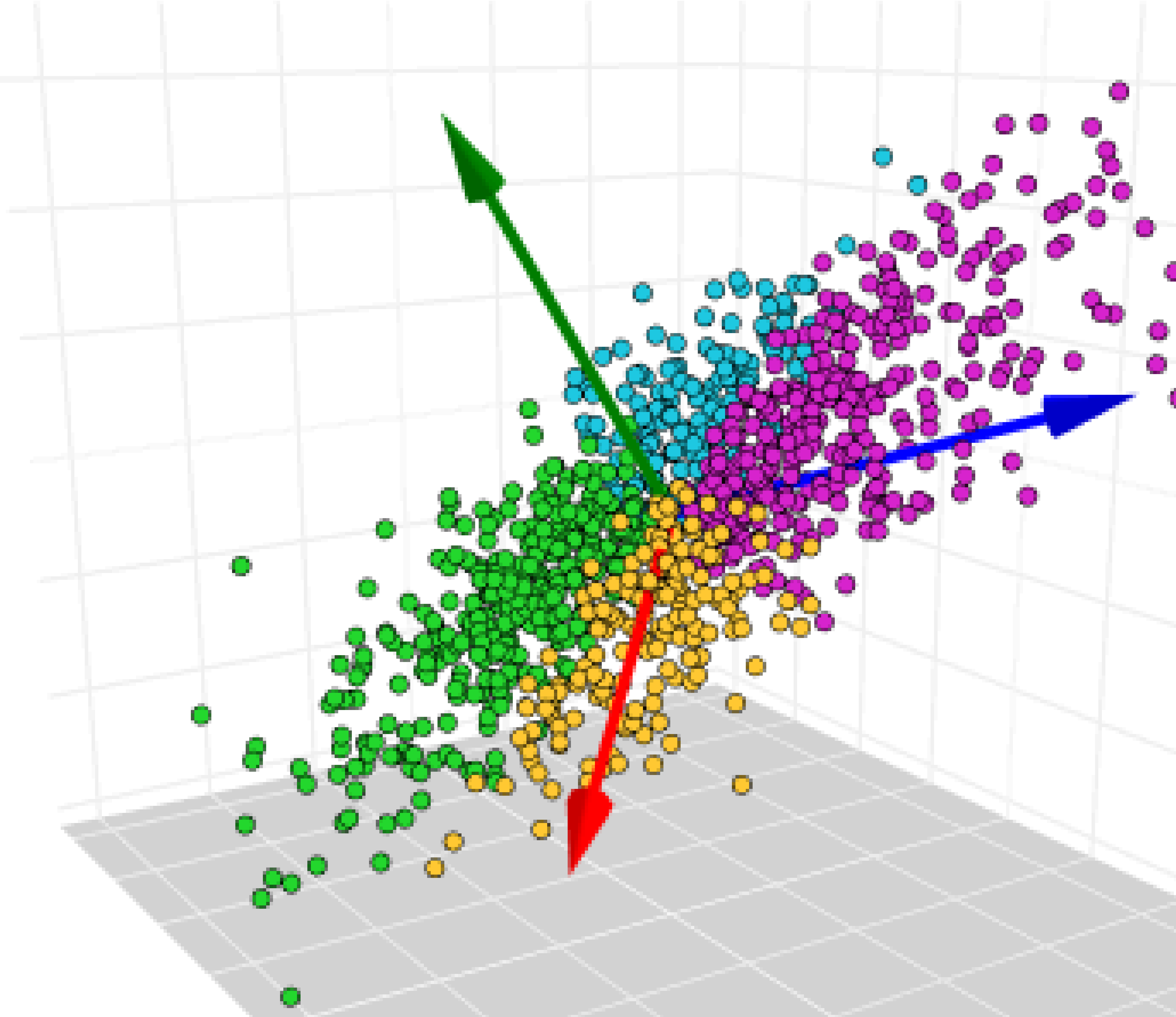
PCA

Principal Component Analysis



Presentado por
Jerónimo Hoyos Botero

¿Qué es el PCA?

El análisis de componentes principales (PCA) es una herramienta con la cuál se realiza una **Reducción de dimensionalidad** sobre unos datos para revelar las estructuras a veces ocultas y simplificadas que a menudo subyacen a él.

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & x_{15} \\ x_{21} & x_{22} & x_{23} & x_{24} & x_{25} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & x_{n4} & x_{n5} \end{bmatrix} \xrightarrow{\text{PCA}} \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \\ \vdots & \vdots \\ z_{n1} & z_{n2} \end{bmatrix}$$

Algebra lineal

Una de las bondades del PCA es el uso de **algebra lineal básica.**

Valores y Vectores Propios

$$Ax = \lambda x$$

Matrices Ortogonales

$$A^{-1} = A^T$$

Matrices Simetricas

$$A = A^T$$

Diagonalización Ortogonal

$$A = EDE^T$$

Otras herramientas matemáticas

Promedio

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Varianza

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Covarianza

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

El lagrangiano

$$max\ R(x,y)$$

$$B(x,y) = b$$

$$\nabla R = \lambda \nabla B$$

$$\mathcal{L}(x,y,\lambda) = R(x,y) - \lambda(B(x,y) - b)$$

Motivación

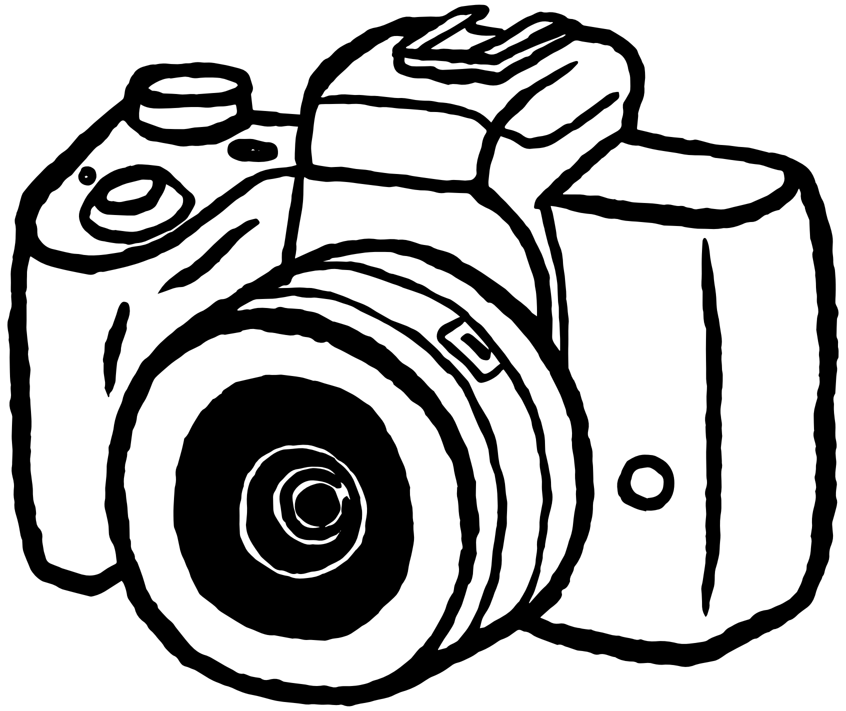
Grabar recuerdos

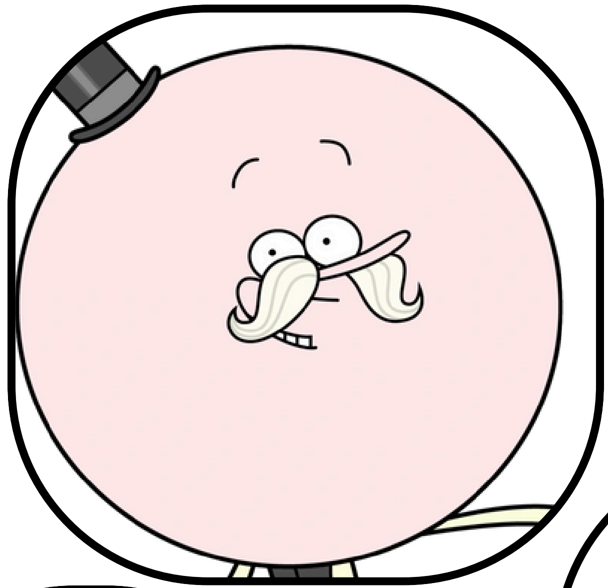
Motivación

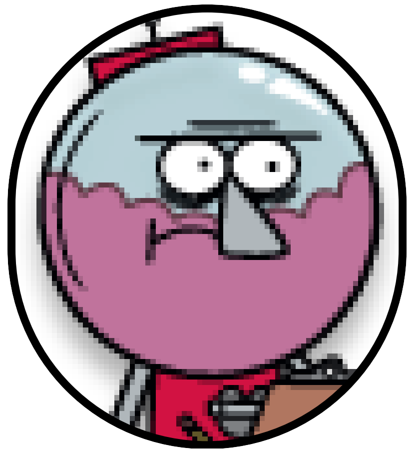
Grabar recuerdos

Motivación

Tomar una foto grupal

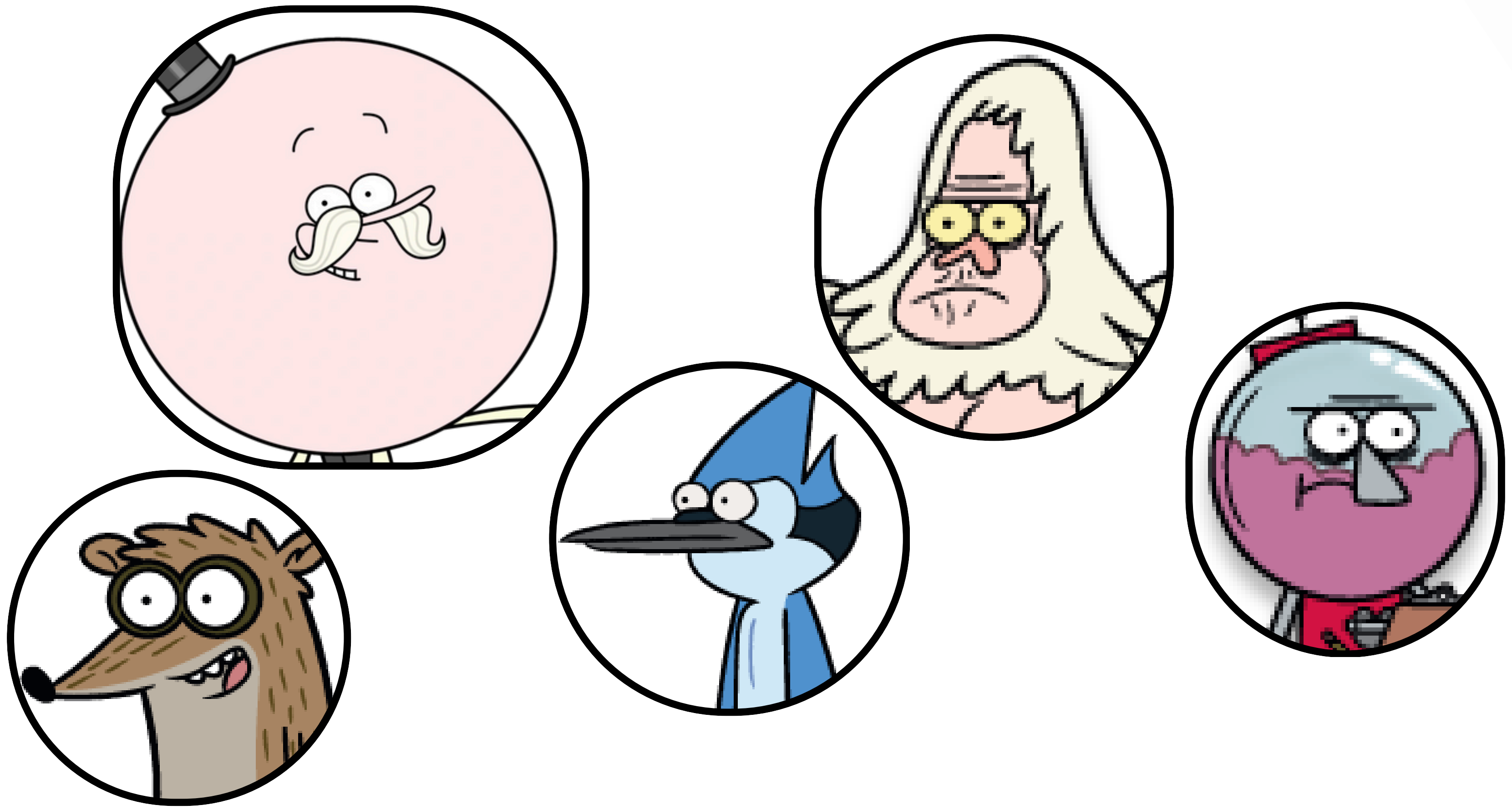






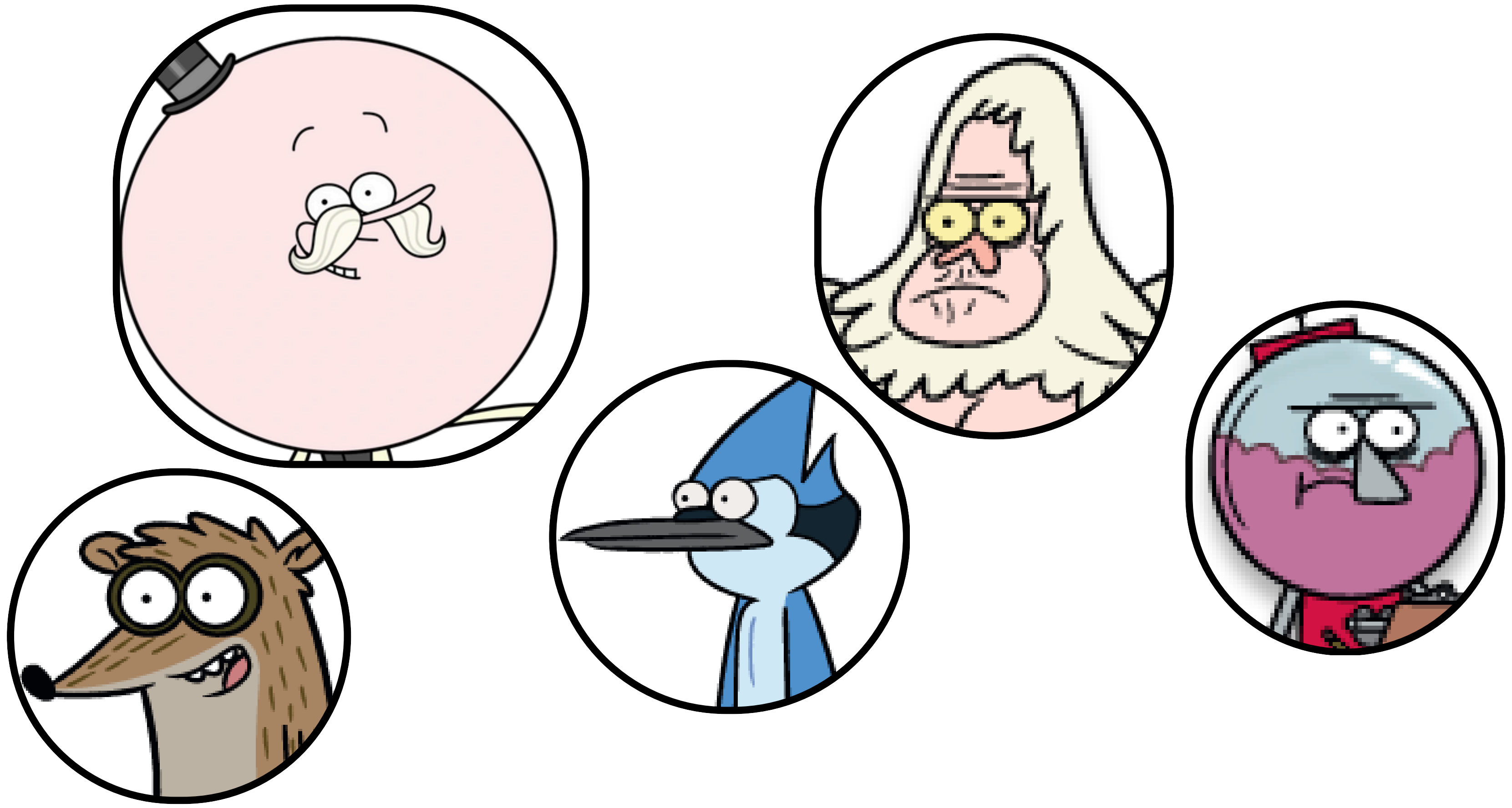
Motivación

Elegir mejor ángulo



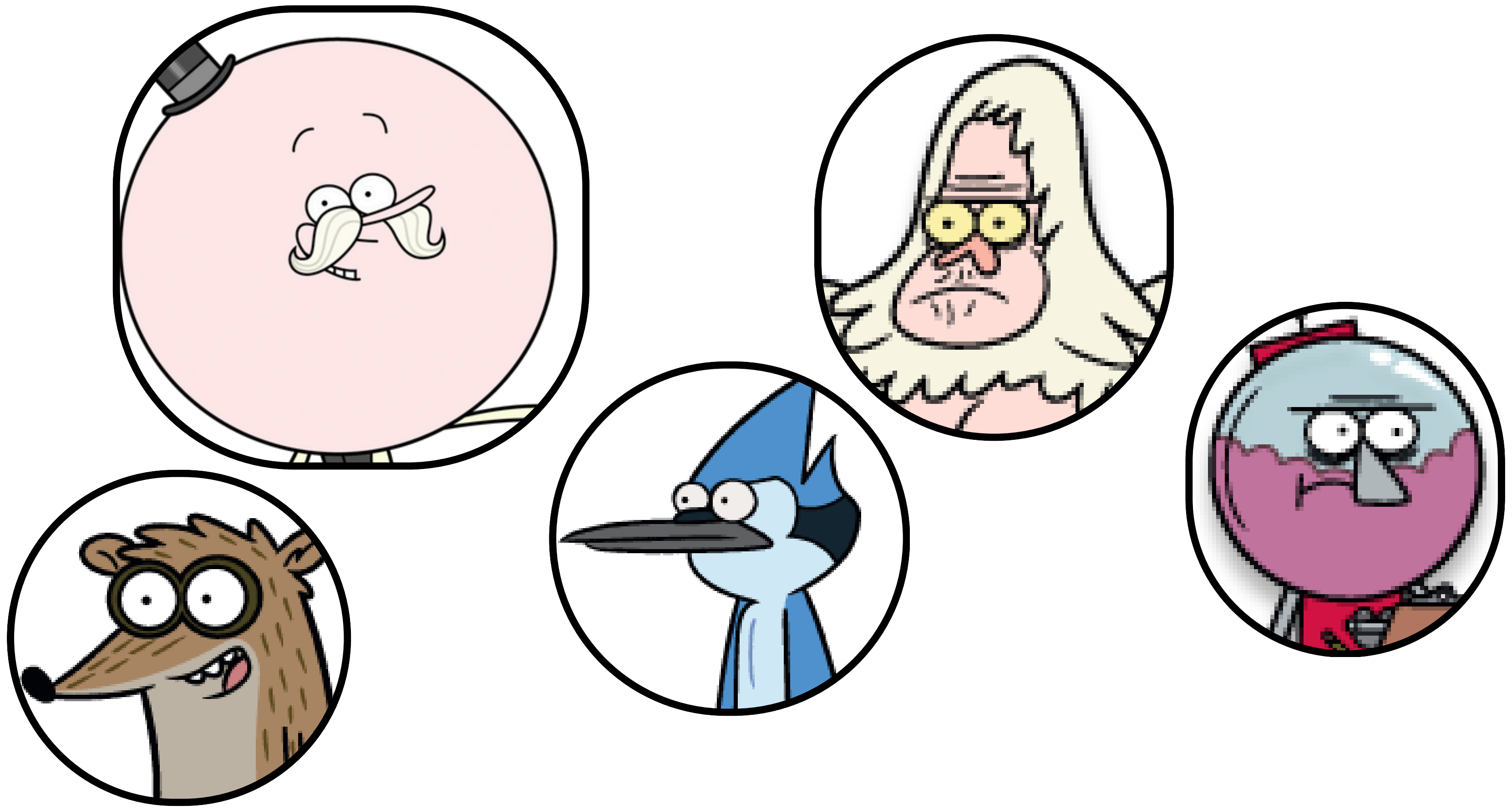
Motivación

Elegir mejor ángulo



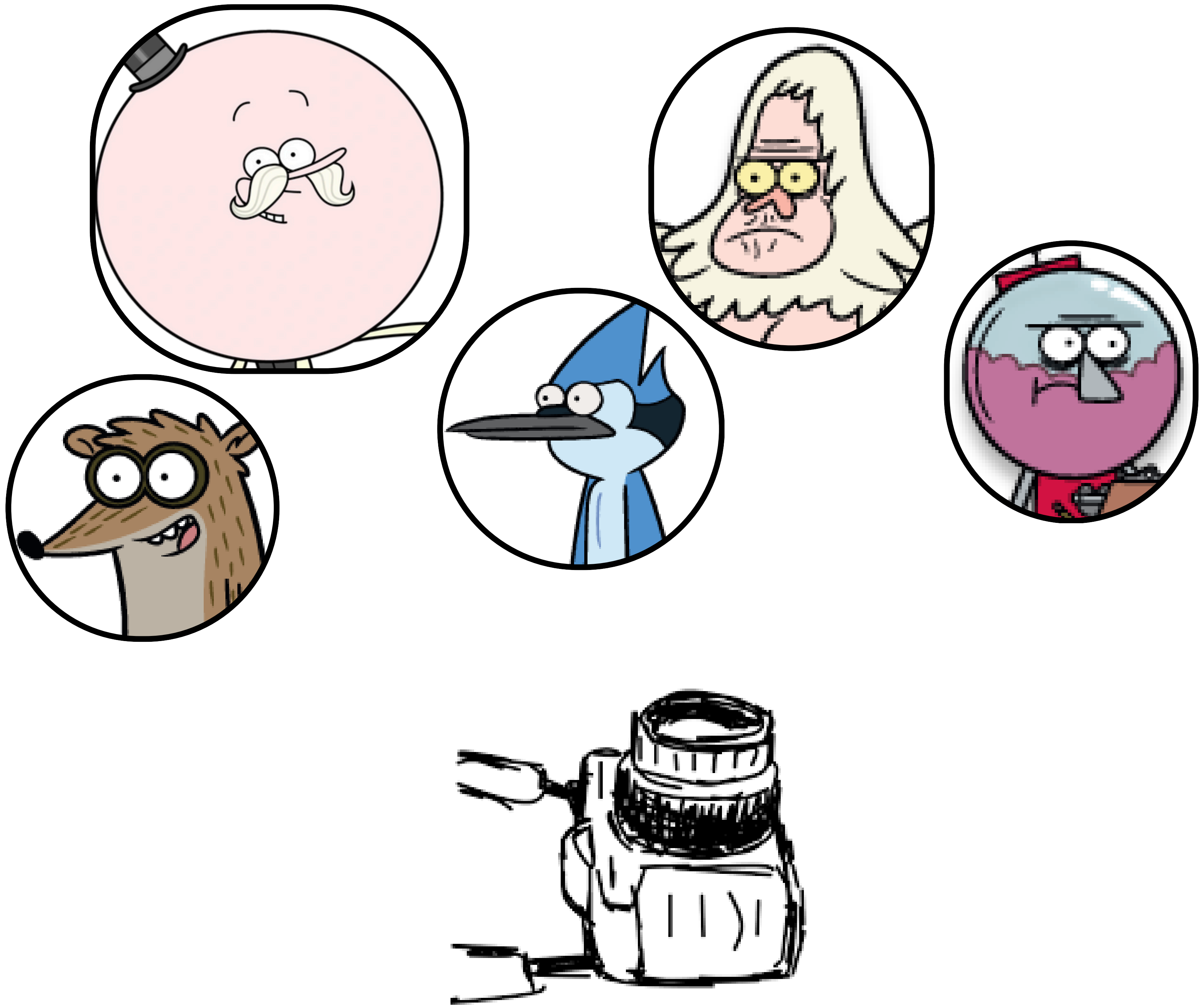
Motivación

Elegir mejor ángulo



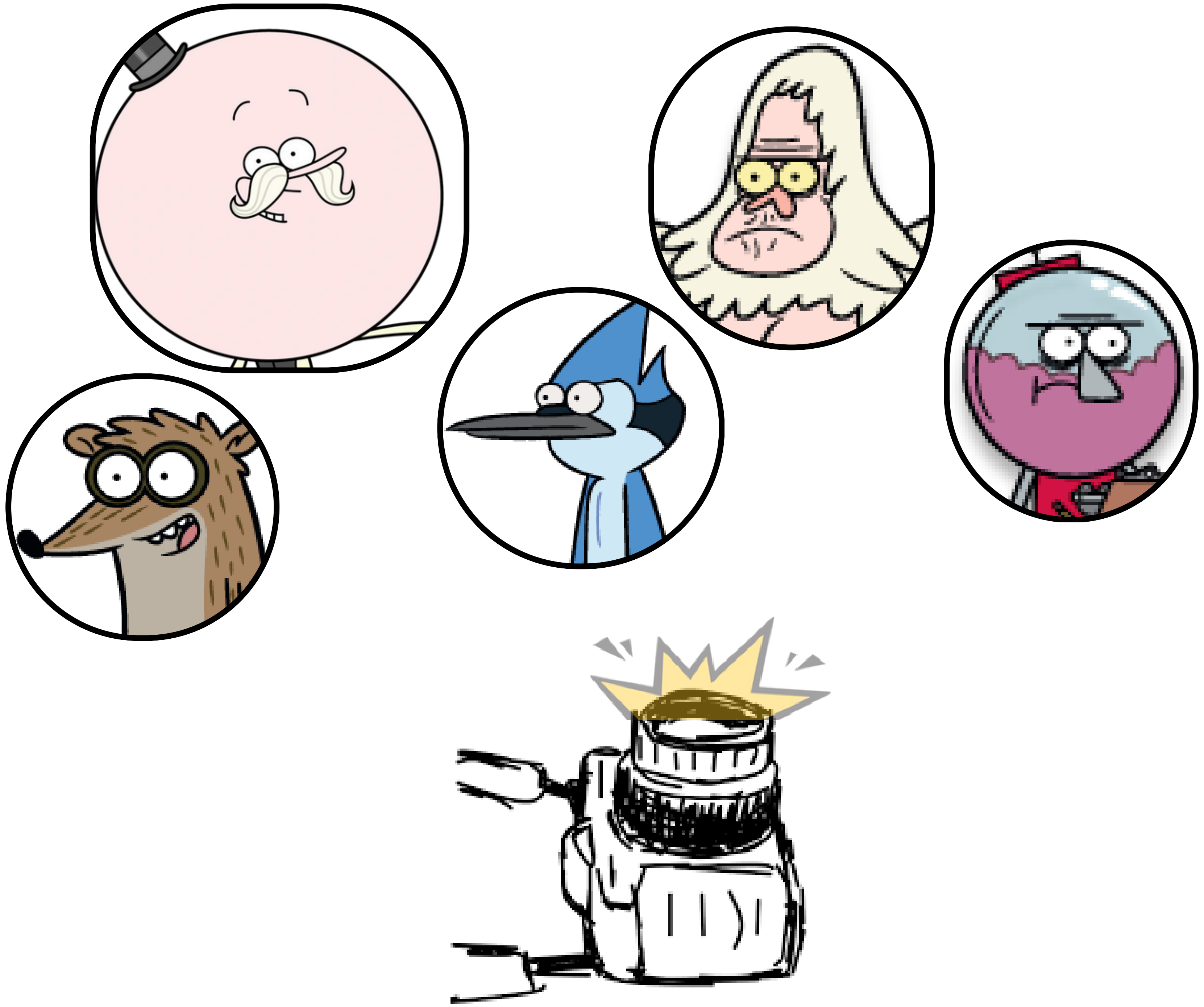
Motivación

Hasta encontrar un óptimo



Motivación

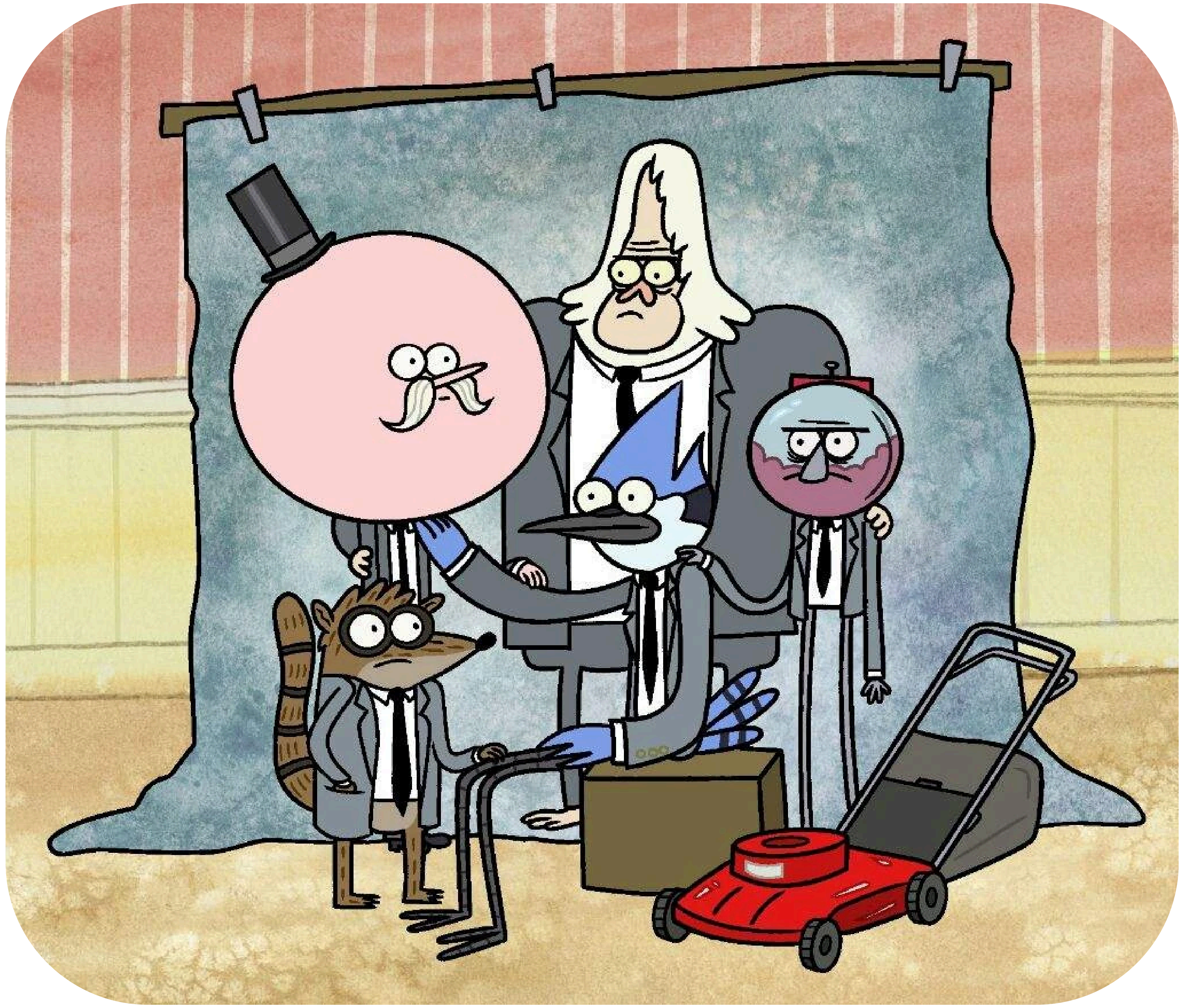
Se obtiene la componente principal



Motivación

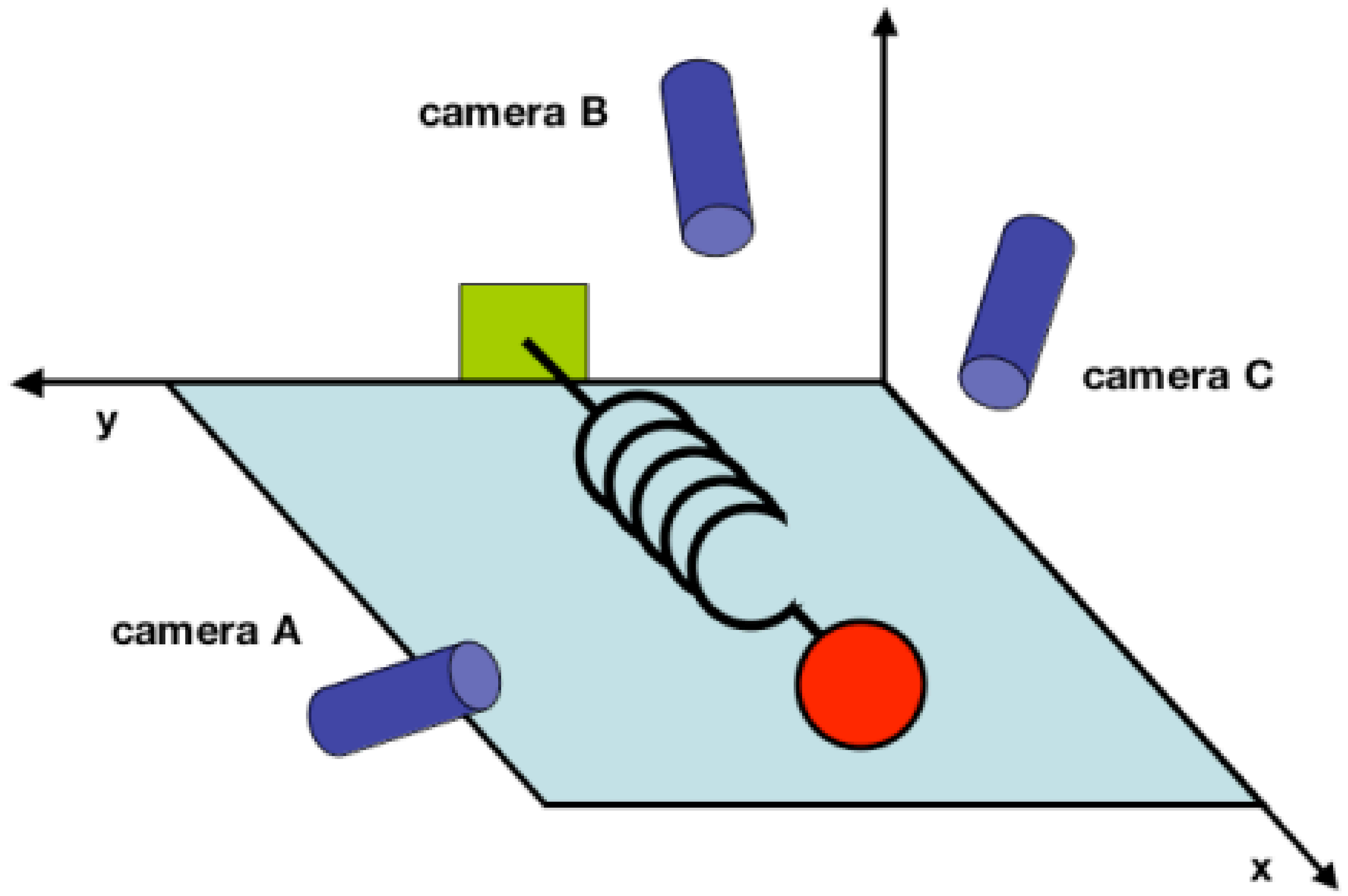
Bajando la dimensionalidad

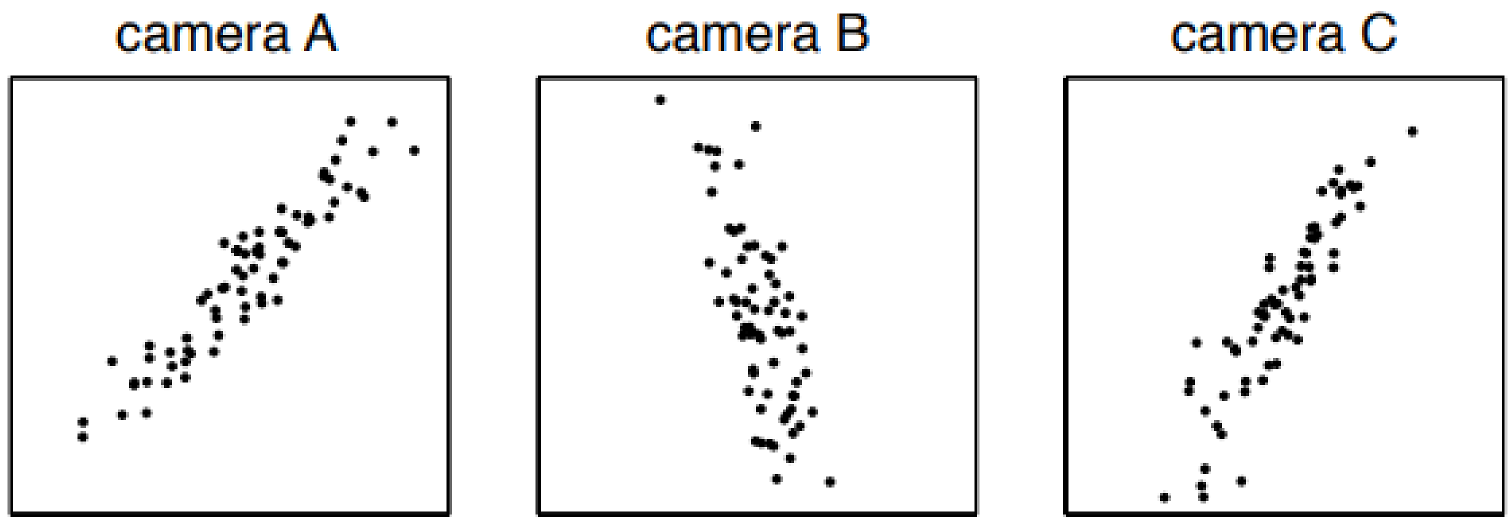
3D → 2D



Motivación

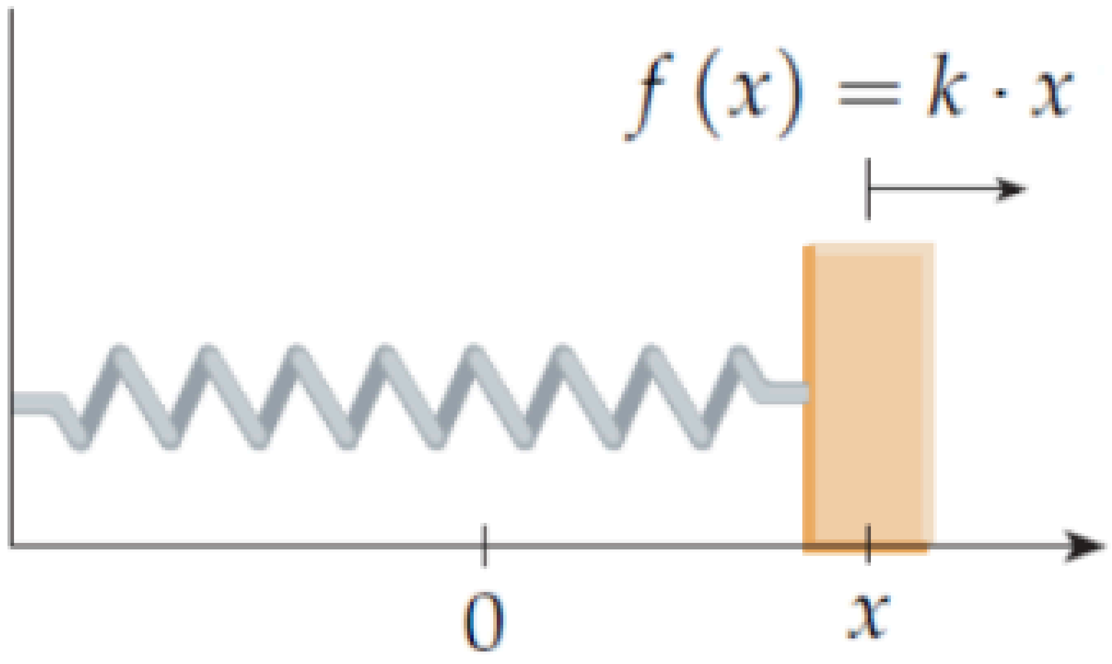
Experimento de la dinamica del resorte



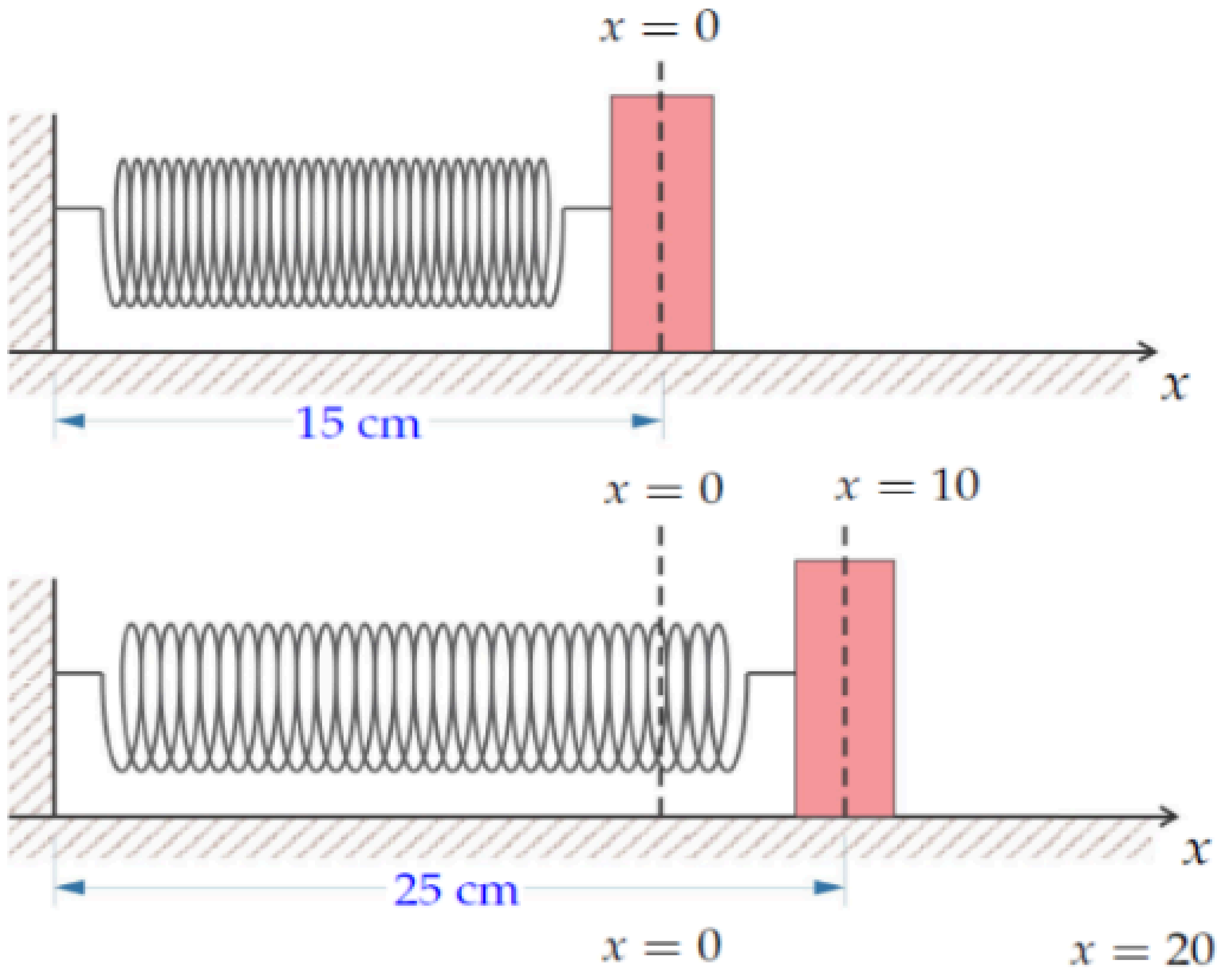


Motivación

Determinamos la mejor posición de la cámara (es decir, la base óptima) y proyectamos los datos sobre ella.



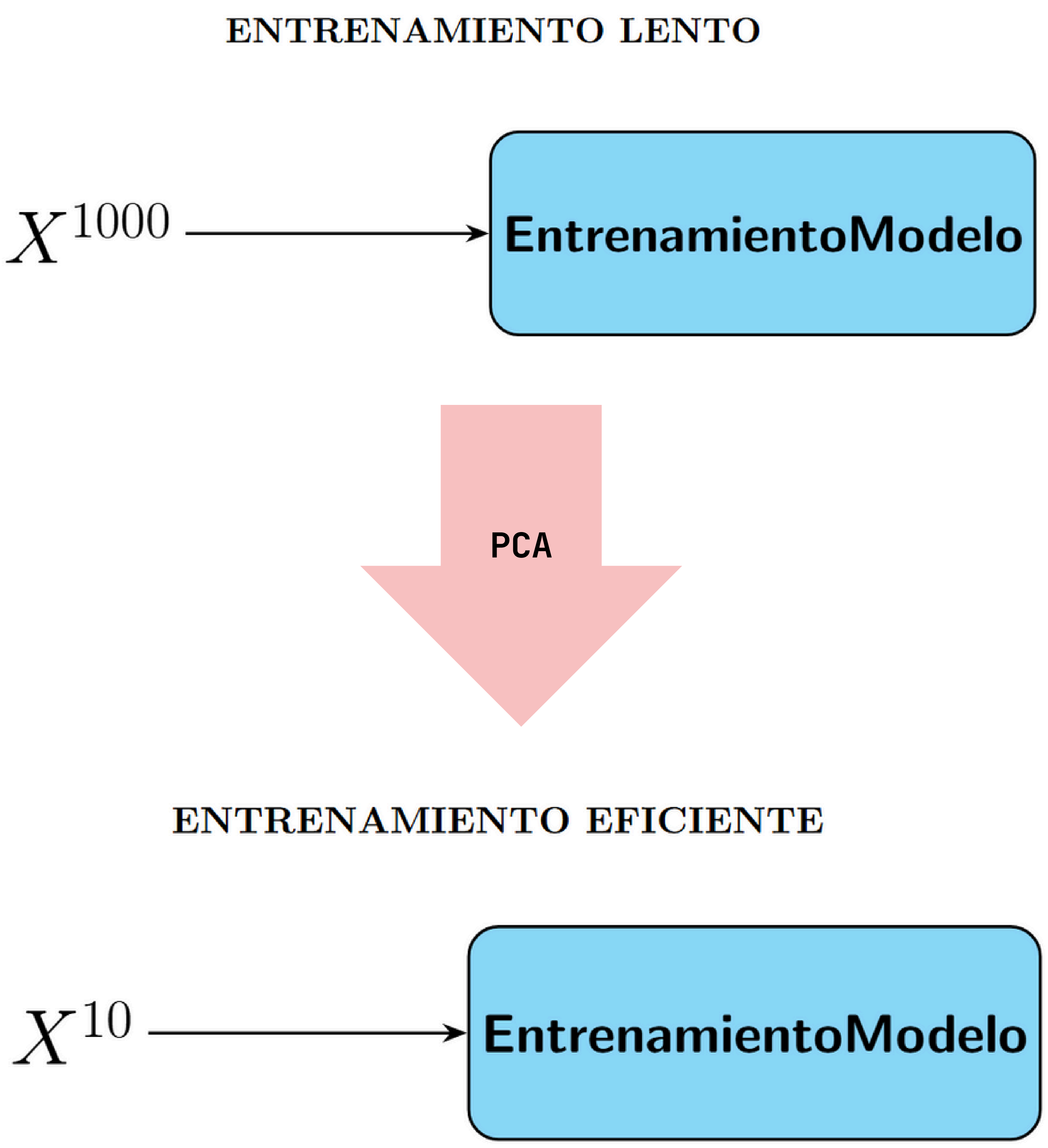
Ley de Hooke



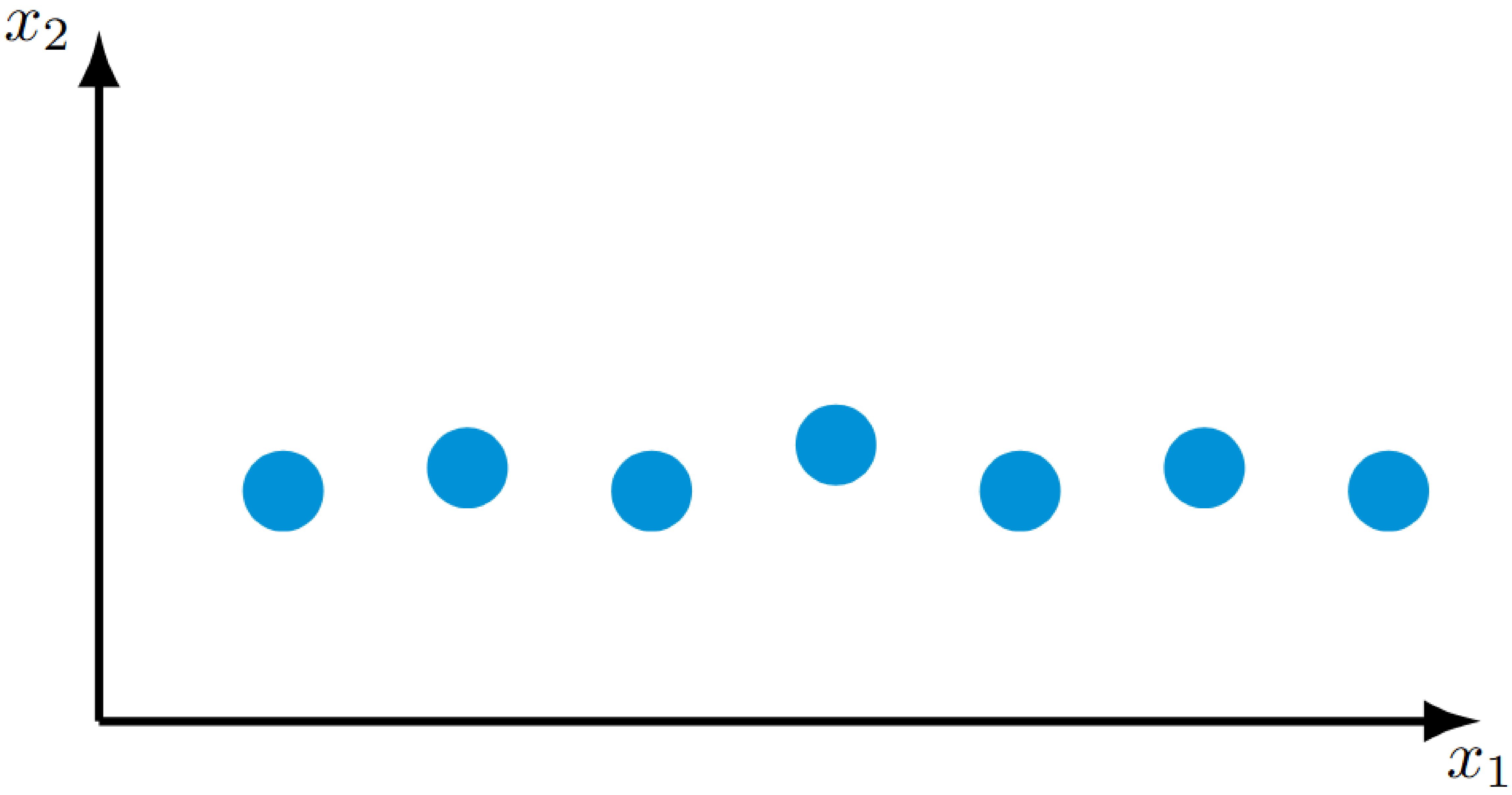
Motivación

Si la aplicación del PCA se realiza de manera óptima sobre un conjunto de datos, puede generar una mejora significativa en el rendimiento del modelo de aprendizaje automático.

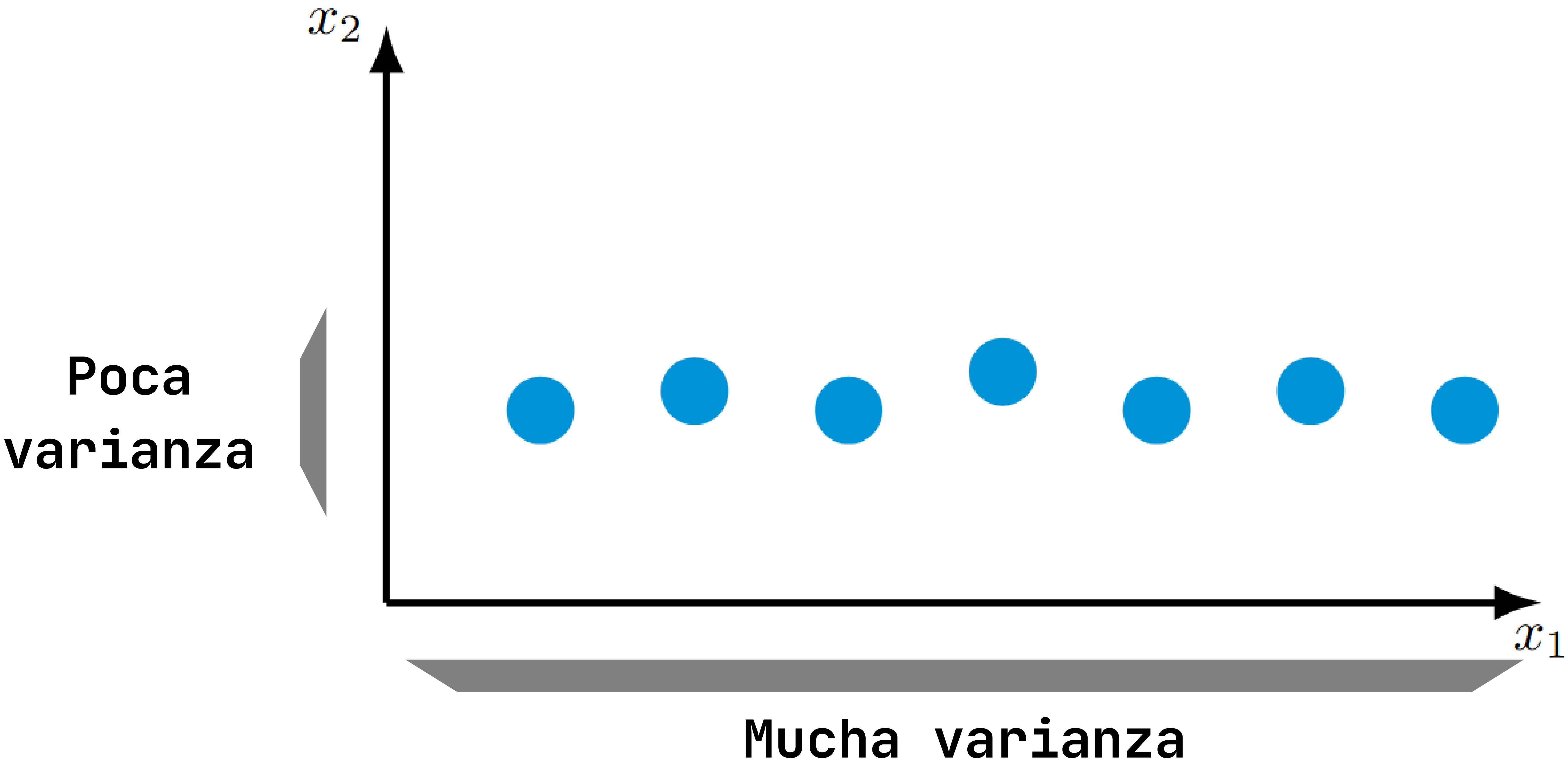
98% de la información --->



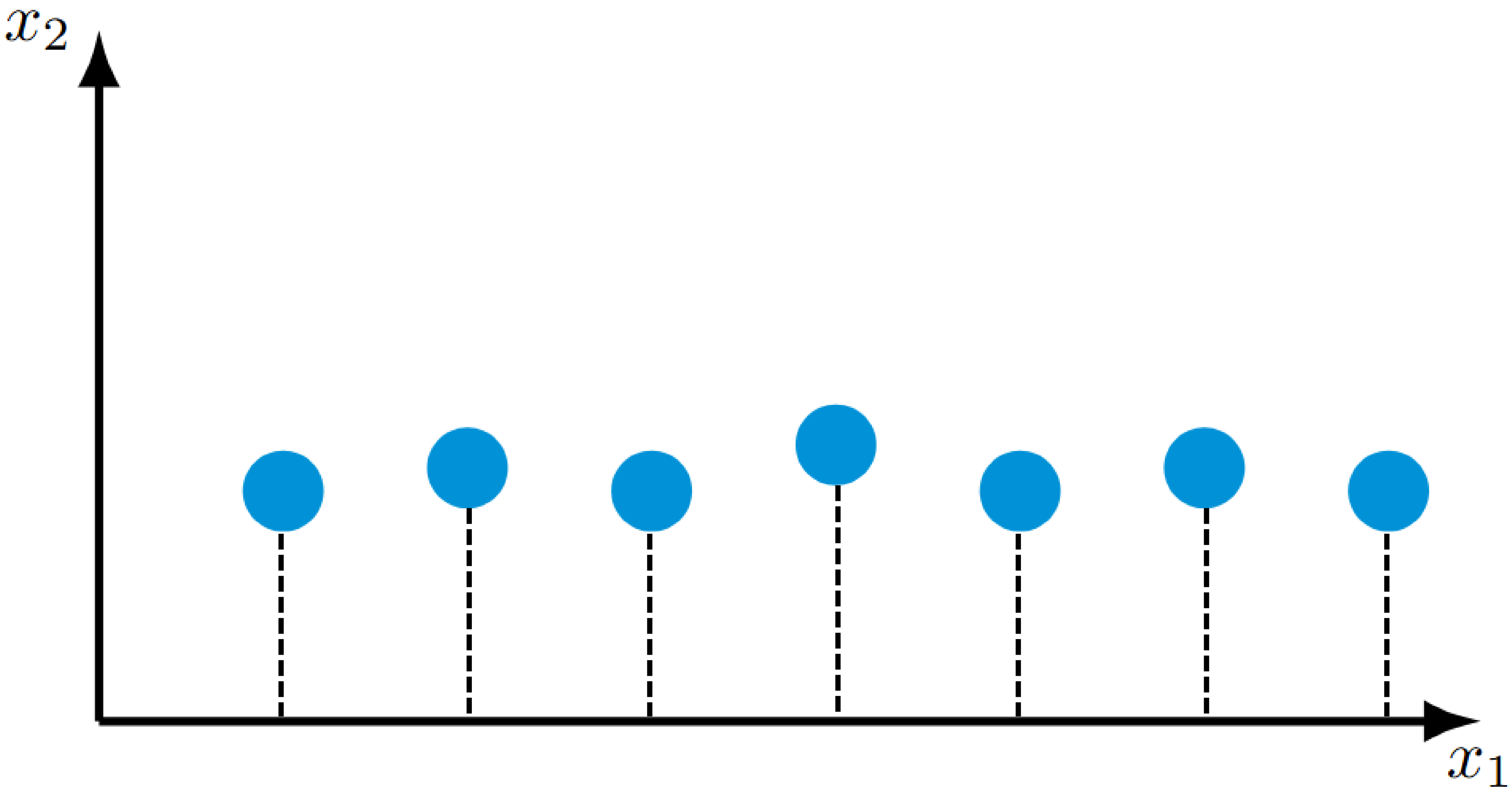
Obtener la mejor foto



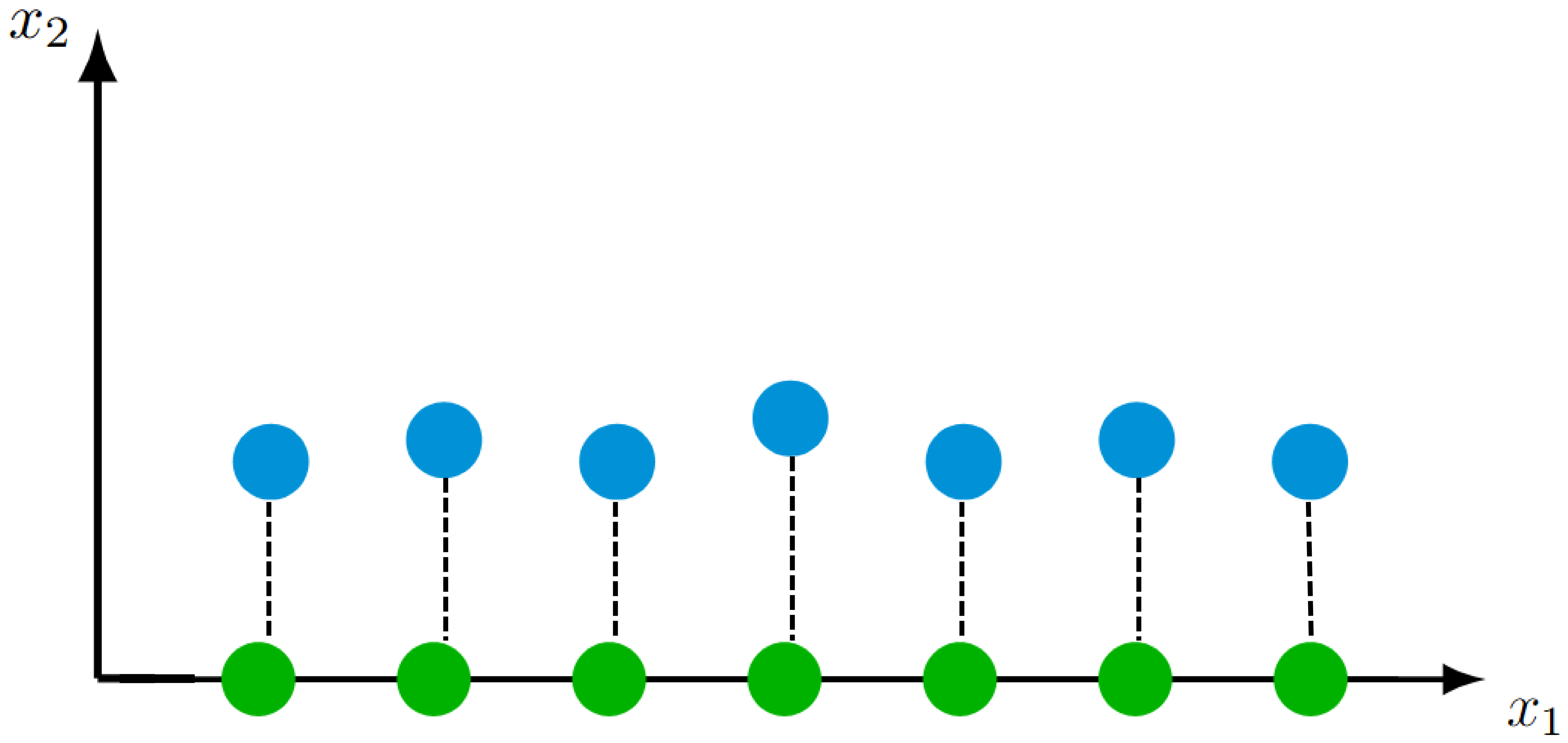
Obtener la mejor foto



Obtener la mejor foto

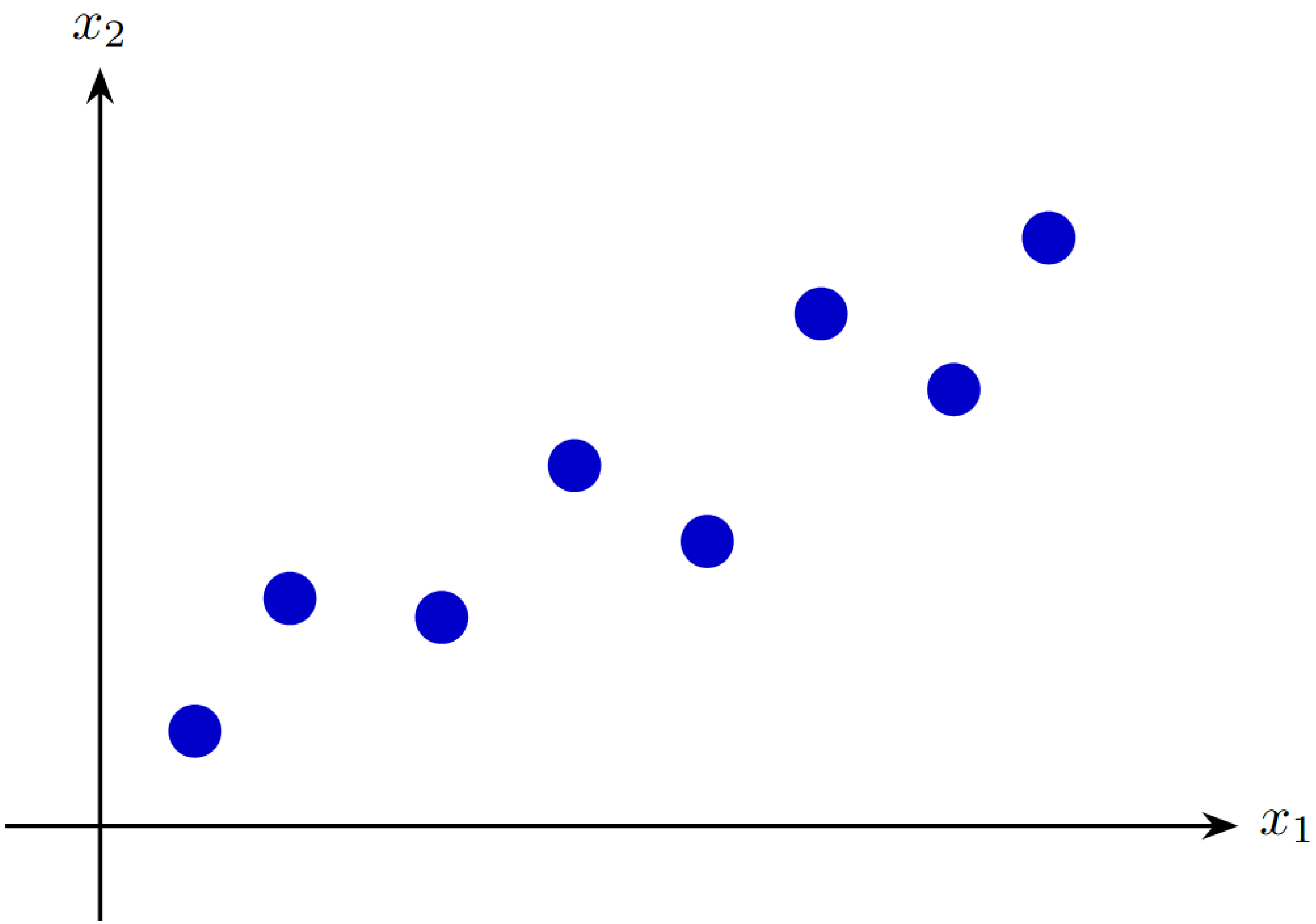


Obtener la mejor foto



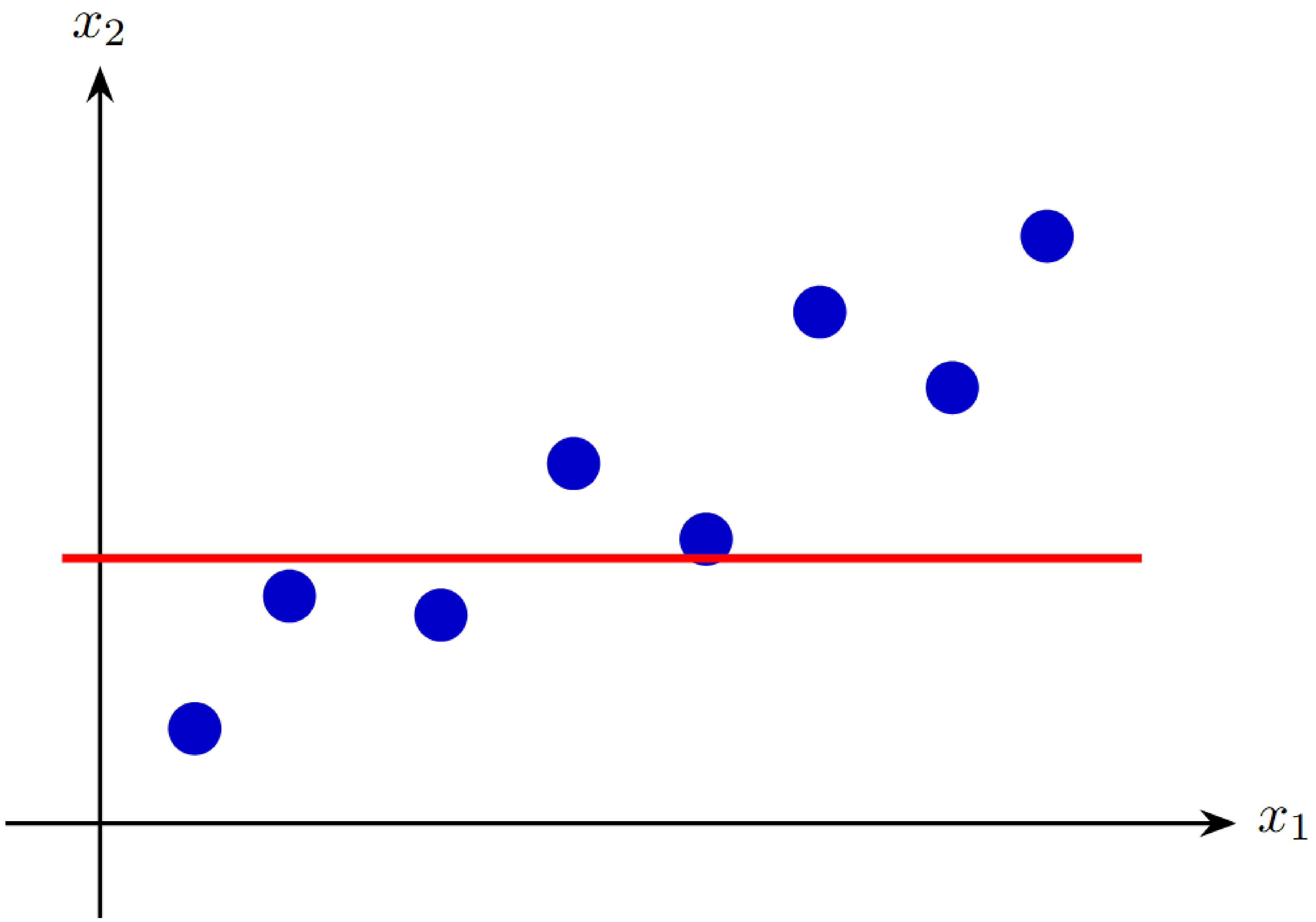
Obtener la mejor foto

No siempre será tan fácil



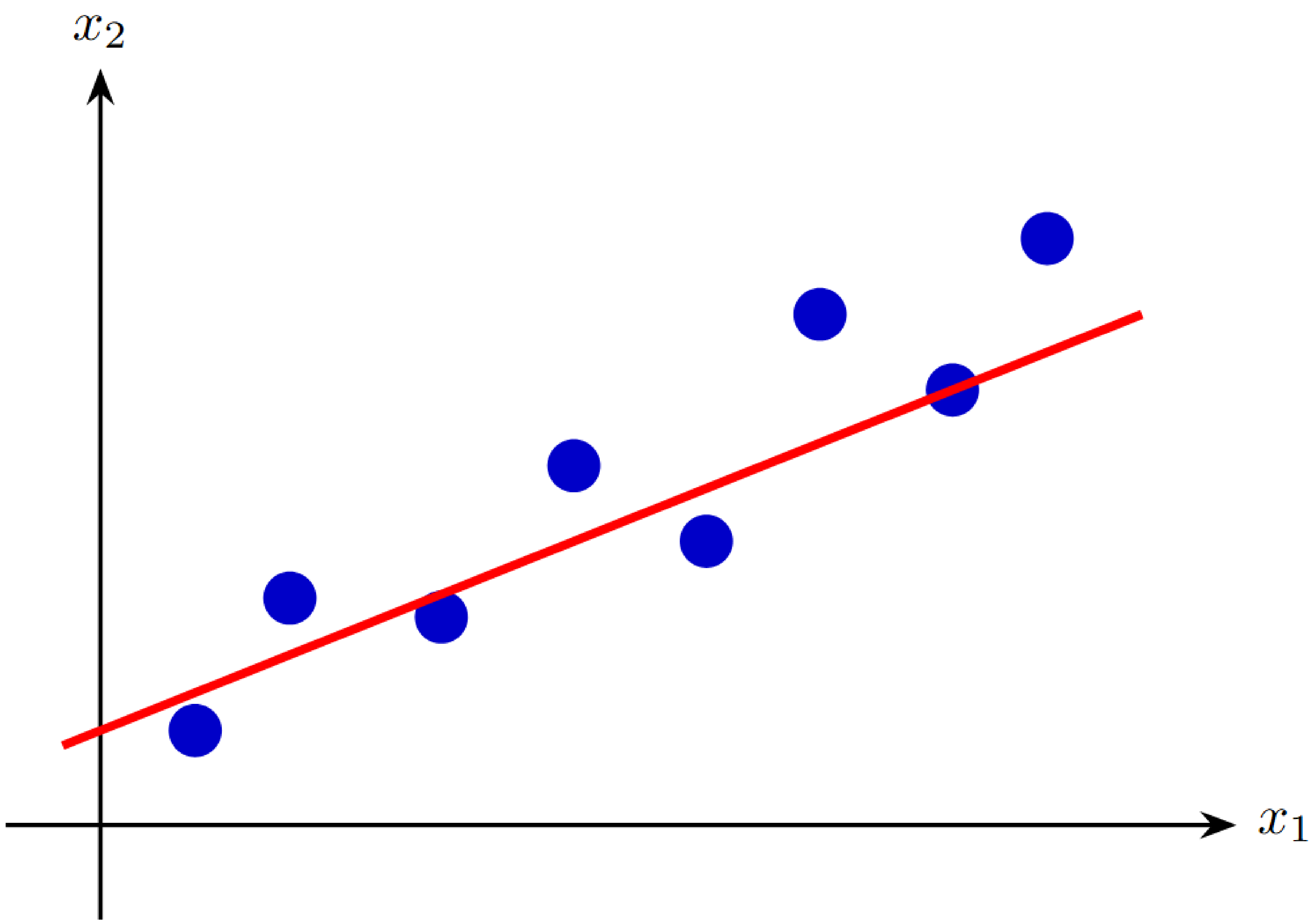
Obtener la mejor foto

No siempre será tan fácil



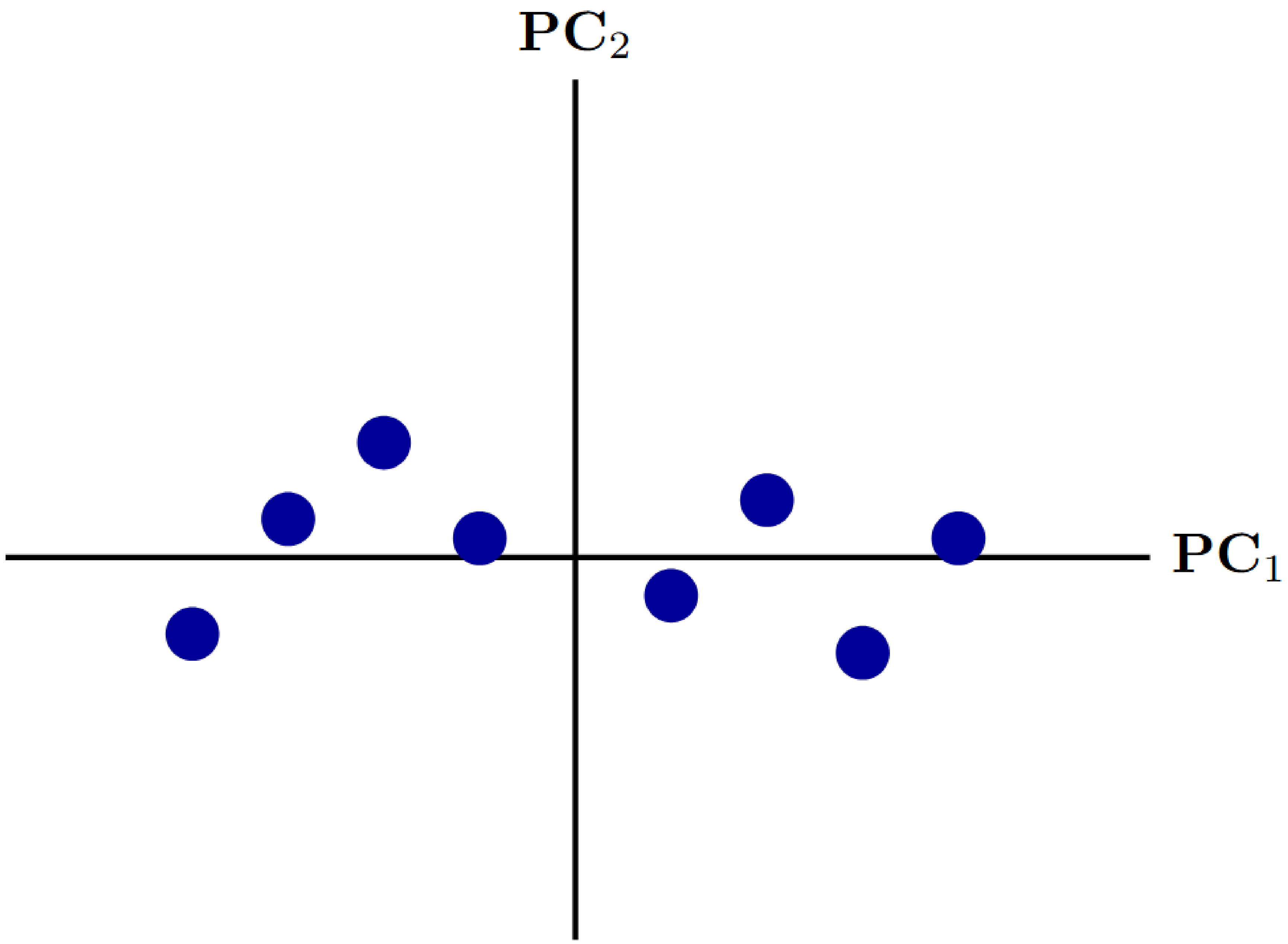
Obtener la mejor foto

No siempre será tan fácil



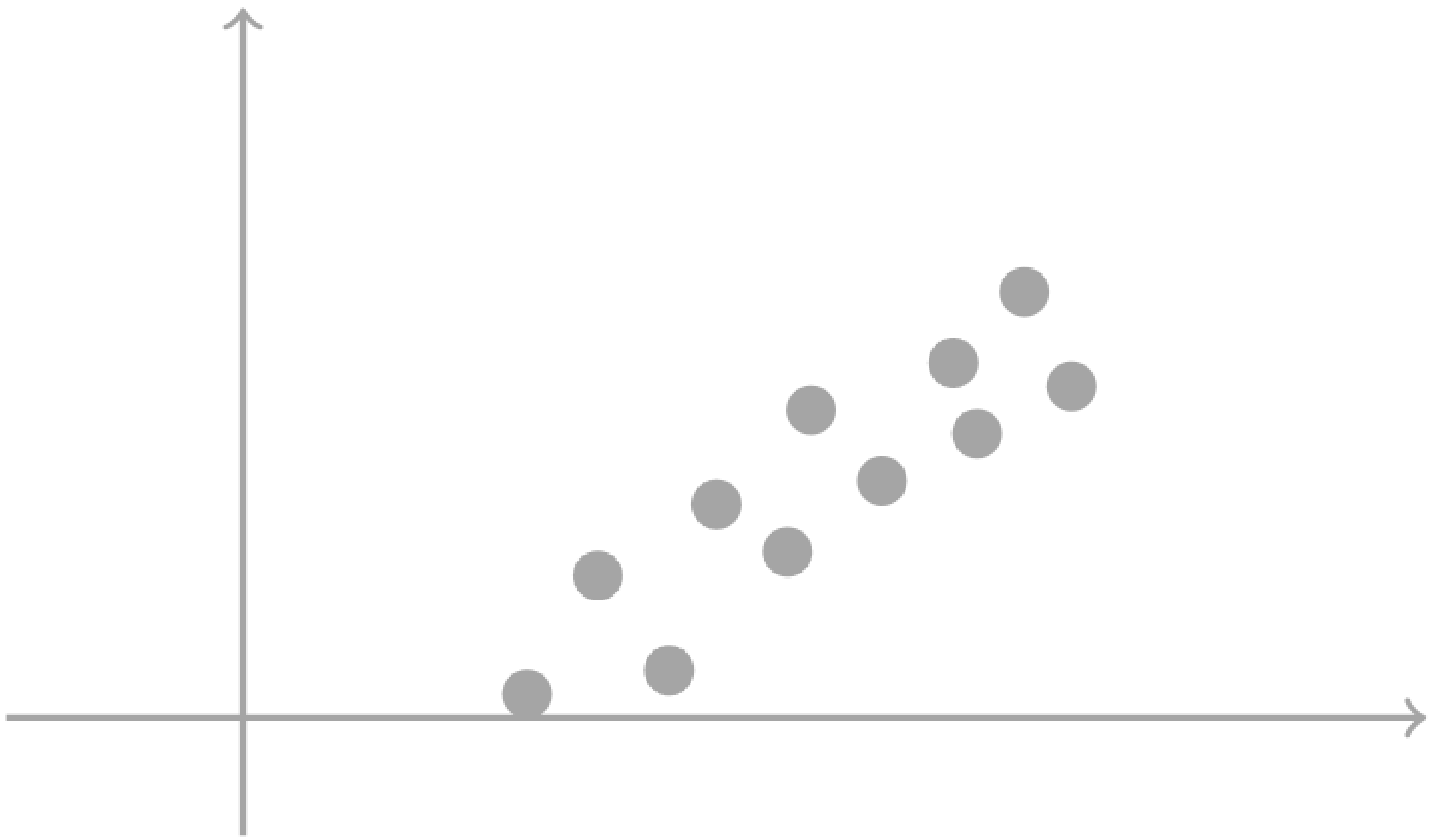
Obtener la mejor foto

No siempre será tan fácil



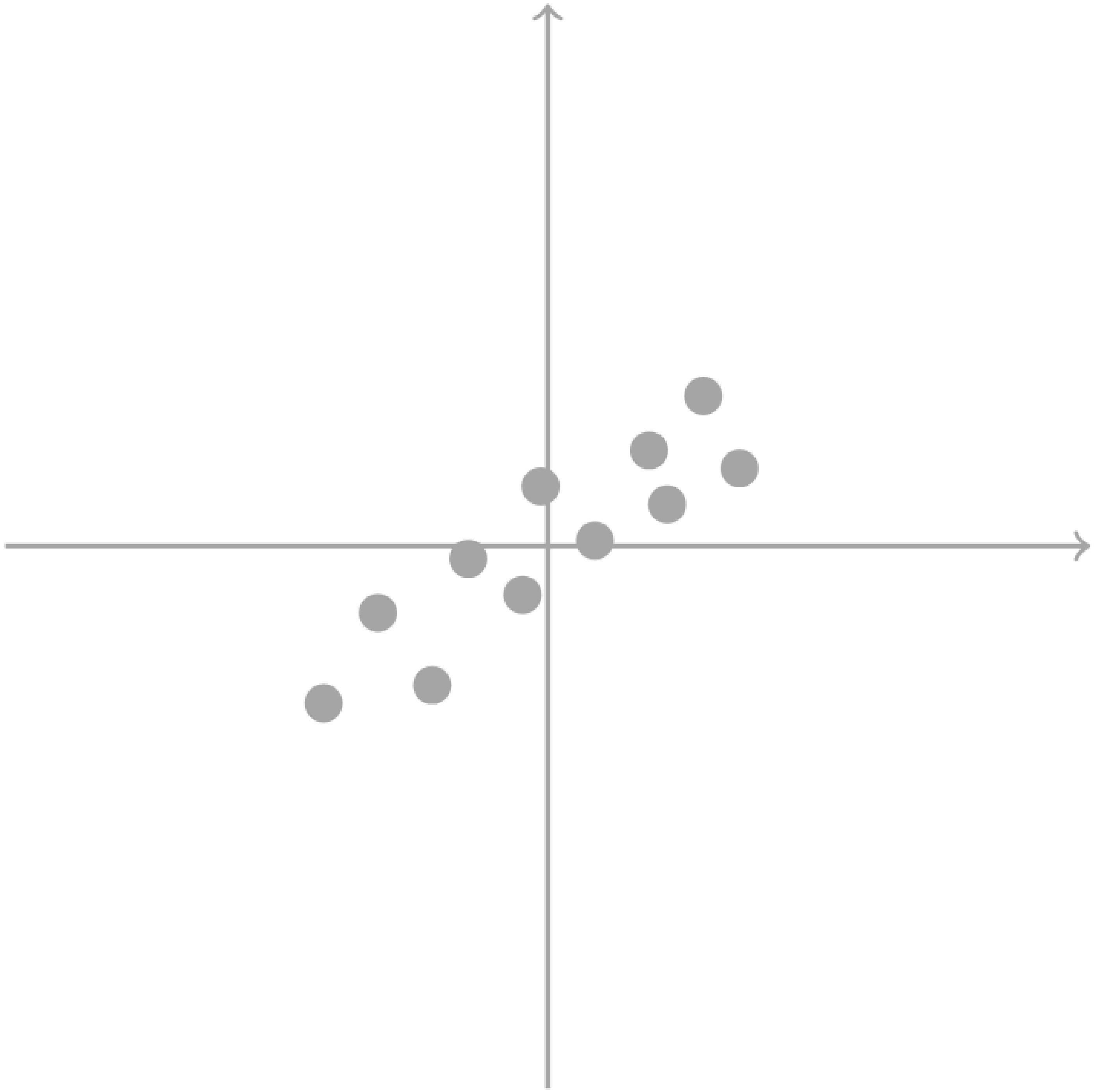
Deducción matemática

Tenemos un conjunto de datos



Deducción matemática

Los centramos restando el promedio

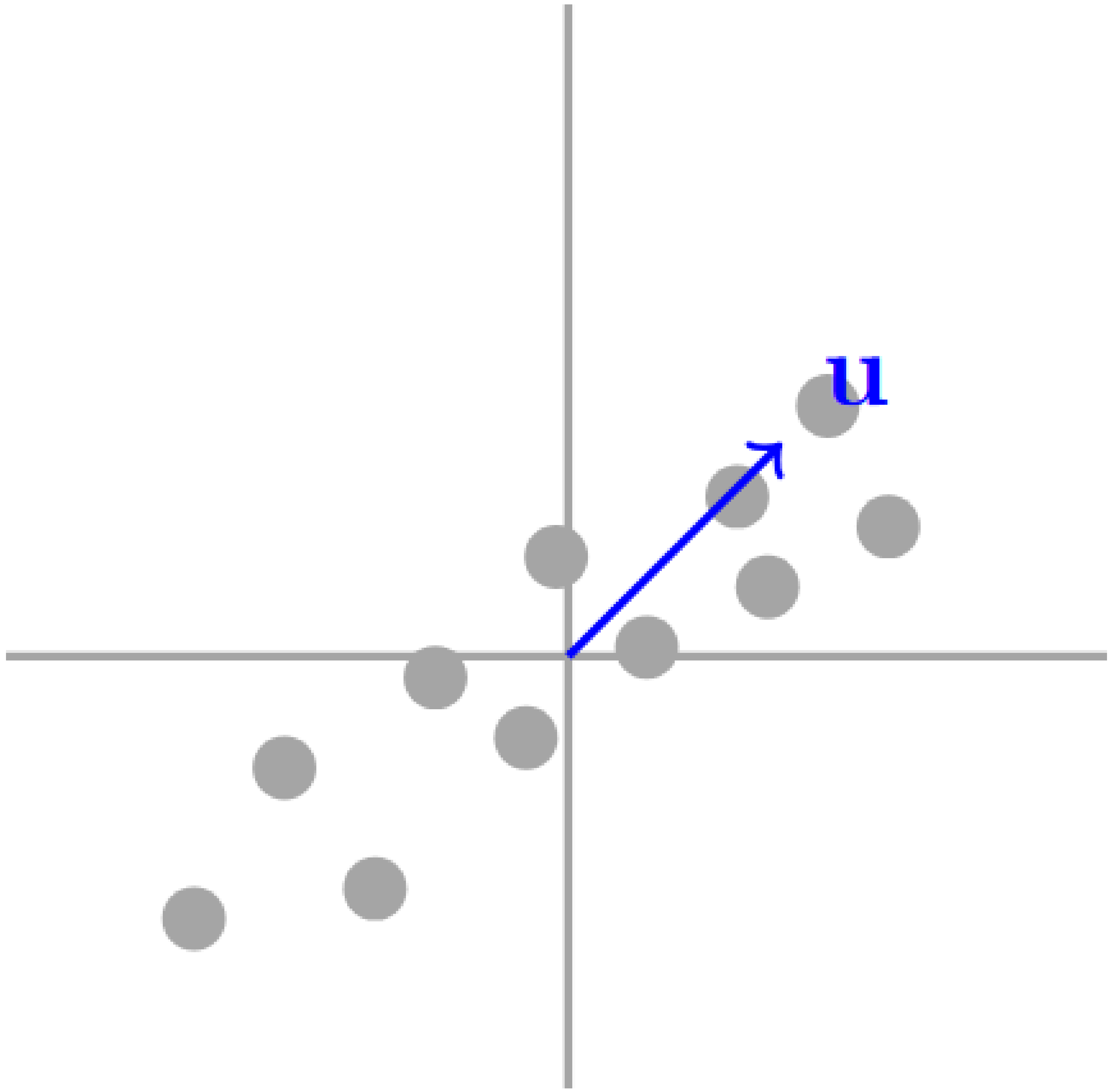


$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$x_i \leftarrow x_i - \bar{x}$$

Deducción matemática

Supongamos un vector unitario u



$$||u||^2 = u^t u = 1$$

Deducción matemática

Calculemos la varianza

$$\mathrm{Var}(u) = \frac{1}{n}\sum_{i=1}^{n}(u^\top x_i)^2$$

Deducción matemática

Realicemos una manipulación algebraica

$$\mathrm{Var}(u) = \frac{1}{n}\sum_{i=1}^{n}(u^\top x_i)(x_i^\top u)$$

Deducción matemática

Realicemos una manipulación algebraica

$$\operatorname{Var}(u) = u^\top \left[\frac{1}{n} \sum_{i=1}^{n} x_i x_i^\top \right] u$$

Deducción matemática

Ahora planteamos nuestro problema de optimización

$$\text{Max} \quad f(u) = u^\top \mathbf{C} u$$

$$||u||^2 = u^t u = 1$$

Deducción matemática

Le agradecemos a Lagrange por haber existido

$$\mathcal{L}(u, \lambda) = u^\top \mathbf{C} u - \lambda(u^\top u - 1)$$

$$\frac{\partial \mathcal{L}}{\partial u} = \mathbf{2C}u - \mathbf{2\lambda}u = 0$$

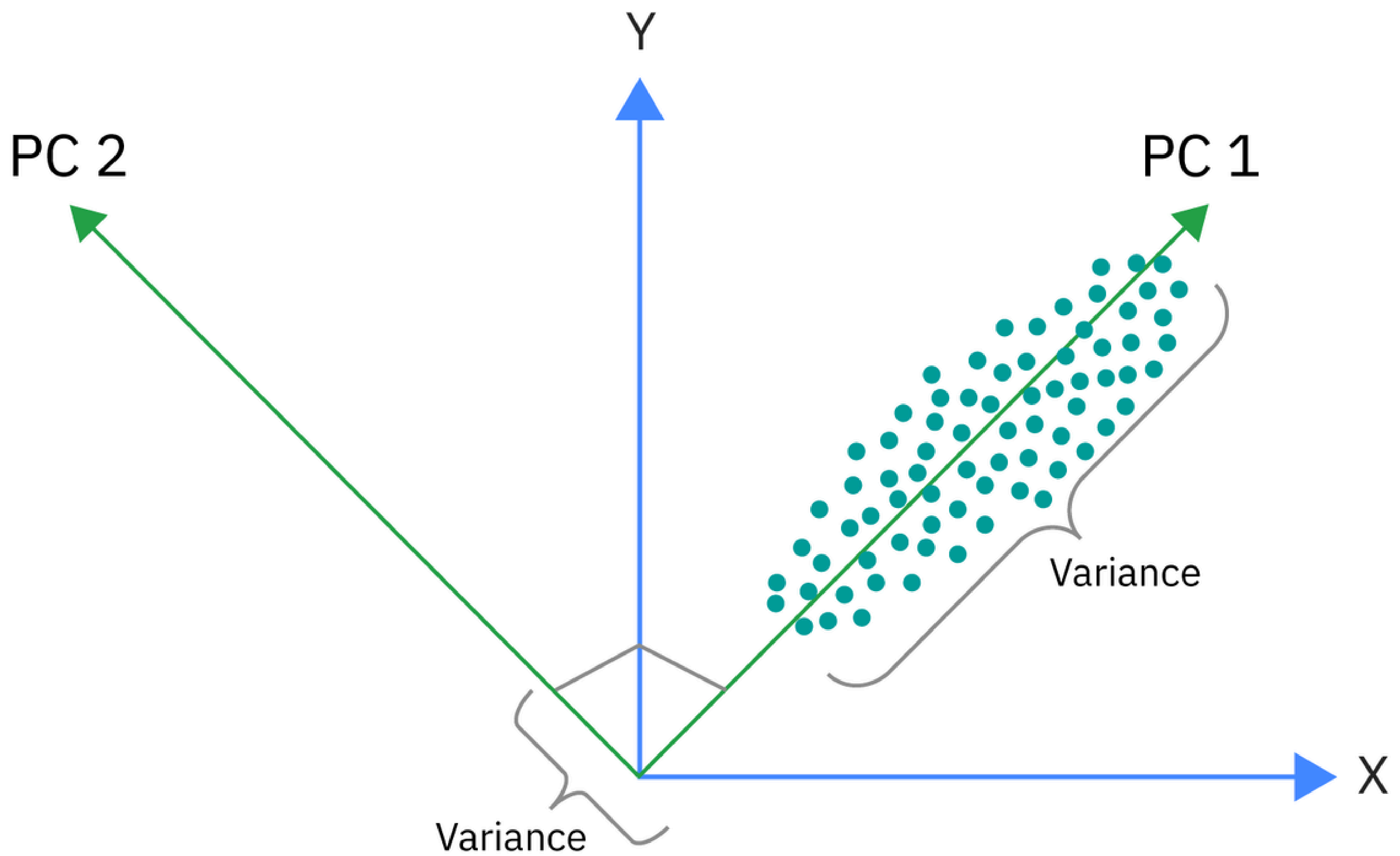
Deducción matemática

Dando como resultado un señor conocido

$$\mathbf{C}u = \lambda u$$

Supuestos

- Linealidad
- Las grandes varianzas tienen una estructura importante
- Los componentes principales son ortogonales



Resumen

1. Organice los datos como una matriz d×n.
2. Reste la media para cada tipo de medición.
3. Calcular la SVD o los vectores propios de la covarianza.

$$\textbf{Complexity} = \mathrm{O}(\mathrm{nd}^2 + d^3)$$

$$n = \text{Number of samples} \qquad d = \text{Number of features}$$

Aplicación

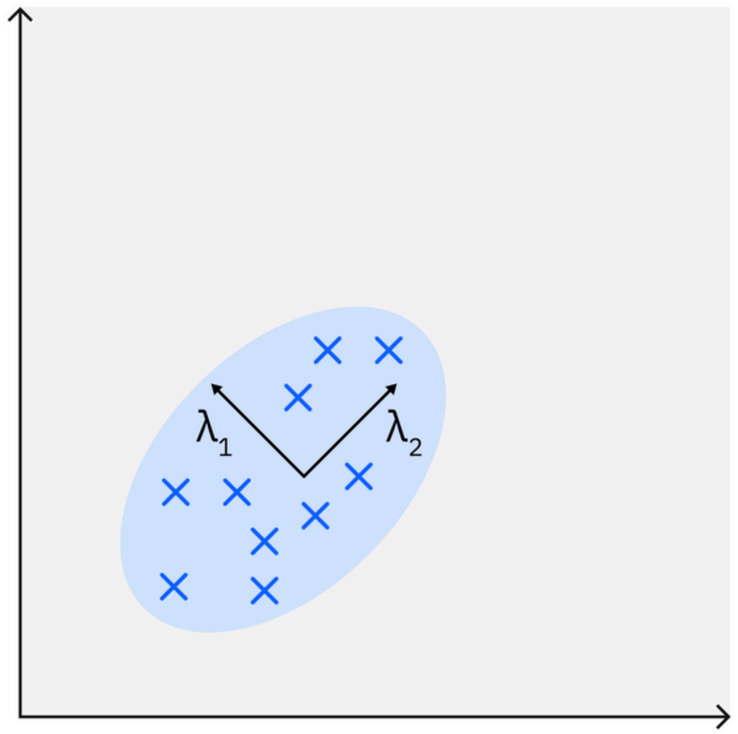
¿Cuál es la principal ventaja del PCA?

- Reducir el número de variables de los datos eliminando las que son ruido, lo que también puede reducir el sobreajuste.
- Mejorar el rendimiento de los algoritmos centrándose solo en las características relevantes.
- Mejorar la visualización de los datos para comprenderlos mejor.

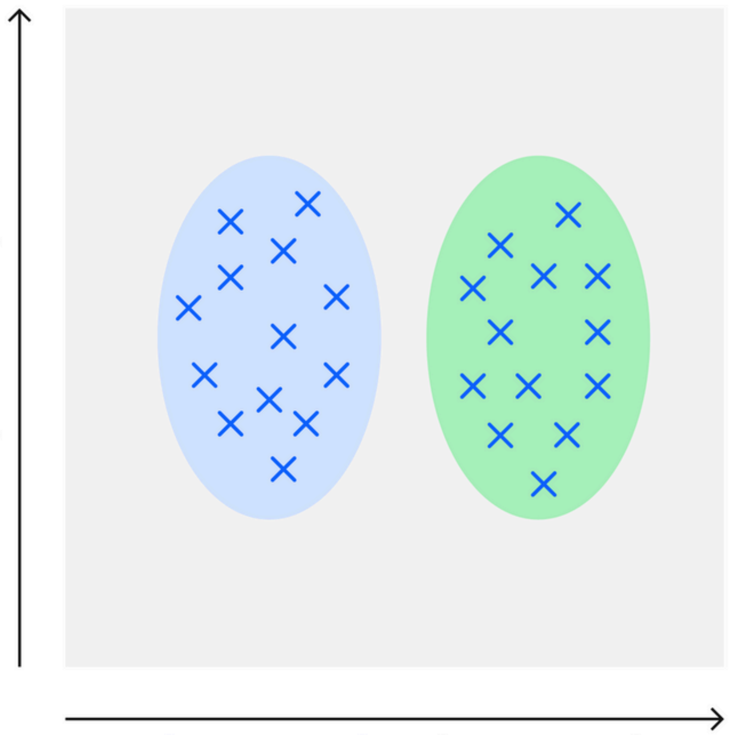
Aplicación

¿Cuándo usarlo?

Se recomienda utilizar el PCA cuando se trate con variables fuertemente correlacionadas. En caso de correlación débil, el PCA puede no ser lo mejor para reducir los datos



PCA



LDA

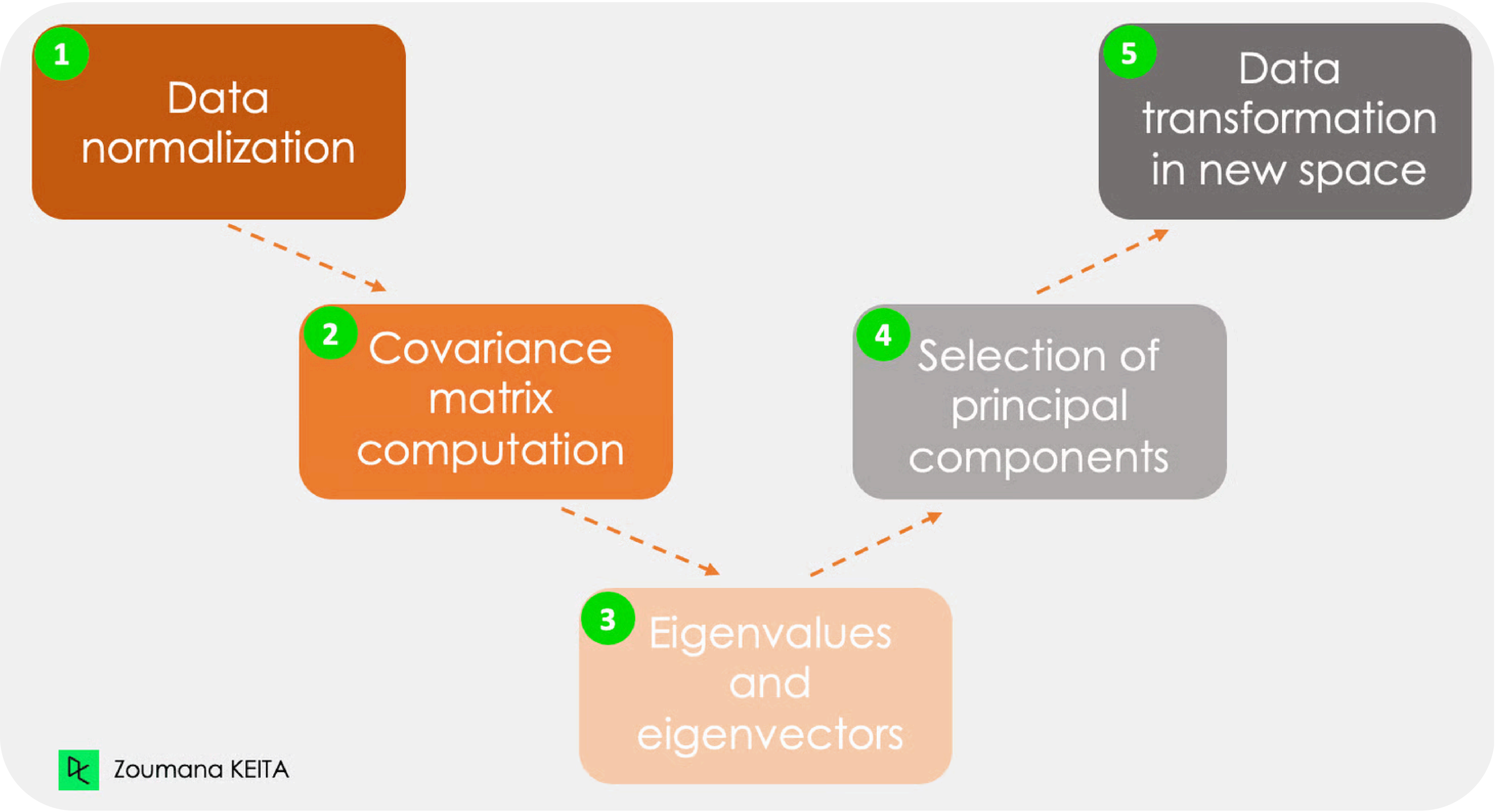
Aplicación

¿Cuáles son las limitaciones del PCA?

- Solo funciona con variables numéricas.
- No funciona bien cuando las variables no están fuertemente correlacionadas
- Es sensible a la escala de las características y el resultado se ve afectado por los valores atípicos.

Ejemplo práctico

Adaptación:Tutorial de análisis de componentes principales en R, DataCamp



¡MUCHAS Gracias!

Material y diapositivas de la clase