UTEC Posgrado

UTEC Posgrado

UTEC Posgrado

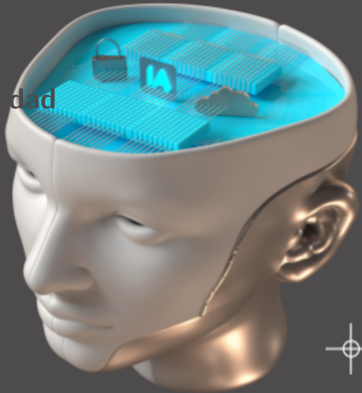
Una escuela para
Early Adopters

MACHINE LEARNING

ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

Motivación

Contexto y necesidad **de la reducción de dimension dad**



El problema de la alta dimensionalidad

En Machine Learning trabajamos con datos de **alta dimensionalidad**:

- Imágenes: 100×100 píxeles $\Rightarrow D = 10{,}000$ dimensiones

El problema de la alta dimensionalidad

En Machine Learning trabajamos con datos de **alta dimensionalidad**:

- Imágenes: 100×100 píxeles $\Rightarrow D = 10{,}000$ dimensiones
- Genómica: $\sim 20{,}000$ genes medidos simultáneamente

El problema de la alta dimensionalidad

En Machine Learning trabajamos con datos de **alta dimensionalidad**:

- Imágenes: 100×100 píxeles $\Rightarrow D = 10{,}000$ dimensiones
- Genómica: $\sim 20{,}000$ genes medidos simultáneamente
- Procesamiento de lenguaje: vocabularios de $10{,}000+$ palabras

El problema de la alta dimensionalidad

En Machine Learning trabajamos con datos de **alta dimensionalidad**:

- Imágenes: 100×100 píxeles $\Rightarrow D = 10{,}000$ dimensiones
- Genómica: $\sim 20{,}000$ genes medidos simultáneamente
- Procesamiento de lenguaje: vocabularios de $10{,}000+$ palabras
- Sensores IoT: cientos de mediciones por segundo

El problema de la alta dimensionalidad

En Machine Learning trabajamos con datos de **alta dimensionalidad**:

- Imágenes: 100×100 píxeles $\Rightarrow D = 10{,}000$ dimensiones
- Genómica: $\sim 20{,}000$ genes medidos simultáneamente
- Procesamiento de lenguaje: vocabularios de $10{,}000+$ palabras
- Sensores IoT: cientos de mediciones por segundo

Problema: La *maldición de la dimensionalidad*

- Distancias pierden significado
- Necesidad exponencial de datos
- Alto costo computacional

¿Qué es PCA?

PCA (Principal Component Analysis) o *Análisis de Componentes Principales*:

- Técnica de **reducción de dimensionalidad**

¿Qué es PCA?

PCA (Principal Component Analysis) o *Análisis de Componentes Principales*:

- Técnica de **reducción de dimensionalidad**
- Encuentra un subespacio de **menor dimensión** $M < D$

¿Qué es PCA?

PCA (Principal Component Analysis) o *Análisis de Componentes Principales*:

- Técnica de **reducción de dimensionalidad**
- Encuentra un subespacio de **menor dimensión** $M < D$
- Proyecta los datos en ese subespacio

¿Qué es PCA?

PCA (Principal Component Analysis) o *Análisis de Componentes Principales*:

- Técnica de **reducción de dimensionalidad**
- Encuentra un subespacio de **menor dimensión** $M < D$
- Proyecta los datos en ese subespacio
- **Maximiza la varianza** de los datos proyectados

¿Qué es PCA?

PCA (Principal Component Analysis) o *Análisis de Componentes Principales*:

- Técnica de **reducción de dimensionalidad**
- Encuentra un subespacio de **menor dimensión** $M < D$
- Proyecta los datos en ese subespacio
- **Maximiza la varianza** de los datos proyectados
- Equivalentemente: **minimiza el error** de reconstrucción

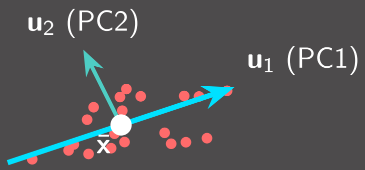
¿Qué es PCA?

PCA (Principal Component Analysis) o *Análisis de Componentes Principales*:

- Técnica de **reducción de dimensionalidad**
- Encuentra un subespacio de **menor dimensión** $M < D$
- Proyecta los datos en ese subespacio
- **Maximiza la varianza** de los datos proyectados
- Equivalentemente: **minimiza el error** de reconstrucción

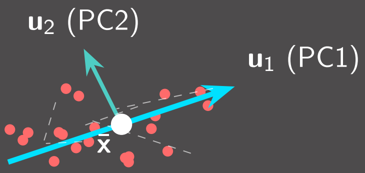
▷ Es una transformación **lineal** y **ortogonal**

Intuición geométrica de PCA



- PC1: dirección de **máxima varianza**

Intuición geométrica de PCA



- PC1: dirección de **máxima varianza**
- Las proyecciones sobre PC1 capturan la mayor información

Aplicaciones de PCA

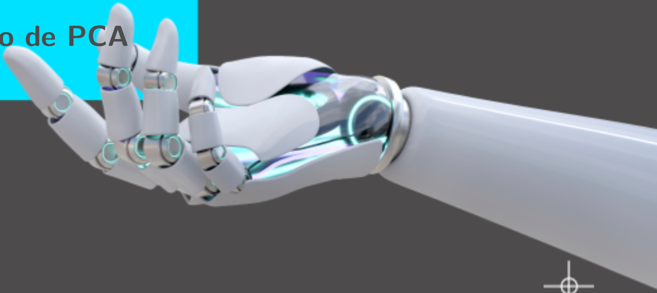
Área	Aplicación
Compresión de datos	Reducir tamaño de imágenes, videos
Visualización	Proyectar datos de D dimensiones a 2D/3D
Preprocesamiento	Eliminar correlaciones, reducir ruido
Extracción de características	Reducir features para clasificación
Reconocimiento facial	Eigenfaces (Turk & Pentland, 1991)
Bioinformática	Análisis de expresión génica

UTEC Posgrado

Una escuela para
Early Adopters

Desarrollo

Conceptos fundamentales **y algoritmo de PCA**



Concepto 1: Media muestral

(1) Concepto:

La **media muestral** $\bar{\mathbf{x}}$ es el centro de masa de los datos. Es el punto promedio en el espacio de características.

(2) Fórmula:

Para un conjunto de datos $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ con $\mathbf{x}_n \in \mathbb{R}^D$:

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n$$

donde N es el número de observaciones y D es la dimensionalidad.

Concepto 1: Media muestral (ejemplo)

(3) Dataset ejemplo:

Consideremos $N = 4$ puntos en $D = 2$ dimensiones:

Observación	x_1	x_2
$\mathbf{x}_1$	1	2
$\mathbf{x}_2$	3	4
$\mathbf{x}_3$	3	6
$\mathbf{x}_4$	5	8

(4) Cálculo paso a paso:

$$\bar{x}_1 = \frac{1}{4}(1 + 3 + 3 + 5) = \frac{12}{4} = 3$$

$$\bar{x}_2 = \frac{1}{4}(2 + 4 + 6 + 8) = \frac{20}{4} = 5$$

Centrado de datos

Antes de aplicar PCA, **centramos** los datos restando la media:

$$\tilde{\mathbf{x}}_n = \mathbf{x}_n - \bar{\mathbf{x}}$$

Para nuestro ejemplo:

n	$\mathbf{x}_n$	$-$	$\bar{\mathbf{x}}$	$= \tilde{\mathbf{x}}_n$
1	$\begin{pmatrix} 1 \\ 2 \end{pmatrix}$	$-$	$\begin{pmatrix} 3 \\ 5 \end{pmatrix}$	$= \begin{pmatrix} -2 \\ -3 \end{pmatrix}$
2	$\begin{pmatrix} 3 \\ 4 \end{pmatrix}$	$-$	$\begin{pmatrix} 3 \\ 5 \end{pmatrix}$	$= \begin{pmatrix} 0 \\ -1 \end{pmatrix}$
3	$\begin{pmatrix} 3 \\ 6 \end{pmatrix}$	$-$	$\begin{pmatrix} 3 \\ 5 \end{pmatrix}$	$= \begin{pmatrix} 0 \\ 1 \end{pmatrix}$
4	$\begin{pmatrix} 5 \\ \text{[illegible]} \end{pmatrix}$	$-$	$\begin{pmatrix} 3 \\ 5 \end{pmatrix}$	$= \begin{pmatrix} 2 \\ \text{[illegible]} \end{pmatrix}$

Concepto 2: Matriz de covarianza

(1) Concepto:

La **matriz de covarianza S** mide cómo varían conjuntamente las diferentes dimensiones de los datos.

(2) Fórmula:

Para datos centrados $\tilde{\mathbf{x}}_n$:

$$\mathbf{S} = \frac{1}{N}\sum_{n=1}^{N}\tilde{\mathbf{x}}_n\tilde{\mathbf{x}}_n^{\mathrm{T}} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^{\mathrm{T}}$$

$\mathbf{S}$ es una matriz simétrica de $D \times D$ donde:

- Diagonal: varianzas de cada dimensión
- Fuera de diagonal: covarianzas entre dimensiones

Concepto 2: Matriz de covarianza (ejemplo)

(3) Dataset: usamos los datos centrados anteriores

(4) Cálculo paso a paso:

$$
\begin{aligned}
\mathbf{S} &= \frac{1}{4}\sum_{n=1}^{4} \tilde{\mathbf{x}}_n \tilde{\mathbf{x}}_n^{\mathrm{T}} \\
&= \frac{1}{4}\left[\begin{pmatrix}-2\\-3\end{pmatrix}\begin{pmatrix}-2 & -3\end{pmatrix} + \begin{pmatrix}0\\-1\end{pmatrix}\begin{pmatrix}0 & -1\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix}\begin{pmatrix}0 & 1\end{pmatrix} + \begin{pmatrix}2\\3\end{pmatrix}\begin{pmatrix}2 & 3\end{pmatrix}\right] \\
&= \frac{1}{4}\left[\begin{pmatrix}4 & 6\\6 & 9\end{pmatrix} + \begin{pmatrix}0 & 0\\0 & 1\end{pmatrix} + \begin{pmatrix}0 & 0\\0 & 1\end{pmatrix} + \begin{pmatrix}4 & 6\\6 & 9\end{pmatrix}\right] \\
&= \frac{1}{4}\begin{pmatrix}8 & 12\\12 & 20\end{pmatrix} = \begin{pmatrix}2 & 3\\3 & 5\end{pmatrix}
\end{aligned}
$$

Concepto 3: Eigenvectores y eigenvalores

(1) Concepto:

Para una matriz $\mathbf{A}$, un **eigenvector v** es un vector que al ser multiplicado por $\mathbf{A}$ solo cambia de escala:

(2) Fórmula:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

donde λ es el **eigenvalor** asociado.

Concepto 3: Eigenvectores y eigenvalores

(1) Concepto:

Para una matriz $\mathbf{A}$, un **eigenvector** $\mathbf{v}$ es un vector que al ser multiplicado por $\mathbf{A}$ solo cambia de escala:

(2) Fórmula:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

donde λ es el **eigenvalor** asociado.

- Para hallar eigenvalores: $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$

Concepto 3: Eigenvectores y eigenvalores

(1) Concepto:

Para una matriz $\mathbf{A}$, un **eigenvector v** es un vector que al ser multiplicado por $\mathbf{A}$ solo cambia de escala:

(2) Fórmula:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

donde λ es el **eigenvalor** asociado.

- Para hallar eigenvalores: $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$
- Para hallar eigenvectores: $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$

Concepto 3: Eigenvectores y eigenvalores

(1) Concepto:

Para una matriz $\mathbf{A}$, un **eigenvector v** es un vector que al ser multiplicado por $\mathbf{A}$ solo cambia de escala:

(2) Fórmula:

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$$

donde λ es el **eigenvalor** asociado.

- Para hallar eigenvalores: $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$
- Para hallar eigenvectores: $(\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0}$
- Los eigenvectores de $\mathbf{S}$ son **ortogonales** (porque $\mathbf{S}$ es simétrica)

Concepto 3: Eigenvectores y eigenvalores

(1) Concepto:

Para una matriz $\mathbf{A}$, un **eigenvector v** es un vector que al ser multiplicado por $\mathbf{A}$ solo cambia de escala:

(2) Fórmula:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

donde λ es el **eigenvalor** asociado.

- Para hallar eigenvalores: $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$
- Para hallar eigenvectores: $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$
- Los eigenvectores de $\mathbf{S}$ son **ortogonales** (porque $\mathbf{S}$ es simétrica)
- Los eigenvalores de $\mathbf{S}$ indican la **varianza** en la dirección del eigenvector

Concepto 3: Eigenvectores (ejemplo paso 1)

(3) Dataset: $\mathbf{S} = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$

(4) Cálculo de eigenvalores:

$$\det(\mathbf{S} - \lambda \mathbf{I}) = 0$$
$$\det \begin{pmatrix} 2-\lambda & 3 \\ 3 & 5-\lambda \end{pmatrix} = 0$$
$$(2-\lambda)(5-\lambda) - 9 = 0$$
$$\lambda^2 - 7\lambda + 1 = 0$$

Fórmula cuadrática:

$$\lambda = \frac{7 \pm \sqrt{49-4}}{2} = \frac{7 \pm \sqrt{45}}{2} = \frac{7 \pm 6{,}708}{2}$$

$\Rightarrow \lambda_1 = 6{,}854 \quad$ y $\quad \lambda_2 = 0{,}146$

Concepto 3: Eigenvectores (ejemplo paso 2)

Para $\lambda_1 = 6{,}854$:

$$(\mathbf{S} - 6{,}854\mathbf{I})\mathbf{v}_1 = \mathbf{0}$$

$$\begin{pmatrix} -4{,}854 & 3 \\ 3 & -1{,}854 \end{pmatrix} \begin{pmatrix} v_{11} \\ v_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

De la primera ecuación: $-4{,}854 v_{11} + 3 v_{12} = 0 \Rightarrow v_{12} = 1{,}618 v_{11}$

Normalizando ($\|\mathbf{v}_1\| = 1$): $\mathbf{v}_1 = \begin{pmatrix} 0{,}526 \\ 0{,}850 \end{pmatrix}$

Para $\lambda_2 = 0{,}146$: $\mathbf{v}_2 = \begin{pmatrix} 0{,}850 \\ -0{,}526 \end{pmatrix}$

Nota: $\mathbf{v}_1^{\mathrm{T}} \mathbf{v}_2 = 0$ (ortogonales)

Concepto 4: Proyección

(1) Concepto:

La **proyección** de un punto $\mathbf{x}$ sobre un vector unitario $\mathbf{u}$ es:

(2) Fórmula:

$$z = \mathbf{u}^{\mathrm{T}}\mathbf{x}$$

Esta es la **coordenada** del punto en la dirección $\mathbf{u}$.

La reconstrucción desde la proyección es:

$$\tilde{\mathbf{x}} = z\mathbf{u} = (\mathbf{u}^{\mathrm{T}}\mathbf{x})\mathbf{u}$$

Para proyectar sobre M componentes principales $\{\mathbf{u}_1, \ldots, \mathbf{u}_M\}$:

$$\mathbf{z} = \mathbf{U}^{\mathrm{T}}\mathbf{x}, \quad \text{donde } \mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_M]$$

Concepto 4: Proyección (ejemplo)

(3) Dataset: Proyectemos $\tilde{\mathbf{x}}_1 = \begin{pmatrix} -2 \\ -3 \end{pmatrix}$ sobre PC1

(4) Cálculo:

Usando $\mathbf{u}_1 = \begin{pmatrix} 0{,}526 \\ 0{,}850 \end{pmatrix}$:

$$\begin{aligned} z_1 &= \mathbf{u}_1^{\mathrm{T}} \tilde{\mathbf{x}}_1 \\ &= \begin{pmatrix} 0{,}526 & 0{,}850 \end{pmatrix} \begin{pmatrix} -2 \\ -3 \end{pmatrix} \\ &= (0{,}526)(-2) + (0{,}850)(-3) \\ &= -1{,}052 - 2{,}550 = -3{,}602 \end{aligned}$$

Si proyectamos todos los puntos sobre PC1:

Formulación de máxima varianza

Objetivo: Encontrar la dirección $\mathbf{u}_1$ que maximiza la varianza de los datos proyectados.

La varianza de los datos proyectados es:

$$\mathrm{Var}(z) = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{u}_1^{\mathrm{T}}\tilde{\mathbf{x}}_n)^2 = \mathbf{u}_1^{\mathrm{T}}\mathbf{S}\mathbf{u}_1$$

Maximizar con restricción $\mathbf{u}_1^{\mathrm{T}}\mathbf{u}_1 = 1$:

$$\max_{\mathbf{u}_1}\left\{\mathbf{u}_1^{\mathrm{T}}\mathbf{S}\mathbf{u}_1 + \lambda(1 - \mathbf{u}_1^{\mathrm{T}}\mathbf{u}_1)\right\}$$

Derivando e igualando a cero:

$$\mathbf{S}\mathbf{u}_1 = \lambda\mathbf{u}_1$$

$\Rightarrow$ $\mathbf{u}_1$ es el eigenvector con el mayor eigenvalor!

Formulación de error mínimo

Perspectiva alternativa: PCA minimiza el error de reconstrucción.

Error cuadrático promedio donde $\tilde{\tilde{\mathbf{x}}}_n$ es la reconstrucción:

$$J = \frac{1}{N} \sum_{n=1}^{N} \|\tilde{\mathbf{x}}_n - \tilde{\tilde{\mathbf{x}}}_n\|^2$$

Se puede demostrar que:

$$J = \sum_{i=M+1}^{D} \lambda_i$$

$\Rightarrow$ Minimizar J requiere descartar los eigenvectores con menores eigenvalores!
Ambas formulaciones conducen al **mismo resultado**.

Algoritmo de PCA

Entrada: Datos $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, dimensión objetivo M

Salida : Componentes principales $\mathbf{U}$, datos proyectados $\mathbf{Z}$

Calcular la media: $\bar{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_n$

Centrar los datos: $\tilde{\mathbf{x}}_n = \mathbf{x}_n - \bar{\mathbf{x}}$ para $n = 1, \ldots, N$

Calcular la matriz de covarianza: $\mathbf{S} = \frac{1}{N}\sum_{n=1}^{N} \tilde{\mathbf{x}}_n \tilde{\mathbf{x}}_n^{\mathrm{T}}$

Calcular eigenvalores $\{\lambda_1, \ldots, \lambda_D\}$ y eigenvectores $\{\mathbf{v}_1, \ldots, \mathbf{v}_D\}$ de $\mathbf{S}$

Ordenar eigenvalores en orden decreciente: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_D$

Seleccionar los M eigenvectores correspondientes a los M mayores eigenvalores

Formar matriz $\mathbf{U} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_M]$

Proyectar datos: $\mathbf{z}_n = \mathbf{U}^{\mathrm{T}} \tilde{\mathbf{x}}_n$ para $n = 1, \ldots, N$

return $\mathbf{U}$, $\{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$

Ejemplo completo: Resumen

Retomemos nuestro ejemplo con $N = 4$ puntos en $D = 2$ dimensiones.

Paso 1: Media

$$\bar{\mathbf{x}} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$$

Paso 2: Datos centrados

$$\tilde{\mathbf{X}} = \begin{pmatrix} -2 & 0 & 0 & 2 \\ -3 & -1 & 1 & 3 \end{pmatrix}^{\mathrm{T}}$$

Paso 3: Covarianza

$$\mathbf{S} = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$$

Ejemplo completo: Componentes principales

Paso 4-6: Eigenvalores y eigenvectores

$$\lambda_1 = 6{,}854, \quad \mathbf{u}_1 = \begin{pmatrix} 0{,}526 \\ 0{,}850 \end{pmatrix}$$

$$\lambda_2 = 0{,}146, \quad \mathbf{u}_2 = \begin{pmatrix} 0{,}850 \\ -0{,}526 \end{pmatrix}$$

Interpretación:

- PC1 captura $\frac{6{,}854}{6{,}854+0{,}146} = 97{,}9\,\%$ de la varianza
- PC2 captura solo $2{,}1\,\%$ de la varianza

Si reducimos a $M = 1$: proyectamos solo sobre $\mathbf{u}_1$, reteniendo $97{,}9\,\%$ de la información!

Ejemplo completo: Proyección y reconstrucción

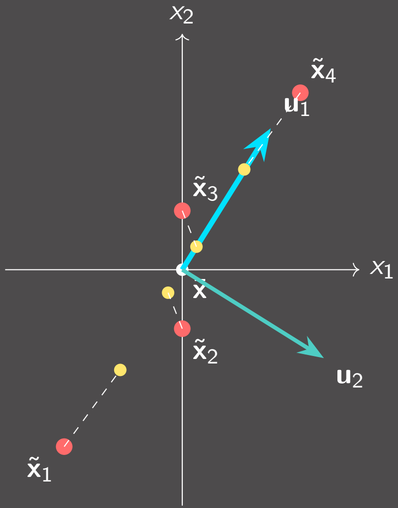
Para reducir a $M = 1$, proyectamos sobre $\mathbf{u}_1$:

n	$\tilde{\mathbf{x}}_n$	$z_n = \mathbf{u}_1^{\mathrm{T}} \tilde{\mathbf{x}}_n$	Reconstrucción $z_n \mathbf{u}_1$
1	$(-2, -3)^{\mathrm{T}}$	$-3{,}602$	$(-1{,}89, -3{,}06)^{\mathrm{T}}$
2	$(0, -1)^{\mathrm{T}}$	$-0{,}850$	$(-0{,}45, -0{,}72)^{\mathrm{T}}$
3	$(0, 1)^{\mathrm{T}}$	$0{,}850$	$(0{,}45, 0{,}72)^{\mathrm{T}}$
4	$(2, 3)^{\mathrm{T}}$	$3{,}602$	$(1{,}89, 3{,}06)^{\mathrm{T}}$

Dimensionalidad reducida:

- Original: 4 puntos $\times$ 2 dimensiones $=$ 8 números
- Comprimido: 4 puntos $\times$ 1 dimensión $+$ $\mathbf{u}_1$ (2 valores) $+$ $\bar{\mathbf{x}}$ (2 valores) $=$ 8 números
- Para N grande: ahorro significativo!

Visualización del ejemplo



Varianza explicada

¿Cómo elegir M?

La **proporción de varianza explicada** por los primeros M componentes:

$$\mathrm{PVE}(M) = \frac{\sum_{i=1}^{M} \lambda_i}{\sum_{i=1}^{D} \lambda_i}$$

Criterios comunes:

- Retener componentes hasta alcanzar 90 %, 95 % o 99 % de varianza
- Regla del codo: donde la curva de varianza acumulada se "aplana"
- Eigenvalores > 1 (regla de Kaiser)

En nuestro ejemplo:

- $M = 1$: PVE $= 97{,}9\,\%$ ✓ Excelente!
- $M = 2$: PVE $= 100\,\%$ (sin reducción)

Propiedades importantes de PCA

1. **Decorrelación:** Los componentes principales están **no correlacionados**

$$\mathrm{Cov}(z_i, z_j) = 0 \quad \text{para } i \neq j$$

Propiedades importantes de PCA

1. **Decorrelación:** Los componentes principales están **no correlacionados**

$$\mathrm{Cov}(z_i, z_j) = 0 \quad \text{para } i \neq j$$

2. **Varianza ordenada:**

$$\mathrm{Var}(z_1) \geq \mathrm{Var}(z_2) \geq \cdots \geq \mathrm{Var}(z_D)$$

Propiedades importantes de PCA

1. **Decorrelación:** Los componentes principales están **no correlacionados**

$$\mathrm{Cov}(z_i, z_j) = 0 \quad \text{para } i \neq j$$

2. **Varianza ordenada:**

$$\mathrm{Var}(z_1) \geq \mathrm{Var}(z_2) \geq \cdots \geq \mathrm{Var}(z_D)$$

3. **Ortogonalidad:** Los eigenvectores son ortogonales

$$\mathbf{u}_i^{\mathrm{T}} \mathbf{u}_j = \delta_{ij}$$

Propiedades importantes de PCA

1. **Decorrelación:** Los componentes principales están **no correlacionados**

$$\mathrm{Cov}(z_i, z_j) = 0 \quad \text{para } i \neq j$$

2. **Varianza ordenada:**

$$\mathrm{Var}(z_1) \geq \mathrm{Var}(z_2) \geq \cdots \geq \mathrm{Var}(z_D)$$

3. **Ortogonalidad:** Los eigenvectores son ortogonales

$$\mathbf{u}_i^{\mathrm{T}} \mathbf{u}_j = \delta_{ij}$$

4. **Optimalidad:** PCA es óptimo bajo ambas formulaciones (máxima varianza y mínimo error)

Propiedades importantes de PCA

1. **Decorrelación:** Los componentes principales están **no correlacionados**

$$\text{Cov}(z_i, z_j) = 0 \quad \text{para } i \neq j$$

2. **Varianza ordenada:**

$$\text{Var}(z_1) \geq \text{Var}(z_2) \geq \cdots \geq \text{Var}(z_D)$$

3. **Ortogonalidad:** Los eigenvectores son ortogonales

$$\mathbf{u}_i^{\mathrm{T}} \mathbf{u}_j = \delta_{ij}$$

4. **Optimalidad:** PCA es óptimo bajo ambas formulaciones (máxima varianza y mínimo error)
5. **Unicidad:** Los eigenvectores son únicos (salvo el signo) si los eigenvalores son distintos

Limitaciones de PCA

1. **Linealidad:** PCA solo encuentra transformaciones **lineales**
   - No captura relaciones no lineales en los datos
   - Solución: Kernel PCA

Limitaciones de PCA

1. **Linealidad:** PCA solo encuentra transformaciones **lineales**
   - No captura relaciones no lineales en los datos
   - Solución: Kernel PCA
2. **Sensibilidad a escala:** Dimensiones con mayor varianza dominan
   - Solución: Estandarizar datos antes de PCA

Limitaciones de PCA

1. **Linealidad:** PCA solo encuentra transformaciones **lineales**
   - No captura relaciones no lineales en los datos
   - Solución: Kernel PCA
2. **Sensibilidad a escala:** Dimensiones con mayor varianza dominan
   - Solución: Estandarizar datos antes de PCA
3. **Interpretabilidad:** Los componentes son combinaciones de todas las variables
   - Pueden ser difíciles de interpretar

Limitaciones de PCA

1. **Linealidad:** PCA solo encuentra transformaciones **lineales**
   - No captura relaciones no lineales en los datos
   - Solución: Kernel PCA
2. **Sensibilidad a escala:** Dimensiones con mayor varianza dominan
   - Solución: Estandarizar datos antes de PCA
3. **Interpretabilidad:** Los componentes son combinaciones de todas las variables
   - Pueden ser difíciles de interpretar
4. **Outliers:** PCA es sensible a valores atípicos
   - Solución: Robust PCA

Limitaciones de PCA

1. **Linealidad:** PCA solo encuentra transformaciones **lineales**
   - No captura relaciones no lineales en los datos
   - Solución: Kernel PCA
2. **Sensibilidad a escala:** Dimensiones con mayor varianza dominan
   - Solución: Estandarizar datos antes de PCA
3. **Interpretabilidad:** Los componentes son combinaciones de todas las variables
   - Pueden ser difíciles de interpretar
4. **Outliers:** PCA es sensible a valores atípicos
   - Solución: Robust PCA
5. **No supervisado:** No usa información de etiquetas
   - Puede no ser óptimo para clasificación
   - Alternativa: Linear Discriminant Analysis (LDA)

Conclusiones

Resumen y reflexiones **del Análisis de Componentes Principales**

Resumen: PCA paso a paso

1. **Calcular la media $\bar{\mathbf{x}}$** de los datos

Resumen: PCA paso a paso

1. **Calcular la media** $\bar{\mathbf{x}}$ de los datos
2. **Centrar** los datos: $\tilde{\mathbf{x}}_n = \mathbf{x}_n - \bar{\mathbf{x}}$

Resumen: PCA paso a paso

1. **Calcular la media** $\bar{\mathbf{x}}$ de los datos
2. **Centrar** los datos: $\tilde{\mathbf{x}}_n = \mathbf{x}_n - \bar{\mathbf{x}}$
3. **Calcular la matriz de covarianza** $\mathbf{S}$

Resumen: PCA paso a paso

1. **Calcular la media** $\bar{\mathbf{x}}$ de los datos
2. **Centrar** los datos: $\tilde{\mathbf{x}}_n = \mathbf{x}_n - \bar{\mathbf{x}}$
3. **Calcular la matriz de covarianza** $\mathbf{S}$
4. **Calcular eigenvectores y eigenvalores** de $\mathbf{S}$

Resumen: PCA paso a paso

1. **Calcular la media** $\bar{\mathbf{x}}$ de los datos
2. **Centrar** los datos: $\tilde{\mathbf{x}}_n = \mathbf{x}_n - \bar{\mathbf{x}}$
3. **Calcular la matriz de covarianza** $\mathbf{S}$
4. **Calcular eigenvectores y eigenvalores** de $\mathbf{S}$
5. **Ordenar** eigenvalores en orden decreciente

Resumen: PCA paso a paso

1. **Calcular la media** $\bar{\mathbf{x}}$ de los datos
2. **Centrar** los datos: $\tilde{\mathbf{x}}_n = \mathbf{x}_n - \bar{\mathbf{x}}$
3. **Calcular la matriz de covarianza** $\mathbf{S}$
4. **Calcular eigenvectores y eigenvalores** de $\mathbf{S}$
5. **Ordenar** eigenvalores en orden decreciente
6. **Seleccionar** los M eigenvectores principales

Resumen: PCA paso a paso

1. **Calcular la media** $\bar{\mathbf{x}}$ de los datos
2. **Centrar** los datos: $\tilde{\mathbf{x}}_n = \mathbf{x}_n - \bar{\mathbf{x}}$
3. **Calcular la matriz de covarianza** $\mathbf{S}$
4. **Calcular eigenvectores y eigenvalores** de $\mathbf{S}$
5. **Ordenar** eigenvalores en orden decreciente
6. **Seleccionar** los M eigenvectores principales
7. **Proyectar** los datos: $\mathbf{z}_n = \mathbf{U}^{\mathrm{T}}\tilde{\mathbf{x}}_n$

Resumen: PCA paso a paso

1. **Calcular la media** $\bar{\mathbf{x}}$ de los datos
2. **Centrar** los datos: $\tilde{\mathbf{x}}_n = \mathbf{x}_n - \bar{\mathbf{x}}$
3. **Calcular la matriz de covarianza** $\mathbf{S}$
4. **Calcular eigenvectores y eigenvalores** de $\mathbf{S}$
5. **Ordenar** eigenvalores en orden decreciente
6. **Seleccionar** los M eigenvectores principales
7. **Proyectar** los datos: $\mathbf{z}_n = \mathbf{U}^{\mathrm{T}}\tilde{\mathbf{x}}_n$

Resultado: Datos de dimensión D reducidos a dimensión M, reteniendo la máxima varianza posible.

Conceptos clave aprendidos

- **Media muestral**: Centro de los datos

Conceptos clave aprendidos

- **Media muestral**: Centro de los datos
- **Matriz de covarianza**: Cómo varían juntas las dimensiones

Conceptos clave aprendidos

- **Media muestral**: Centro de los datos
- **Matriz de covarianza**: Cómo varían juntas las dimensiones
- **Eigenvectores**: Direcciones principales de varianza

Conceptos clave aprendidos

- **Media muestral**: Centro de los datos
- **Matriz de covarianza**: Cómo varían juntas las dimensiones
- **Eigenvectores**: Direcciones principales de varianza
- **Eigenvalores**: Magnitud de varianza en cada dirección

Conceptos clave aprendidos

- **Media muestral**: Centro de los datos
- **Matriz de covarianza**: Cómo varían juntas las dimensiones
- **Eigenvectores**: Direcciones principales de varianza
- **Eigenvalores**: Magnitud de varianza en cada dirección
- **Proyección**: Transformación a menor dimensión

Conceptos clave aprendidos

- **Media muestral**: Centro de los datos
- **Matriz de covarianza**: Cómo varían juntas las dimensiones
- **Eigenvectores**: Direcciones principales de varianza
- **Eigenvalores**: Magnitud de varianza en cada dirección
- **Proyección**: Transformación a menor dimensión
- **Varianza explicada**: Qué tanto se retiene con M componentes

Conceptos clave aprendidos

- **Media muestral**: Centro de los datos
- **Matriz de covarianza**: Cómo varían juntas las dimensiones
- **Eigenvectores**: Direcciones principales de varianza
- **Eigenvalores**: Magnitud de varianza en cada dirección
- **Proyección**: Transformación a menor dimensión
- **Varianza explicada**: Qué tanto se retiene con M componentes

PCA unifica dos perspectivas:

1. Maximizar varianza proyectada
2. Minimizar error de reconstrucción

¿Cuándo usar PCA?

Escenario	Decisión
Features correlacionadas	✓ Usar PCA
Visualización ($D \gg 3$)	✓ Usar PCA
Compresión de datos	✓ Usar PCA
Preprocesamiento ML	✓ Considerar PCA
Relaciones no lineales	× Kernel PCA / t-SNE
Problema supervisado	× Considerar LDA
Interpretabilidad crítica	× Evitar PCA

Implementación práctica

PCA está disponible en todas las bibliotecas populares de ML:

Python (scikit-learn):

```
from sklearn.decomposition import PCA
pca = PCA(n_components=2)
X_reduced = pca.fit_transform(X)
```

R:

```
pca_result <- prcomp(X, scale = TRUE)
X_reduced <- pca_result$x[, 1:2]
```

MATLAB:

```
[coeff, score, latent] = pca(X);
X_reduced = score(:, 1:2);
```

UTEC Posgrado

UTEC Posgrado