

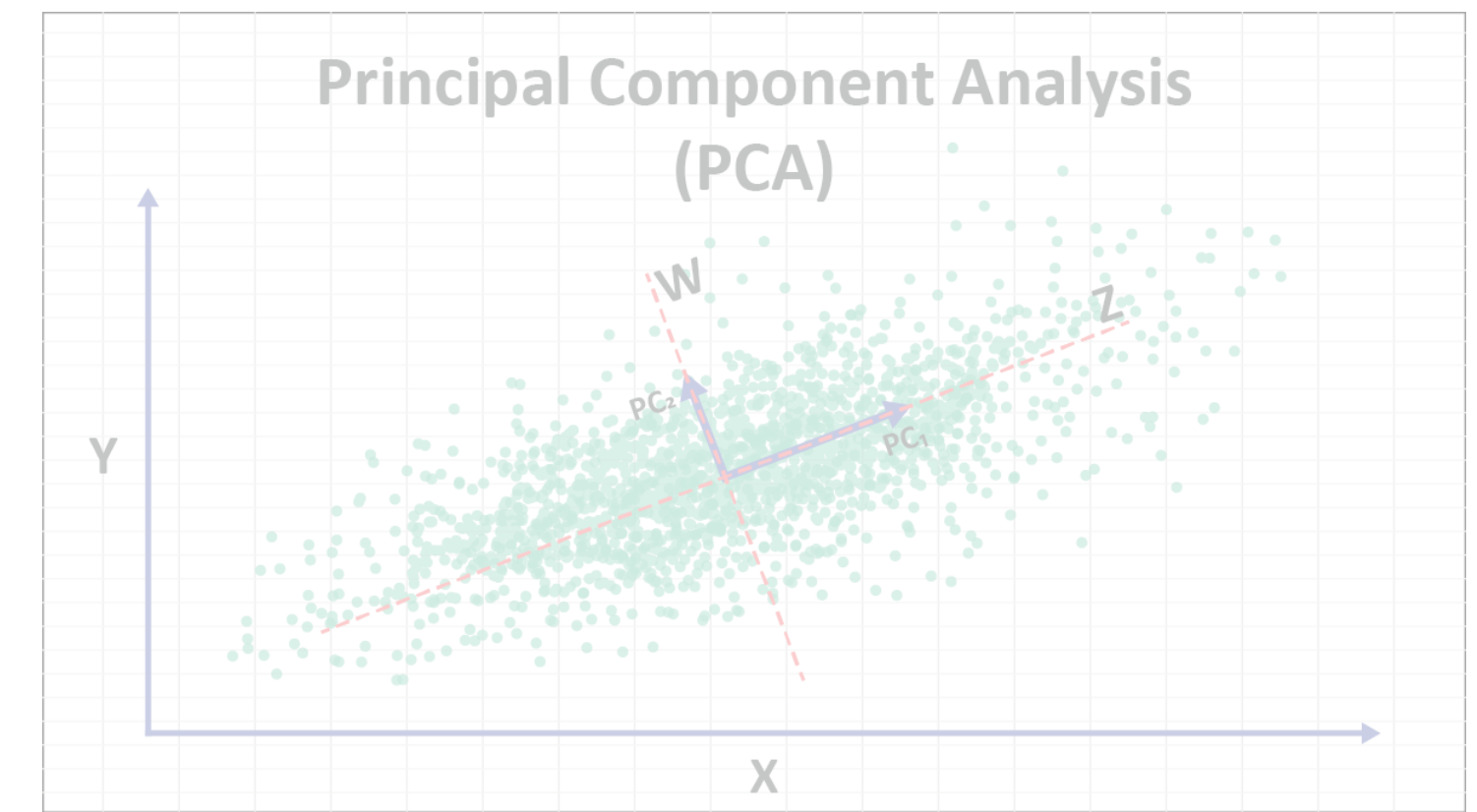


# Análisis de Componentes Principales (PCA)

Licenciatura en Inteligencia Artificial y Ciencia de Datos, CUGDL,  
Universidad de Guadalajara.

Guadalajara, Jal., agosto de 2025

# Introducción



- En estadística y aprendizaje automático, el **Análisis de Componentes Principales** (PCA) es una técnica que transforma un conjunto de variables posiblemente correlacionadas en un nuevo conjunto de variables llamadas *componentes principales*. Estas componentes son **linealmente independientes** y se ordenan de manera que las primeras capturan la mayor parte de la variabilidad presente en los datos.
- El objetivo de PCA es encontrar una proyección de los datos en un espacio de menor dimensión que conserve la mayor cantidad posible de información (varianza). Gracias a esto, se convierte en una herramienta muy útil para reducir la **dimensionalidad** de un dataset, simplificando su análisis y visualización.

# Introducción

## El problema de la dimensión

- Normalmente, un conjunto de datos está formado por una muestra aleatoria de  $n$  observaciones descritas mediante  $m$  variables. Esto se puede representar de forma matricial como,

$$X = \{x_{i,j}\}_{j=1,2,\dots,m}^{i=1,2,\dots,n}$$

- Donde cada fila corresponde a una observación y cada columna a una variable.
- A partir de estas variables uno puede construir una **matriz de correlaciones**, que incluye todas las posibles combinaciones entre las  $m$  variables. El número de pares distintos crece de acuerdo con la fórmula de combinatoria

$${}_m C_2 = \binom{m}{2} = \frac{m!}{2!(m-2)!}$$

- Esto implica que, conforme aumenta el número de variables  $m$ , el número de cálculos necesarios crece de manera cuadrática, lo que hace evidente la utilidad de técnicas como PCA para simplificar el problema.

# Introducción

## Pearson

- El concepto de “información” que aporta una variable dentro de un dataset está directamente relacionado con su **varianza**. En general, a mayor variabilidad en los datos, se considera que la variable contiene más información. Este enfoque fue introducido por **Karl Pearson** en el siglo XIX, al estudiar cómo representar datos multivariados de forma más eficiente.
- Cuando varias variables están altamente correlacionadas, comparten parte de la misma información, lo que genera redundancia. Para resolver este problema, es posible transformar el conjunto original en un nuevo conjunto de variables que sean **no correlacionadas entre sí**.
- Estas nuevas variables, llamadas **componentes principales**, se construyen como combinaciones lineales de las variables originales y se ordenan de acuerdo con la cantidad de varianza que explican en la muestra.

# Componentes principales

## Cálculo

- Utilizando la siguiente notación, queremos pasar de un set de variables originales  $\{x_1, x_2, \dots, x_m\}$  a un set de componentes  $\{y_1, y_2, \dots, y_m\}$ . Cada una de las variables  $y$  será una combinación lineal de las variables originales  $x$ , de tal forma que:

$$y_i = \sum_{j=1}^m a_{ij}x_j = \vec{a}_i \cdot \vec{x}; \quad \|\vec{a}_i\| = 1 \quad \forall i = 1, 2, \dots, m$$

- Se pide que el orden de las componentes  $y_i$  vaya en orden descendente con respecto a su varianza.
- Un método habitual para maximizar una función de variables sujeta a restricciones es el método de **multiplicadores de Lagrange**.

# Multiplicadores de Lagrange

## Método

- El **método de multiplicadores de Lagrange** permite encontrar los máximos y mínimos de una función multivariable de la forma,

$$f(x, y, \dots)$$

- cuando los valores de sus variables están sujetos a una restricción. Dicha restricción debe expresarse como,

$$g(x, y, \dots) = c$$

- donde  $g$  es una función multivariable definida sobre los mismos parámetros de  $f$ , y  $c$  es una constante. Este procedimiento introduce un nuevo parámetro, llamado **multiplicador de Lagrange**, que permite incorporar la restricción dentro de la función original y resolver el problema de optimización de manera conjunta.

# Multiplicadores de Lagrange

## Pasos

1. Introducir una nueva variable  $\lambda$  y definir una nueva función  $\mathcal{L}$  de la forma:

$$\mathcal{L}(x, y, \dots, \lambda) = f(x, y, \dots) - \lambda (g(x, y, \dots) - c)$$

Esta nueva función  $\mathcal{L}$  es llamado el "**Lagrangiano**" y la nueva variable  $\lambda$  es llamado "**multiplicador de Lagrange**".

2. Poner el gradiente de  $\mathcal{L}$  igual a un vector de cero, i.e., encontrar los puntos críticos,

$$\nabla \mathcal{L}(x, y, \dots, \lambda) = \vec{0}$$

3. Cada solución a esta ecuación tendrá una forma  $(x_0, y_0, \dots, \lambda_0)$ . Se introduce cada solución (excepto la  $\lambda$ ) en la función original  $f(x, y, \dots)$ , y cualquiera que dé el mayor (o el menor) valor, será el máximo (o mínimo) de  $f$ .

# Multiplicadores de Lagrange

## Ejemplo

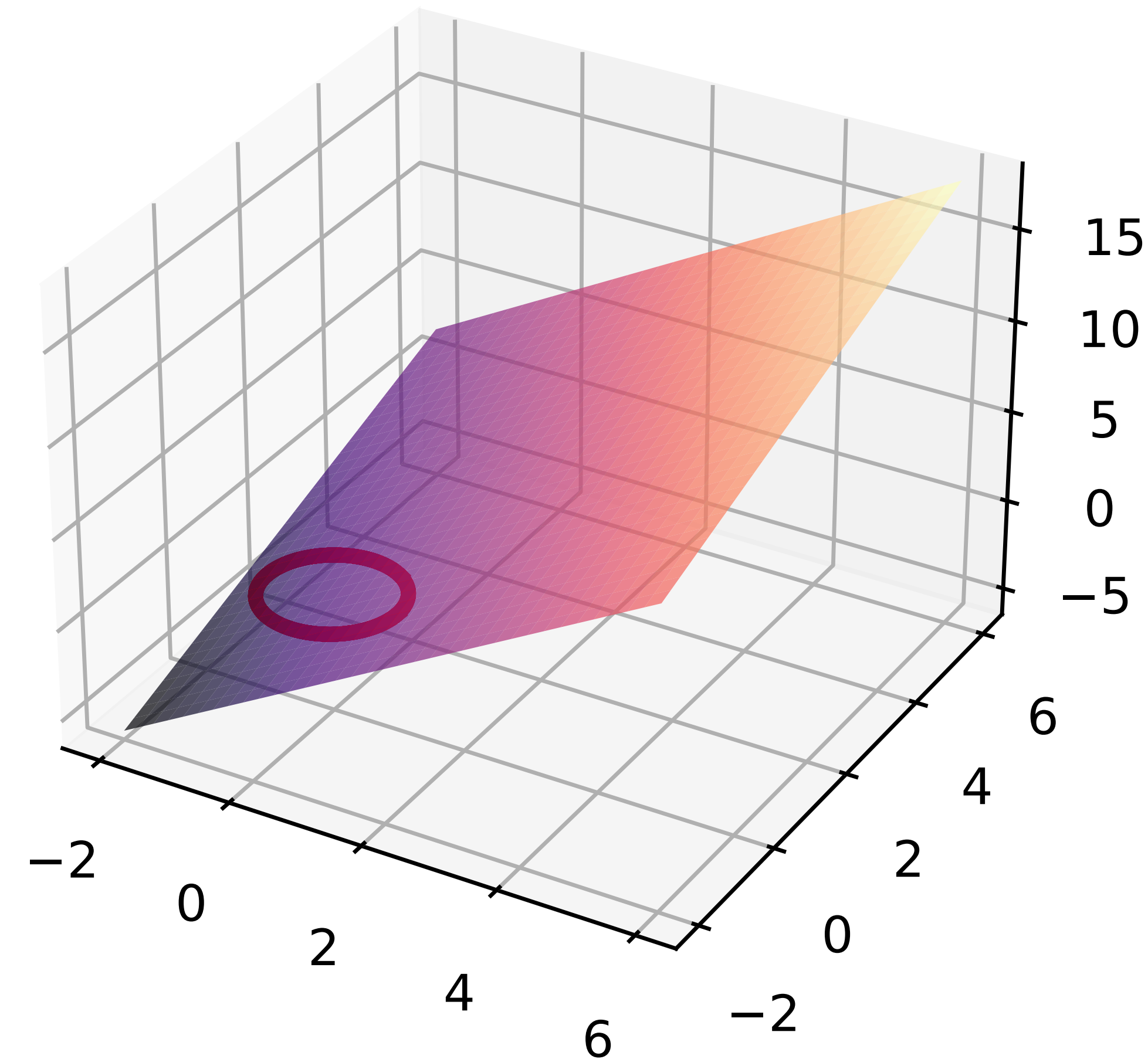
$$f(x,y) = 2x + y \text{ and circle}$$

- Queremos maximizar una función,

$$f(x,y) = 2x + y$$

- Pero estamos limitados a que los puntos  $(x,y)$  sólo pueden estar en un círculo de radio 1,

$$g(x,y) = x^2 + y^2 = 1$$



# Multiplicadores de Lagrange

## Ejemplo

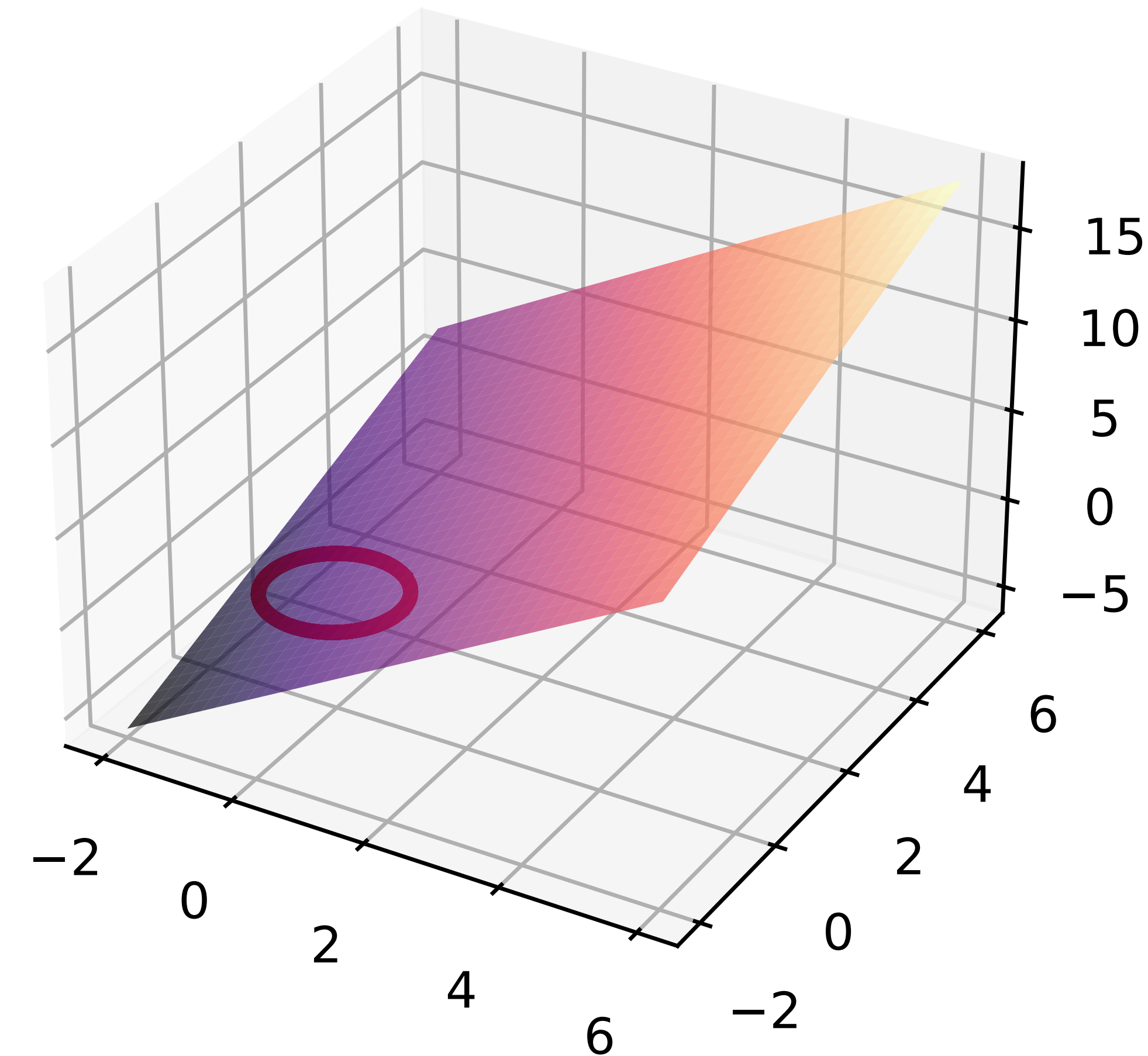
$f(x,y) = 2x + y$  and circle

- Construimos el Lagrangiano,

$$\begin{aligned}\mathcal{L} &= f(x, y) - \lambda(g(x, y) - c) \\ &= 2x + y - \lambda(x^2 + y^2 - 1)\end{aligned}$$

- Igualamos el gradiente de esta función a un vector de ceros,

$$\nabla \mathcal{L}(x, y, \lambda) = \begin{pmatrix} \frac{\partial}{\partial x} \mathcal{L} \\ \frac{\partial}{\partial y} \mathcal{L} \\ \frac{\partial}{\partial \lambda} \mathcal{L} \end{pmatrix} = \begin{pmatrix} 2 - 2\lambda x \\ 1 - 2\lambda y \\ -x^2 - y^2 + 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$



# Multiplicadores de Lagrange

## Ejemplo

$f(x,y) = 2x + y$  and circle

$$\nabla \mathcal{L}(x, y, z) = \begin{pmatrix} \frac{\partial}{\partial x} \mathcal{L} \\ \frac{\partial}{\partial y} \mathcal{L} \\ \frac{\partial}{\partial \lambda} \mathcal{L} \end{pmatrix} = \begin{pmatrix} 2 - 2\lambda x \\ 1 - 2\lambda y \\ -x^2 - y^2 + 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

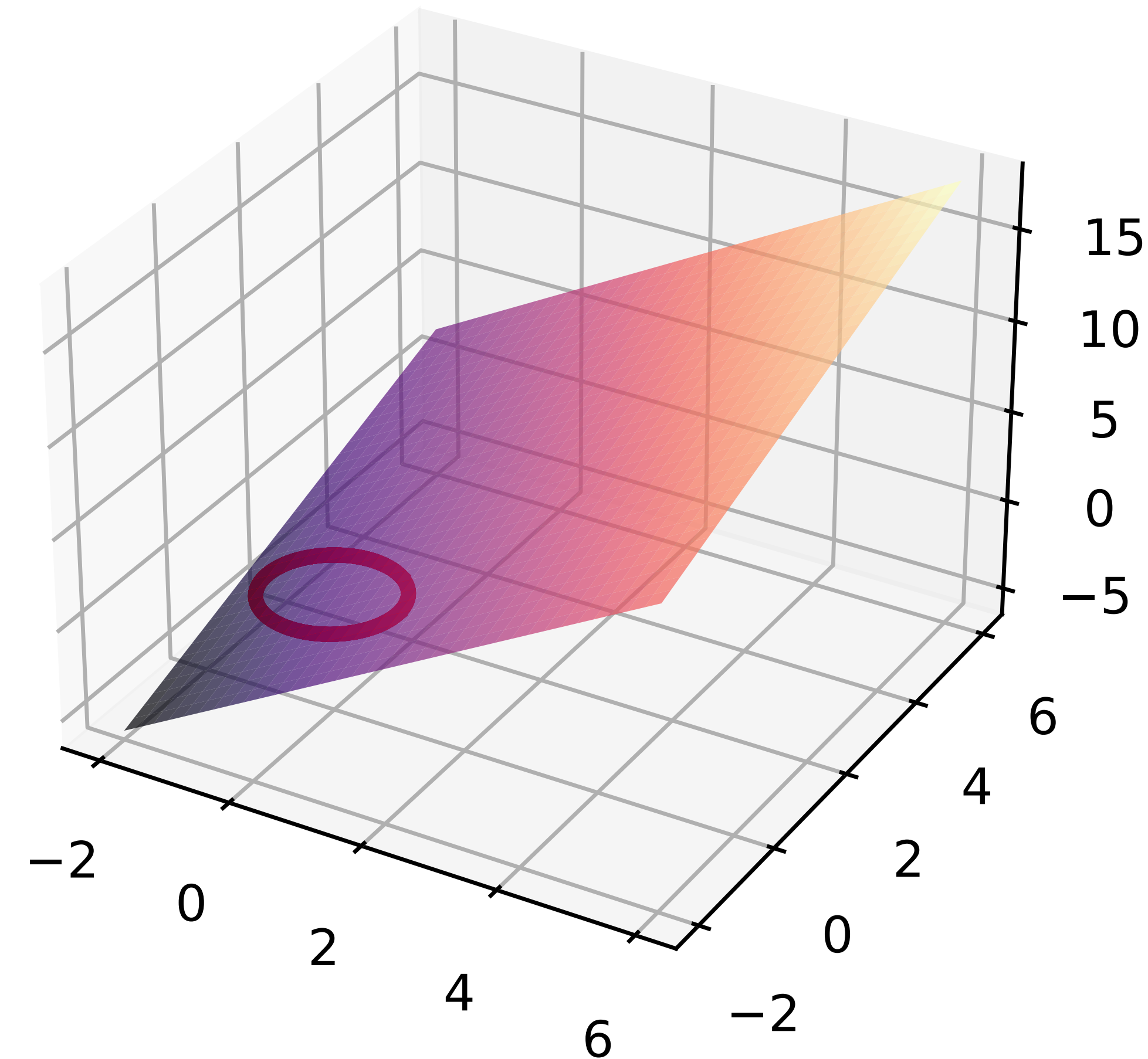
- Esta sistema de ecuaciones se satisface con,

$$x_0 = \frac{1}{\lambda_0}, y_0 = \frac{1}{2\lambda_0}$$

- Introduciendo esto en la ecuación  $x^2 + y^2 = 1$ , se obtiene cuadrática para  $\lambda_0$ ,

$$\frac{1}{\lambda_0^2} + \frac{1}{4\lambda_0^2} = 1$$

- Y entonces  $\lambda_0$  puede tener dos soluciones  $\lambda_0 = \frac{\pm\sqrt{5}}{2}$



# Multiplicadores de Lagrange

## Ejemplo

$$f(x,y) = 2x + y \text{ and circle}$$

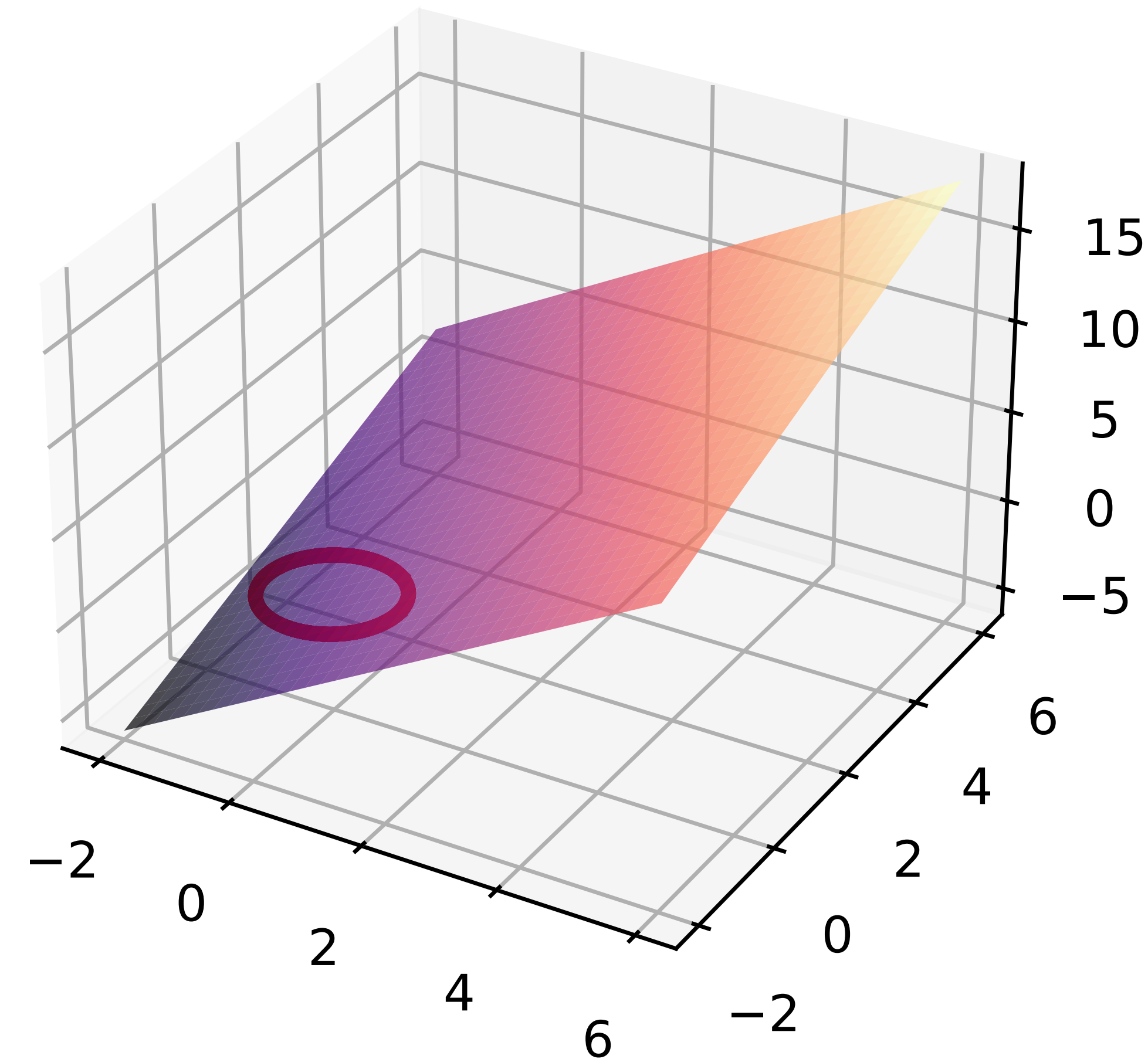
- Por lo tanto, tenemos dos soluciones,

$$\begin{aligned}(x_0, y_0) &= \left( \frac{1}{\lambda_0}, \frac{1}{2\lambda_0} \right) \\ &= \left( \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right) \text{ ó } \left( \frac{-2}{\sqrt{5}}, \frac{-1}{\sqrt{5}} \right)\end{aligned}$$

- Introduciendo estos vectores en  $f(x,y)$  vemos que el primero de éstos dá el resultado mayor, y por tanto el máximo.

$$f\left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right) = \sqrt{5} \text{ máximo}$$

$$f\left(\frac{-2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}\right) = -\sqrt{5} \text{ mínimo}$$



# Componentes principales

## Cálculo (Introducción)

- Teníamos una serie de variables  $x_i$  (correlacionadas) y queremos obtener otra serie de variables  $y_i$  (no correlacionadas),

$$(x_1, x_2, \dots, x_m) \rightarrow (y_1, y_2, \dots, y_m)$$

donde cada  $y_i$  es una combinación lineal de las variables originales  $x_i$ ,

$$y_i = \sum_{j=1}^m a_{ij}x_j = \mathbf{a}_i^T \cdot \mathbf{x} = (a_{i1}, a_{i2}, \dots, a_{im}) \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$$

- Tenemos la restricción,

$$\|\mathbf{a}_i\| = \mathbf{a}_i^T \mathbf{a}_i = 1 \quad \forall i = 1, 2, \dots, m.$$

Esto asegura que la varianza de cada componente principal esté determinada únicamente por la dirección de  $\mathbf{a}_i$  y no por su escala.

# Componentes principales

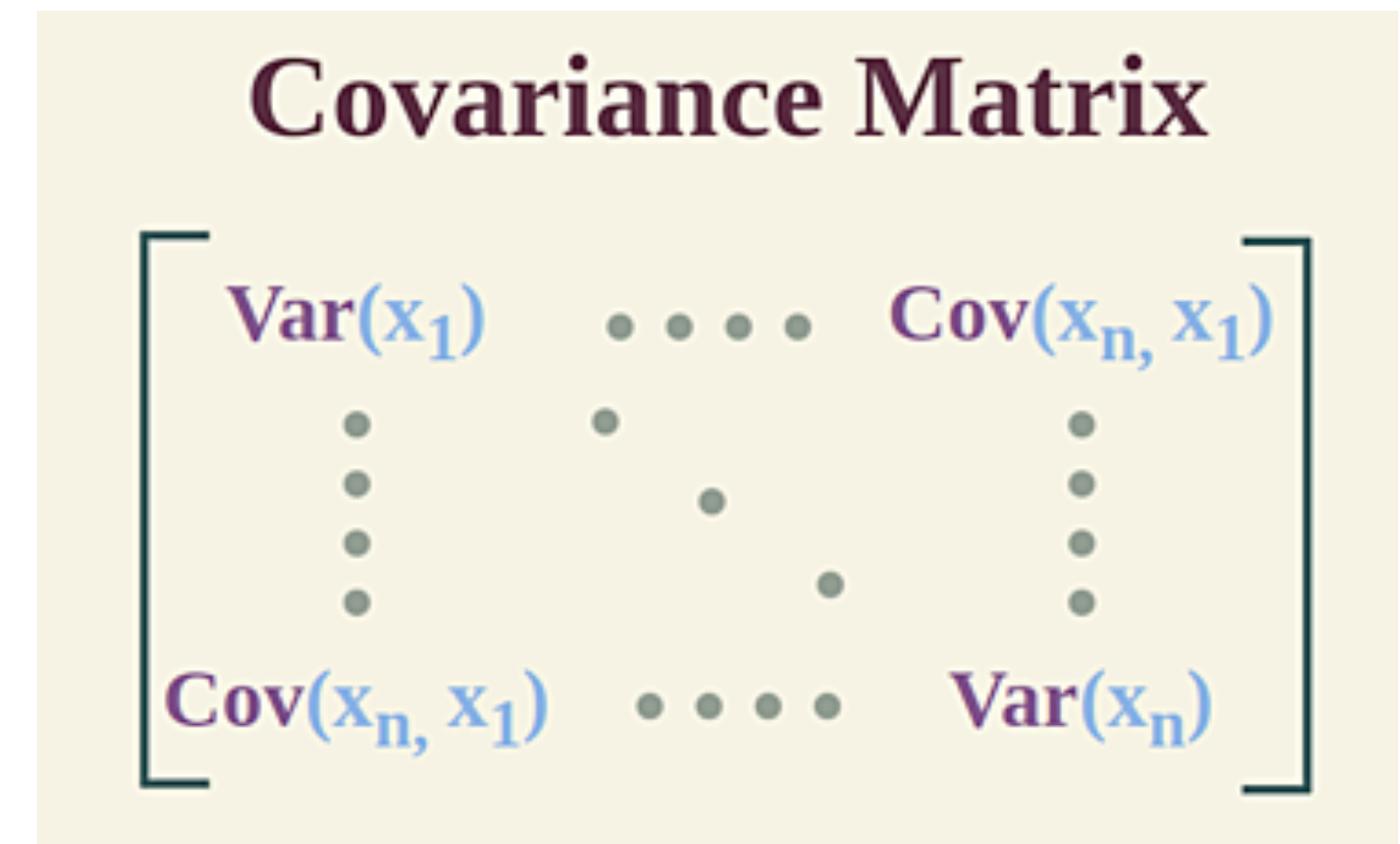
## Varianza

- Se elige el primer valor  $\mathbf{a}_1$  de modo que se maximize la varianza de  $y_1$ ,

$$\text{var}(y_1) = \text{var}(\mathbf{a}_1 \cdot \mathbf{x}) = \mathbf{a}_1^T \cdot \Sigma \cdot \mathbf{a}_1$$

donde  $\Sigma$  es la matriz de covarianza asociada a las variables  $x_i$

$$\begin{aligned} \text{Var}[X] &= E \begin{bmatrix} (X_1 - E[X_1])(X_1 - E[X_1]) & \dots & (X_1 - E[X_1])(X_K - E[X_K]) \\ \vdots & \ddots & \vdots \\ (X_K - E[X_K])(X_1 - E[X_1]) & \dots & (X_K - E[X_K])(X_K - E[X_K]) \end{bmatrix} \\ &= \begin{bmatrix} E[(X_1 - E[X_1])^2] & \dots & E[(X_1 - E[X_1])(X_K - E[X_K])] \\ \vdots & \ddots & \vdots \\ E[(X_K - E[X_K])(X_1 - E[X_1])] & \dots & E[(X_K - E[X_K])^2] \end{bmatrix} \\ &= \begin{bmatrix} \text{Var}[X_1] & \dots & \text{Cov}[X_1, X_K] \\ \vdots & \ddots & \vdots \\ \text{Cov}[X_K, X_1] & \dots & \text{Var}[X_K] \end{bmatrix} \end{aligned}$$



# Componentes principales

## Cálculo (Lagrange multipliers)

- Queremos el valor máximo de  $\mathbf{a}_1^T \cdot \Sigma \cdot \mathbf{a}_1$  sujeto a una restricción  $\mathbf{a}_1^T \mathbf{a}_1 = 1$ .
- Construimos el Lagrangiano

$$\mathcal{L}(\mathbf{a}_1) = \mathbf{a}_1^T \cdot \Sigma \cdot \mathbf{a}_1 - \lambda(\mathbf{a}_1^T \mathbf{a}_1 - 1)$$

- Y luego buscamos el punto crítico

$$\frac{\partial \mathcal{L}}{\partial \mathbf{a}_1} = \Sigma \cdot \mathbf{a}_1 + \mathbf{a}_1^T \cdot \Sigma - \lambda \mathbf{a}_1 - \mathbf{a}_1^T \lambda = 0$$

donde  $\Sigma$  es una matriz simétrica, con propiedad  $\Sigma = \Sigma^T$  y  $(AB)^T = B^T A^T$ .

$$\Rightarrow 2\Sigma \cdot \mathbf{a}_1 - 2\lambda \mathbb{I} \cdot \mathbf{a}_1 = 0$$

$$\Rightarrow (\Sigma - \lambda \mathbb{I})\mathbf{a}_1 = 0$$

# Componentes principales

## Cálculo (Valores propios)

- Entonces, tenemos que resolver un sistema de ecuaciones

$$\Sigma - \lambda \mathbb{I} = 0$$

- Para que este sistema tenga una solución no trivial (donde todos los  $a_i$  son ceros), esta matriz tiene que tener un determinante igual a cero (matriz singular) (Teorema de Rouché-Frobenius),

$$\det | \Sigma - \lambda \mathbb{I} | = 0$$

- Y esta es una definición tal cual para encontrar los valores propios  $\lambda$  de la matriz  $\Sigma$ . El multiplicador de Lagrange  $\lambda$  que necesitamos coincide con un valor propio de la matriz de covarianza  $\Sigma$ .

# Componentes principales

## Cálculo (Alternativo, Eigenvalores y Eigenvectores)

$$\frac{\partial \mathcal{L}}{\partial \mathbf{a}_1} = \Sigma \cdot \mathbf{a}_1 + \mathbf{a}_1^T \cdot \Sigma - \lambda \mathbf{a}_1 - \mathbf{a}_1^T \lambda = 0 \quad \Rightarrow \quad \Sigma \mathbf{a}_1 = \lambda \mathbf{a}_1$$

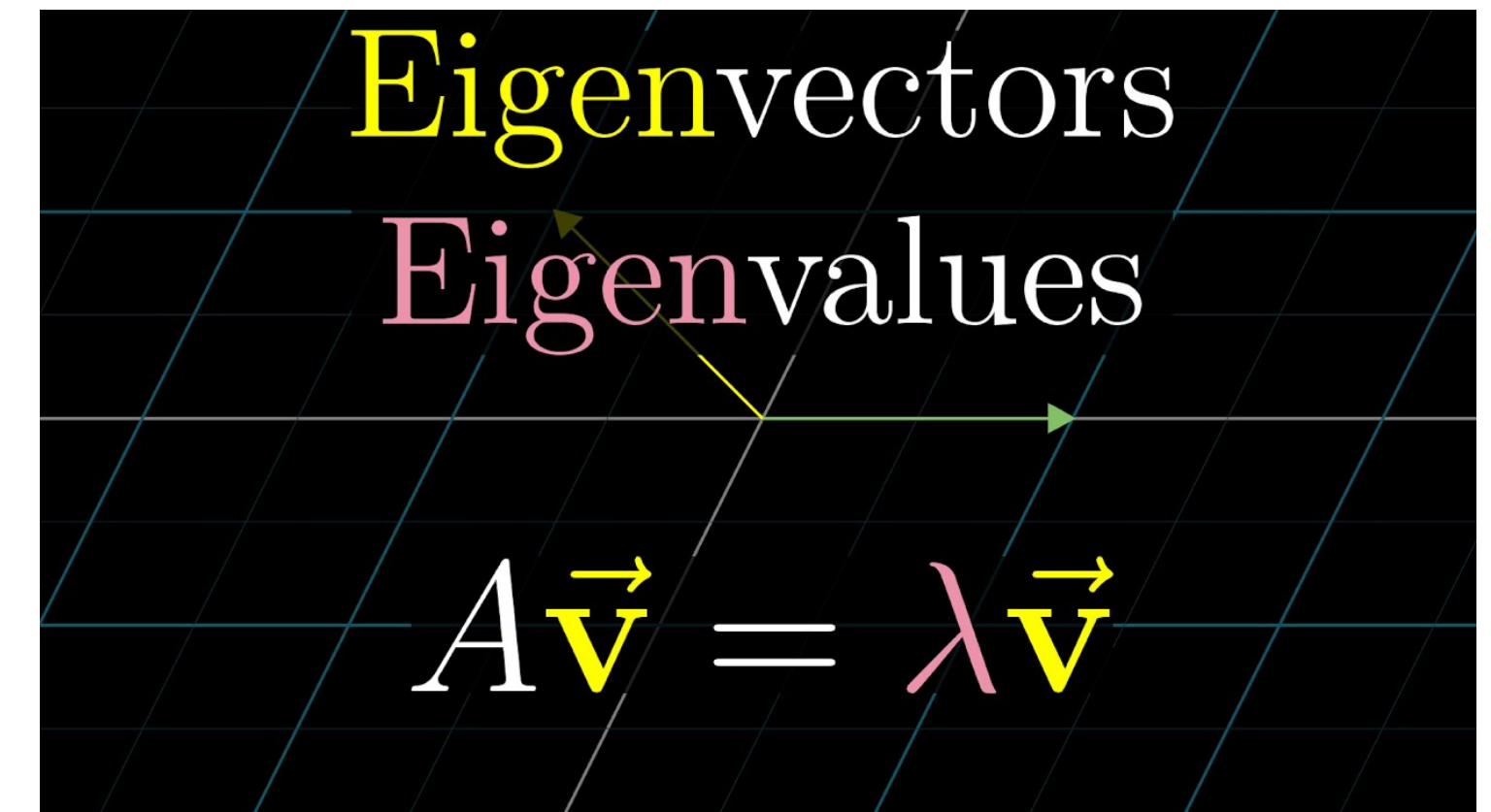
$$\frac{\partial \mathcal{L}}{\partial \lambda} = \mathbf{a}_1^T \mathbf{a}_1 - 1 = 0$$

- Otra forma de verlo puede ser como ecuación de eigenvectores,

$$\Sigma \mathbf{a}_1 = \lambda \mathbf{a}_1$$

- $\Sigma$  es una matriz de orden  $m$ . Dimensiones  $m \times m$ .
- $\mathbf{a}_1$  es un eigenvector asociado a esta matriz  $\Sigma$ .
- $\lambda$  es el eigenvalor asociado al eigenvector  $\mathbf{a}_1$ .

[Link](#)



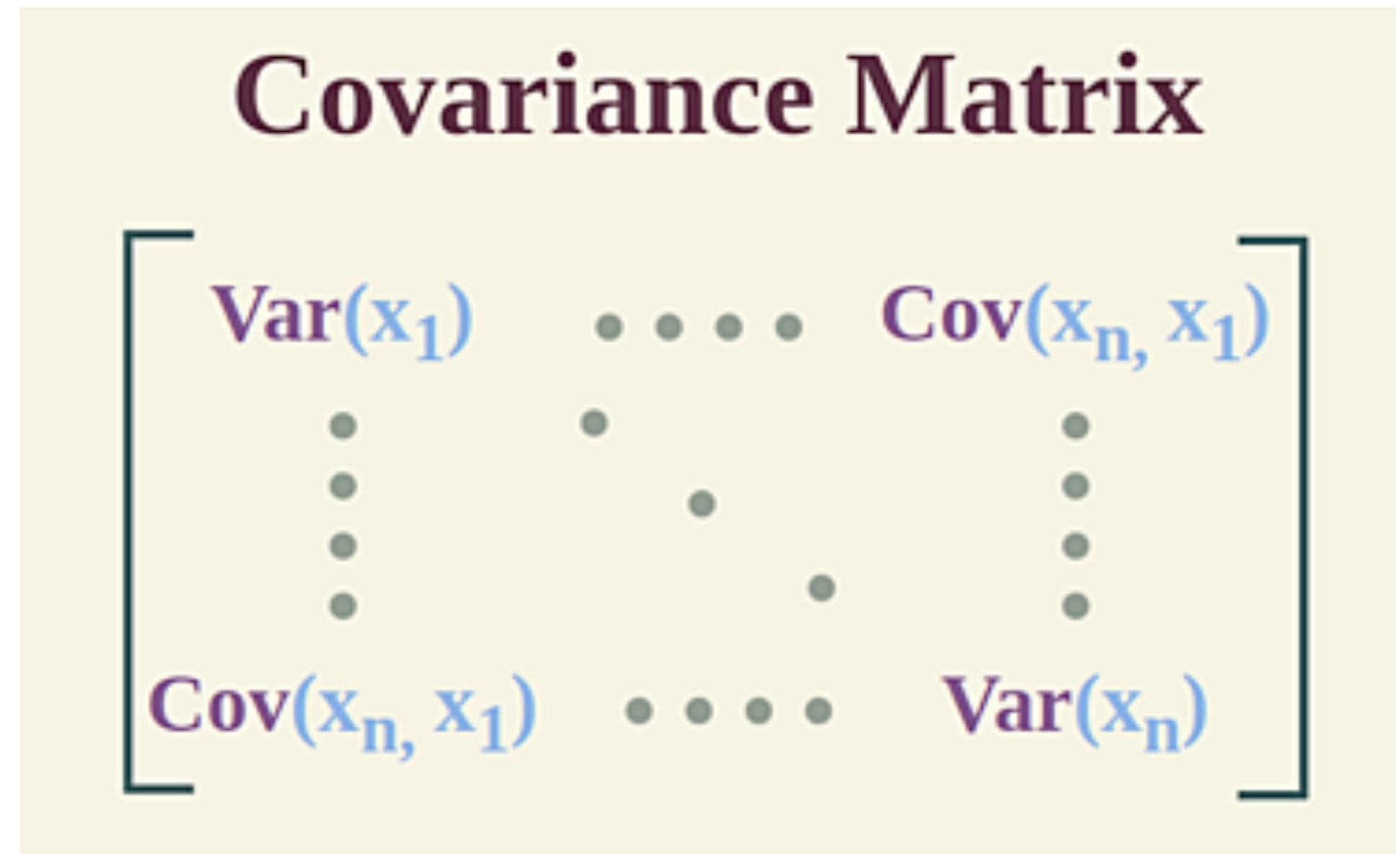
# Componentes principales

## Matriz de covarianza

- La matriz de covarianza  $\Sigma$  es de orden  $m$ .
- Es definida positiva, y se cumple que,

$$\mathbf{a}_i \cdot \Sigma \cdot \mathbf{a}_i^T > 0$$

- Por tanto, tiene exactamente  $m$  valores propios (eigenvalores) reales  $\lambda_i$ .



# Componentes principales

## Cálculo (valor propio máximo)

- Regresando a la varianza de la nueva variable  $y_1$ , ya podemos hacer esta sustitución,

$$\text{var}(y_1) = \text{var}(\mathbf{a}_1 \cdot \mathbf{x}) = \mathbf{a}_1^T \cdot (\Sigma \cdot \mathbf{a}_1) = \mathbf{a}_1^T \cdot \lambda \mathbb{I} \cdot \mathbf{a}_1 = \lambda \mathbf{a}_1^T \cdot \mathbf{a}_1 = \lambda$$

Constricción  $\mathbf{a}_1^T \cdot \mathbf{a}_1 = 1$

$$\max(\text{var}(y_1)) = \max(\lambda) = \lambda_1$$

- Entonces, si yo quiero que la varianza de  $y_1$  sea máxima, lo que tengo que buscar es el eigenvalor máximo (en este caso, lo denotamos  $\lambda_1$ ) con su respectivo eigenvector  $\mathbf{a}_1$ .

# Componentes principales

## Cálculo (las otras variables)

- Todo lo que se ha hecho es válido para todas las demás variables, donde se va extrayendo el máximo de los valores propios  $\lambda_i$  restantes,

$$y_1 = \mathbf{a}_1^T \cdot \mathbf{x} \quad \text{--->} \quad \mathbf{a}_1 \text{ es eigenvector de } \Sigma \text{ con eigenvalor } \lambda_1 = \max(\lambda_i)$$

$$y_2 = \mathbf{a}_2^T \cdot \mathbf{x} \quad \text{--->} \quad \mathbf{a}_2 \text{ es eigenvector de } \Sigma \text{ con eigenvalor } \lambda_2 = \max(\lambda_i), \text{ excepto } \lambda_1$$

$\vdots$

$$y_m = \mathbf{a}_m^T \cdot \mathbf{x}$$

# Componentes principales

¿Pero estos  $y_i$  tienen covarianza nula?

- Lo que pedimos es que estas nuevas variables (componentes)  $y_i$  tengan una covarianza nula, i.e.,

$$\text{cov}(y_2, y_1) = \text{cov}(\mathbf{a}_2^T \cdot \mathbf{x}, \mathbf{a}_1^T \cdot \mathbf{x}) = \mathbf{a}_2^T \text{cov}(\mathbf{x}, \mathbf{x}) \mathbf{a}_1 = \mathbf{a}_2^T \cdot \Sigma \cdot \mathbf{a}_1 = 0$$

- Para que la covarianza entre dos variables sea nula, se tiene que cumplir que  $\mathbf{a}_2^T \cdot \Sigma \cdot \mathbf{a}_1 = 0$ .
- Sabemos que  $\Sigma \mathbf{a}_1 = \lambda \mathbf{a}_1$  (ecuación de eigenvectores). Entonces, esta **constricción** lleva a que,

$$\mathbf{a}_2^T (\lambda \mathbf{a}_1) = \lambda \mathbf{a}_2^T \cdot \mathbf{a}_1 = 0$$

- lo que implica que  $\mathbf{a}_2^T \cdot \mathbf{a}_1 = \mathbf{a}_1^T \cdot \mathbf{a}_2 = 0$ .
- Es decir, que estos vectores tienen que ser ortogonales entre sí  $\mathbf{a}_1 \perp \mathbf{a}_2$ .

# Componentes principales

## Cálculo (segundo multiplicador de Lagrange)

- !!! Tenemos otra restricción, tendríamos que agregarlo al Lagrangiano. Para el segundo eigenvector  $\mathbf{a}_2$  tendríamos que el Lagrangiano sería,

$$\mathcal{L}(\mathbf{a}_2) = \mathbf{a}_2^T \cdot \Sigma \cdot \mathbf{a}_2 - \lambda(\mathbf{a}_2^T \cdot \mathbf{a}_2 - 1) - \delta(\mathbf{a}_2^T \cdot \mathbf{a}_1 - 0)$$

- Con el siguiente punto crítico,

$$\frac{\partial \mathcal{L}}{\partial \mathbf{a}_2} = 2\Sigma \cdot \mathbf{a}_2 - 2\lambda\mathbf{a}_2 - \delta\mathbf{a}_1 = 0$$

difícil de despejar, pero hay truco, si multiplicamos  $\mathbf{a}_1^T$  por la izquierda,

$$2\mathbf{a}_1^T \cdot \Sigma \cdot \mathbf{a}_2 - 2\lambda \underbrace{\mathbf{a}_1^T \cdot \mathbf{a}_2}_{=0, \text{ por restricción}} - \delta \underbrace{\mathbf{a}_1^T \cdot \mathbf{a}_1}_{=1, \text{ por restricción}} \Rightarrow \boxed{\delta = 2\mathbf{a}_1^T \cdot \Sigma \cdot \mathbf{a}_2}$$

Segundo multiplicador de Lagrange

# Componentes principales

Cálculo (segundo multiplicador de Lagrange)

$$\delta = 2\mathbf{a}_1^T \cdot \Sigma \cdot \mathbf{a}_2$$

- Nos regresamos a la derivada del Lagrangiano y su punto crítico, dentro del cual podemos sustituir con lo que ya conocemos,

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{a}_2} &= 2\Sigma \cdot \mathbf{a}_2 - 2\lambda \mathbf{a}_2 - \delta \mathbf{a}_1 = 0 \\ &= 2\Sigma \cdot \mathbf{a}_2 - 2\lambda \mathbf{a}_2 - 2\mathbf{a}_1^T \cdot \Sigma \cdot \mathbf{a}_2 \cdot \mathbf{a}_1 = 0 \\ &= 2\Sigma \cdot \mathbf{a}_2 - 2\lambda \mathbf{a}_2 = 0 \quad \text{=0, por restricción} \\ &\Rightarrow (\Sigma - \lambda \mathbb{I}) \cdot \mathbf{a}_2 = 0\end{aligned}$$

- Es decir, **seguimos teniendo la estructura de antes**. Los vectores  $y_i$  que ya calculamos ya de por sí son **ortogonales** entre ellos.

# Componentes principales

## Resumen

- Finalmente, tenemos nuevas variables  $y_i$ . El cálculo involucra las siguientes matrices,

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}; \quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mm} \end{pmatrix}; \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$$

donde  $\mathbf{y}$  es una combinación lineal  $\mathbf{y} = \mathbf{A}\mathbf{x}$ . Esta matriz de coeficientes  $\mathbf{A}$  es lo que se llama una matriz ortogonal, con todos sus elementos ortogonales entre sí  $\mathbf{a}_i^T \cdot \mathbf{a}_i = 1$ ;  $\mathbf{a}_i^T \cdot \mathbf{a}_j = 0$ , y cumple que  $\mathbf{A}^T = \mathbf{A}^{-1}$ .

- La varianza de una de cada una de las variables  $y_i$  está asociada a un valor propio  $\text{var}(y_i) = \lambda_i$ . Se puede representar todos los valores propios en una matriz diagonal,

$$\mathbf{\Delta} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_m \end{pmatrix}$$

- Entonces, toda la varianza de la matriz de variables nuevas  $\mathbf{y}$  puede ser reescrita como  $\mathbf{\Delta} = \text{var}(\mathbf{y}) = \mathbf{A}^T \text{var}(\mathbf{x}) \mathbf{A} = \mathbf{A}^T \mathbf{\Sigma} \mathbf{A}$ , o, finalmente, despejando,

$$\mathbf{\Sigma} = \mathbf{A} \mathbf{\Delta} \mathbf{A}^{-1}$$

# Componentes principales

## Implementación

1. Estandarizar los datos (por cada una de las  $m$  observaciones).
2. Obtener eigenvalores y eigenvectores a partir de la matriz de covarianza  $\Sigma$ .
3. Ordenar los eigenvalores en orden descendente y quedarnos con los que sean mayores, con el objetivos de reducir la dimensionalidad del problema.
4. Construir una matriz de proyección  $\mathbf{W}$  a partir de los eigenvectores.
5. Transformar el dataset original  $\mathbf{x}$  a través de esta matriz de proyección y tener una matriz de nuevas componentes  $\mathbf{y}$ .