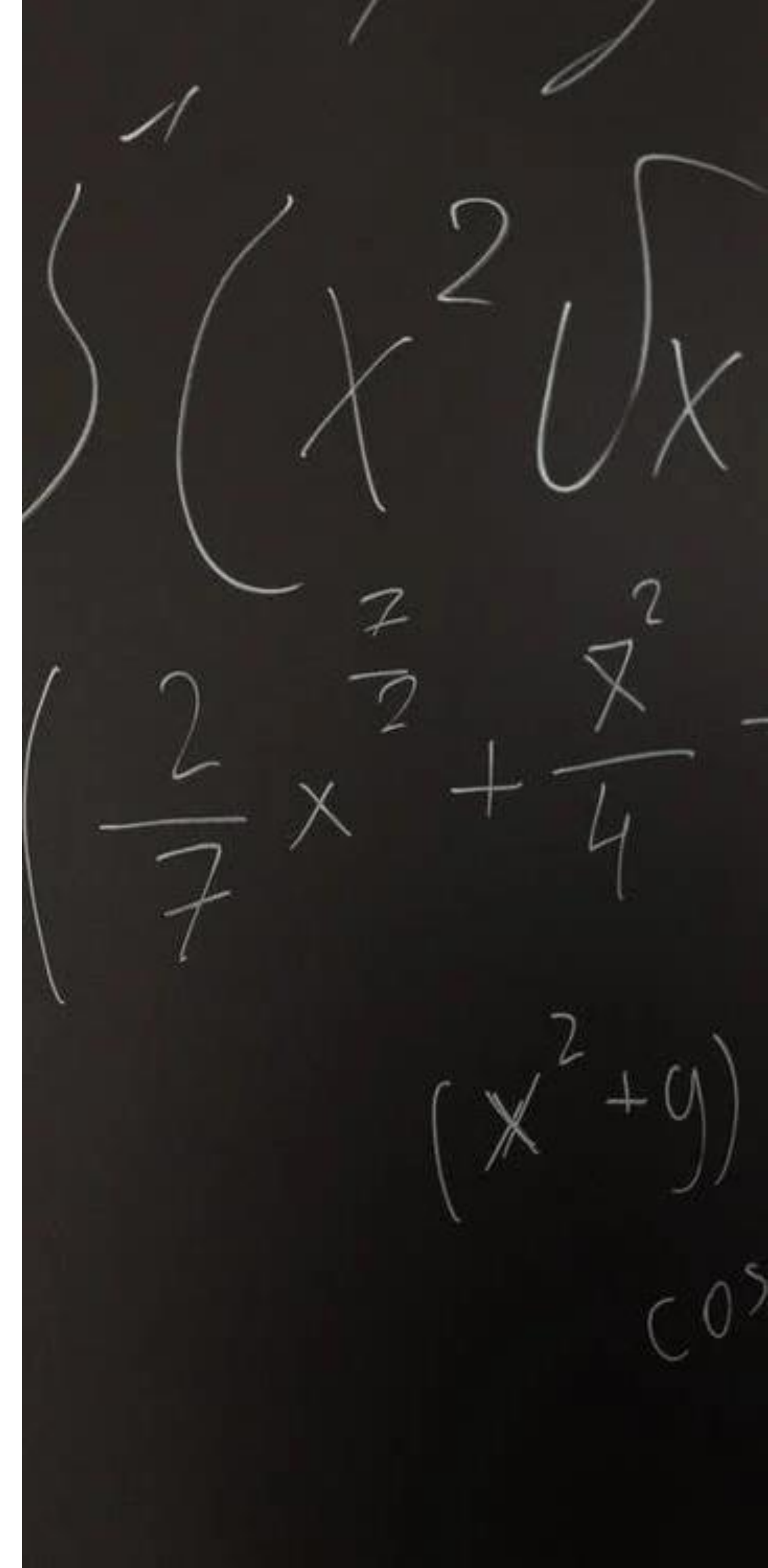
## Modelo

El análisis discriminante lineal (LDA) es un método supervisado para la reducción de la dimensionalidad en problemas de clasificación.

Dadas muestras de dos clases **C1** y **C2**, queremos encontrar la dirección, definida por un vector **w**, tal que cuando los datos se proyecten en **w**, los ejemplos de las dos clases estén tan separados como sea posible.

$$z = \mathbf{w}^T \mathbf{x}$$

es la proyección de **x** sobre **w** y por lo tanto es una reducción de dimensionalidad de **d** a **1**.







# Desarrollo de conceptos

## Modelo

La dispersión de muestras de G1 y G2 después de la proyección son:

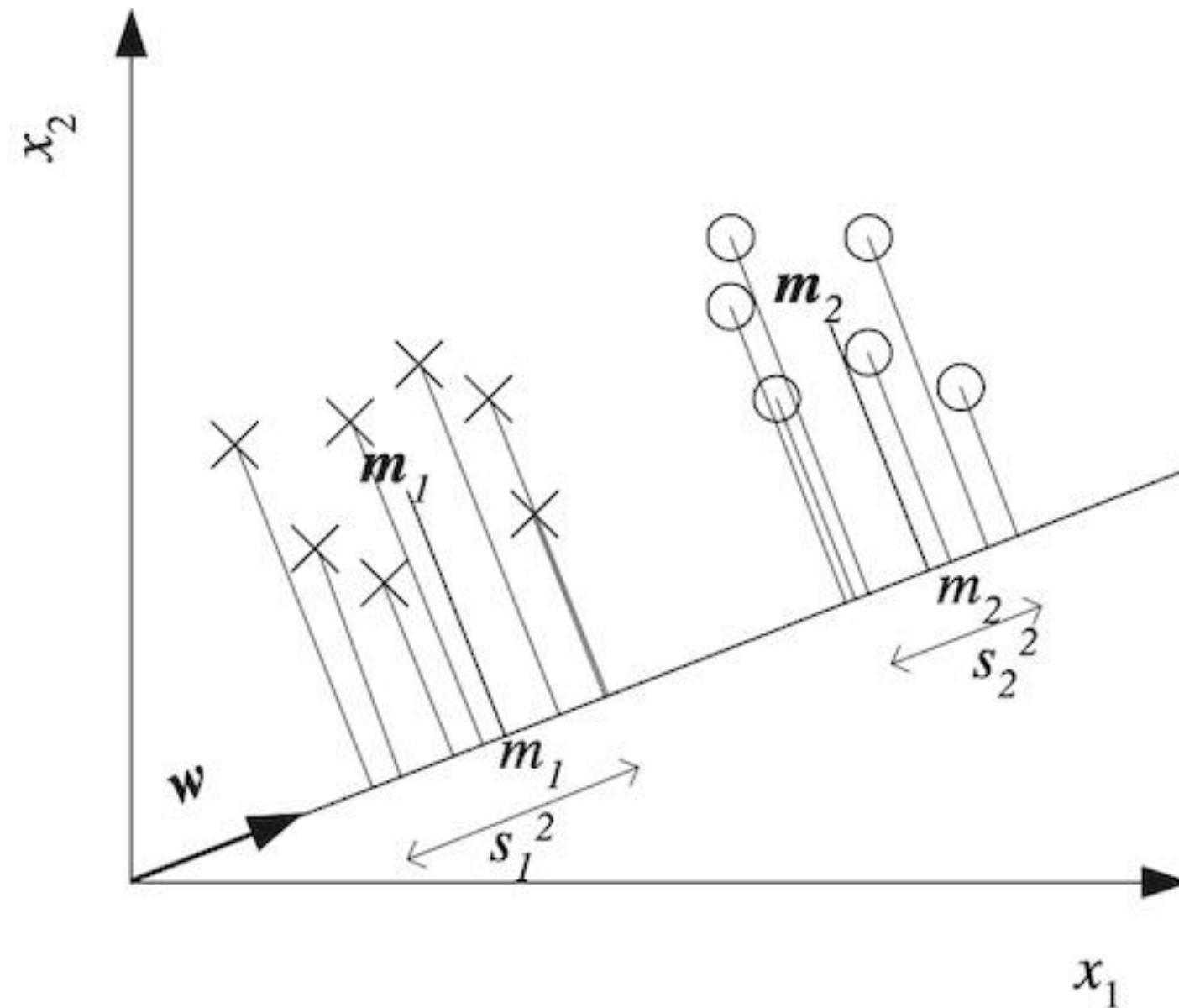


Figure 6.10 Two-dimensional, two-class data projected on  $w$ .

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



# Desarrollo de conceptos

## Casos

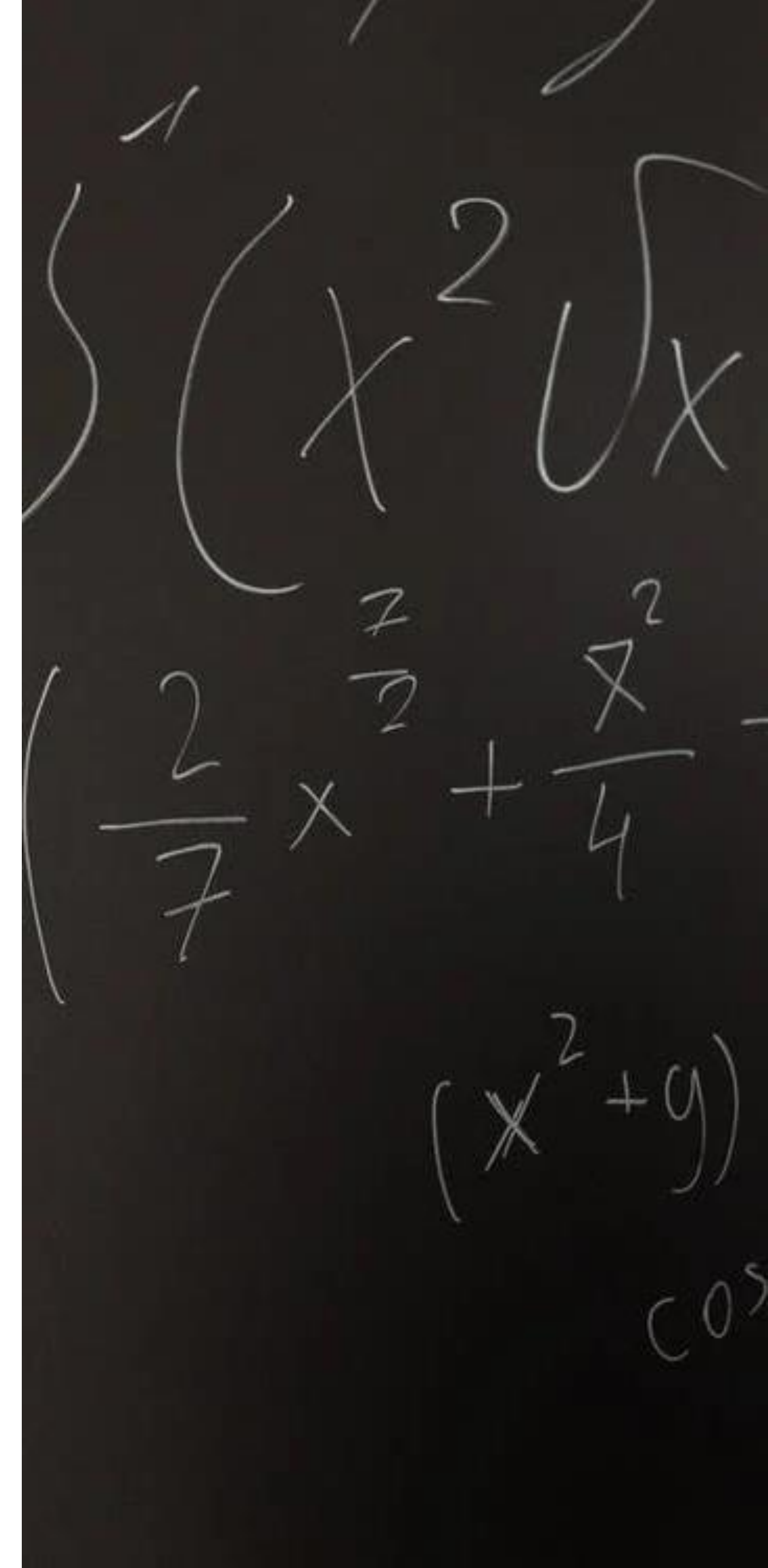
Concesión de créditos

Clasificación de patrones

Reconocer voces y sonidos

Clasificar billetes

Reconocer caracteres escritos







# Solución

- Dentro de Colab
- <https://colab.research.google.com>

```
3 require File.expand_path("../..", __FILE__)
4 # Prevent database truncation if the environment is production
5 abort("The Rails environment is running in production")
6 require 'spec_helper'
7 require 'rspec/rails'
8
9 require 'capybara/rspec'
10 require 'capybara/rails'
11
12 Capybara.javascript_driver = :webkit
13 Category.delete_all; Category.create(:name => "Category")
14 Shoulda::Matchers.configure do |config|
15   config.integrate do |integrate|
16     with.test_framework :rspec
17     with.library :rails
18   end
19 end
20
21 # Add additional requires below this line
22
23 # Requires supporting ruby files with spec/
24 # spec/support/ and its subdirectories
25 # run as spec files by default. The
26 # in _spec.rb will both be required
27 # run twice. It is recommended to
28 # end with _spec.rb. You can use
29 # option on the command line to
30
31 No results found for 'mongoid'
```



## 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>
- Alicia Sanchís Arellano, José A. Gil & Antonio Heras Martínez (2003) The Discriminant Analysis in the prediction of failure for non-life insurance companies, Spanish Journal of Finance and Accounting / Revista Española de Financiación y Contabilidad, 32:116, 183-233, DOI: 10.1080/02102412.2003.10779487
- <https://fei.edu.br/~cet/isas99.pdf>
- <https://quantifyinghealth.com/stepwise-selection/>