



UTECH Posgrado



UTEC Posgrado



MACHINE LEARNING

PRÁCTICA BASE: PCA

Análisis de Componentes Principales

Datos del Alumno

Nombres: _____

Apellidos: _____

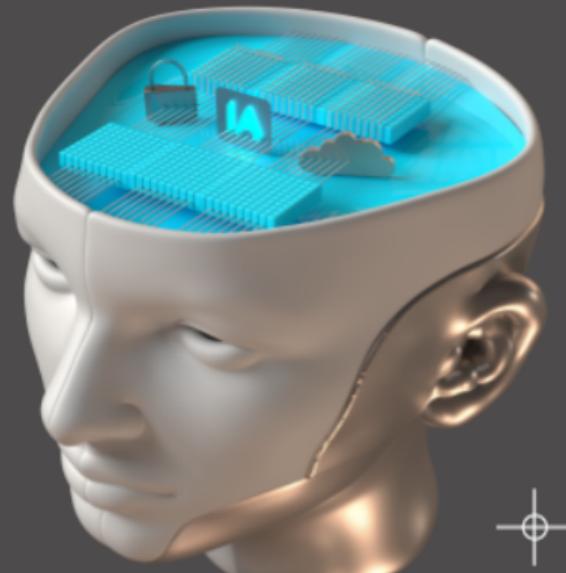
Código: _____

Fecha: _____



Algoritmo PCA

Procedimiento general para **reducir dimensionalidad**



Objetivo: Reducir la dimensionalidad de un conjunto de datos de D dimensiones a M dimensiones ($M < D$), preservando la máxima varianza.

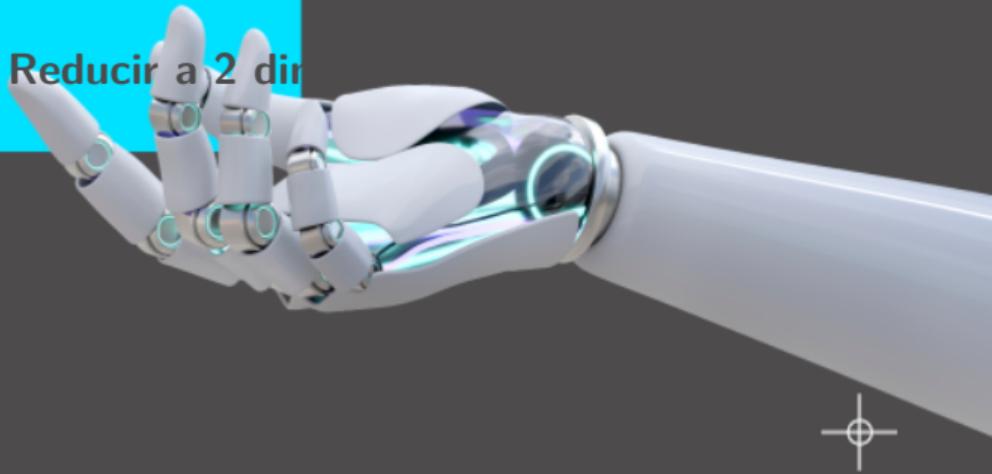
Pasos del algoritmo:

1. **Calcular la media** de cada variable del dataset
2. **Centrar los datos** restando la media
3. **Calcular la matriz de covarianza**
4. **Calcular eigenvalores y eigenvectores** de la matriz de covarianza
5. **Ordenar eigenvalores** de mayor a menor
6. **Seleccionar los M eigenvectores** con mayores eigenvalores
7. **Proyectar los datos** en el nuevo espacio M -dimensional

Referencia: Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*, Cap. 12. Springer.

Ejercicio 1

Dataset: 10 registros, 3 dimensiones **Reducir a 2 dim.**

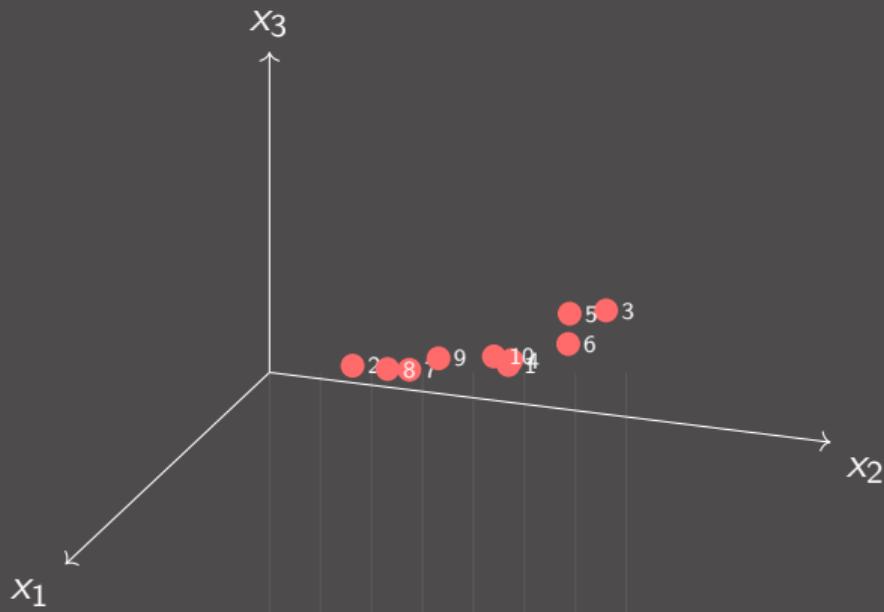


Ejercicio 1: Dataset original

Tenemos 10 observaciones en 3 dimensiones ($D = 3$). Objetivo: reducir a $M = 2$.

ID	x_1	x_2	x_3
1	2.5	2.4	1.2
2	0.5	0.7	0.3
3	2.2	2.9	1.5
4	1.9	2.2	1.0
5	3.1	3.0	1.8
6	2.3	2.7	1.3
7	2.0	1.6	0.9
8	1.0	1.1	0.5
9	1.5	1.6	0.8
10	1.1	1.8	0.7

Visualización 3D: Datos originales



Los 10 puntos en el espacio 3D original.

Paso 1: Calcular la media

Notación matricial:

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

Notación escalar (por variable):

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ij}, \quad j = 1, 2, 3$$

Cálculos:

$$\bar{x}_1 = \frac{1}{10}(2,5 + 0,5 + 2,2 + 1,9 + 3,1 + 2,3 + 2,0 + 1,0 + 1,5 + 1,1) = \frac{18,1}{10} = 1,81$$

$$\bar{x}_2 = \frac{1}{10}(2,4 + 0,7 + 2,9 + 2,2 + 3,0 + 2,7 + 1,6 + 1,1 + 1,6 + 1,8) = \frac{20,0}{10} = 2,00$$

$$\bar{x}_3 = \frac{1}{10}(1,2 + 0,3 + 1,5 + 1,0 + 1,8 + 1,3 + 0,9 + 0,5 + 0,8 + 0,7) = \frac{10,0}{10} = 1,00$$

Paso 2a: Centrar los datos - Fórmulas

Notación matricial:

$$\tilde{\mathbf{X}} = \mathbf{X} - \bar{\mathbf{x}}\mathbf{1}^T$$

Notación escalar (elemento a elemento):

$$\tilde{x}_{ij} = x_{ij} - \bar{x}_j$$

Procedimiento:

- ▶ A cada valor original x_{ij} le restamos la media \bar{x}_j de su variable
- ▶ Los datos centrados tienen media cero
- ▶ Esto facilita el cálculo de la covarianza



Paso 2b: Centrar los datos - Resultados

ID	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3
1	0.69	0.40	0.20
2	-1.31	-1.30	-0.70
3	0.39	0.90	0.50
4	0.09	0.20	0.00
5	1.29	1.00	0.80
6	0.49	0.70	0.30
7	0.19	-0.40	-0.10
8	-0.81	-0.90	-0.50
9	-0.31	-0.40	-0.20
10	-0.71	-0.20	-0.30

Estos son los datos centrados que utilizaremos para calcular la matriz de covarianza.



Paso 3a: Cómo se calcula la covarianza

La **covarianza** entre dos variables x_j y x_k mide cómo varían juntas.

Fórmula:

$$\text{Cov}(x_j, x_k) = \frac{1}{N} \sum_{i=1}^N (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) = \frac{1}{N} \sum_{i=1}^N \tilde{x}_{ij}\tilde{x}_{ik}$$

Ejemplo: $\text{Cov}(x_1, x_2)$:

$$\begin{aligned}\text{Cov}(x_1, x_2) &= \frac{1}{10} \sum_{i=1}^{10} \tilde{x}_{i1}\tilde{x}_{i2} \\ &= \frac{1}{10} [(0,69)(0,40) + (-1,31)(-1,30) + (0,39)(0,90) + \dots] \\ &= \frac{1}{10} [0,276 + 1,703 + 0,351 + 0,018 + 1,290 + 0,343 - 0,076 + 0,729 + 0,124 + \dots] \\ &= \frac{6,15}{10} = 0,615\end{aligned}$$



Paso 3b: Cómo se calcula la matriz de covarianza

La **matriz de covarianza S** es una matriz $D \times D$ que contiene todas las covarianzas.

Notación matricial:

$$\mathbf{S} = \frac{1}{N} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$$

Notación escalar (elemento a elemento):

$$s_{jk} = \text{Cov}(x_j, x_k) = \frac{1}{N} \sum_{i=1}^N \tilde{x}_{ij} \tilde{x}_{ik}$$

Resultado para nuestro ejemplo:

$$\mathbf{S} = \begin{pmatrix} 0,616 & 0,615 & 0,410 \\ 0,615 & 0,717 & 0,516 \\ 0,410 & 0,516 & 0,290 \end{pmatrix}$$

donde $s_{11} = \text{Var}(x_1)$, $s_{12} = \text{Cov}(x_1, x_2)$, etc.

Conceptos clave: Eigenvalores y Eigenvectores

¿Qué son?

Un **eigenvector** (vector propio) de una matriz \mathbf{S} es un vector \mathbf{v} que al ser multiplicado por \mathbf{S} solo cambia en magnitud, no en dirección:

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

El **eigenvalor** (valor propio) λ es el factor de escalamiento.

Interpretación geométrica:

- ▶ El eigenvector \mathbf{v} indica una **dirección**
- ▶ El eigenvalor λ indica cuánta **varianza** hay en esa dirección
- ▶ En PCA, eigenvectores con mayor λ capturan más información



Para qué sirven en PCA?

Utilidad de eigenvalores y eigenvectores:

- ▶ Los **eigenvectores** de la matriz de covarianza definen las **nuevas direcciones** (componentes principales)
- ▶ Estas direcciones son **ortogonales** entre sí (perpendiculares)
- ▶ Los **eigenvalores** indican cuánta varianza captura cada componente
- ▶ Ordenamos por eigenvalor decreciente para priorizar componentes importantes
- ▶ Podemos descartar componentes con eigenvalores pequeños sin perder mucha información

Ejemplo: Si $\lambda_1 = 1,530$ y $\lambda_2 = 0,085$, la primera componente es ≈ 18 veces más importante que la segunda.



Herramienta web para cálculos

Para verificar cálculos de eigenvalores y eigenvectores:

Calculadora online recomendada:

<https://www.symbolab.com/solver/matrix-eigenvalues-calculator>

Alternativas:

- ▶ WolframAlpha: www.wolframalpha.com
- ▶ Matrix Calculator: matrixcalc.org/en/

Nota: Ingrese la matriz de covarianza y la herramienta calculará automáticamente eigenvalores y eigenvectores normalizados.



Concepto: ¿Qué es el determinante?

El **determinante** de una matriz \mathbf{A} es un número que resume propiedades importantes de la matriz.

Interpretación geométrica:

- ▶ Representa el **factor de escala** del volumen cuando se aplica la transformación lineal
- ▶ Si $\det(\mathbf{A}) = 0$, la matriz es **singular** (no invertible)
- ▶ Si $\det(\mathbf{A}) \neq 0$, la matriz es **invertible**

¿Para qué sirve en PCA?

- ▶ Necesitamos resolver $\det(\mathbf{S} - \lambda\mathbf{I}) = 0$
- ▶ Esta ecuación encuentra los valores λ que hacen singular a $(\mathbf{S} - \lambda\mathbf{I})$
- ▶ Esos valores son precisamente los eigenvalores



Ejemplo: Cálculo de determinante 2x2

¿Cómo se calcula?

Para una matriz 2×2 :

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

Ejemplo numérico:

$$\det \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix} = (3)(4) - (2)(1) = 12 - 2 = 10$$

Como $\det = 10 \neq 0$, la matriz es invertible.



Ejemplo: Cálculo de determinante 3x3

Para una matriz 3×3 , usamos **expansión por cofactores** (primera fila):

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = a \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \det \begin{pmatrix} d & e \\ g & h \end{pmatrix}$$

Ejemplo numérico:

$$\begin{aligned}\det \begin{pmatrix} 2 & 1 & 3 \\ 0 & 4 & 1 \\ 1 & 2 & 5 \end{pmatrix} &= 2 \det \begin{pmatrix} 4 & 1 \\ 2 & 5 \end{pmatrix} - 1 \det \begin{pmatrix} 0 & 1 \\ 1 & 5 \end{pmatrix} + 3 \det \begin{pmatrix} 0 & 4 \\ 1 & 2 \end{pmatrix} \\ &= 2[(4)(5) - (1)(2)] - 1[(0)(5) - (1)(1)] + 3[(0)(2) - (4)(1)] \\ &= 2[20 - 2] - 1[0 - 1] + 3[0 - 4] \\ &= 2(18) - 1(-1) + 3(-4) \\ &= 36 + 1 - 12 = 25\end{aligned}$$



£Por qué un polinomio cúbico?

La **dimensión de la matriz** determina el grado del polinomio característico.

Regla general:

- ▶ Matriz $n \times n \Rightarrow$ polinomio de grado n
- ▶ Matriz $2 \times 2 \Rightarrow$ polinomio **cuadrático** (grado 2)
- ▶ Matriz $3 \times 3 \Rightarrow$ polinomio **cúbico** (grado 3)
- ▶ Matriz $4 \times 4 \Rightarrow$ polinomio **cuártico** (grado 4)

En nuestro caso:

- ▶ Tenemos matriz de covarianza **S** de 3×3 (3 variables)
- ▶ Por lo tanto, obtendremos un polinomio cúbico
- ▶ Este polinomio tendrá exactamente 3 raíces (los 3 eigenvalores)



Criterio del polinomio característico

¿Por qué usamos este método?

El polinomio característico $\det(\mathbf{S} - \lambda\mathbf{I}) = 0$ es el método estándar porque:

Fundamento matemático:

- ▶ Si $\mathbf{S}\mathbf{v} = \lambda\mathbf{v}$, entonces $(\mathbf{S} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$
- ▶ Para que exista $\mathbf{v} \neq \mathbf{0}$, la matriz $(\mathbf{S} - \lambda\mathbf{I})$ debe ser singular
- ▶ Una matriz es singular si y solo si su determinante es cero

Garantías:

- ▶ El polinomio de grado n tiene exactamente n raíces (contando multiplicidad)
- ▶ Por tanto, una matriz $n \times n$ tiene exactamente n eigenvalores



Paso 4a: Ecuación característica

Los **eigenvalores** λ de \mathbf{S} se obtienen resolviendo la ecuación característica:

$$\det(\mathbf{S} - \lambda\mathbf{I}) = 0$$

donde \mathbf{I} es la matriz identidad.

Paso a paso:

1. Restar λ de los elementos de la diagonal de \mathbf{S}
2. Calcular el determinante de la matriz resultante
3. Igualar a cero y resolver para λ



Paso 4b: Formando la matriz $\mathbf{S} - \lambda\mathbf{I}$

Para nuestra matriz de covarianza:

$$\mathbf{S} - \lambda\mathbf{I} = \begin{pmatrix} 0,616 & 0,615 & 0,410 \\ 0,615 & 0,717 & 0,516 \\ 0,410 & 0,516 & 0,290 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0,616 - \lambda & 0,615 & 0,410 \\ 0,615 & 0,717 - \lambda & 0,516 \\ 0,410 & 0,516 & 0,290 - \lambda \end{pmatrix}$$



Paso 4c: Calcular el determinante

Para una matriz 3×3 , el determinante se calcula como:

$$\begin{aligned}\det(\mathbf{S} - \lambda\mathbf{I}) &= \\ (0,616 - \lambda)[(0,717 - \lambda)(0,290 - \lambda) - (0,516)(0,516)] &\\ - (0,615)[(0,615)(0,290 - \lambda) - (0,516)(0,410)] &\\ + (0,410)[(0,615)(0,516) - (0,717 - \lambda)(0,410)] &\end{aligned}$$

Expandiendo y simplificando, obtenemos un polinomio cúbico:

$$-\lambda^3 + 1,623\lambda^2 - 0,268\lambda + 0,010 = 0$$



Paso 4d: Resolver el polinomio

Resolver el polinomio cúbico:

$$-\lambda^3 + 1,623\lambda^2 - 0,268\lambda + 0,010 = 0$$

Usando métodos numéricos (Newton-Raphson, calculadora online, etc.):

Eigenvalores obtenidos:

$$\lambda_1 = 1,530 \quad (\text{mayor} - 94.4 \% \text{ varianza})$$

$$\lambda_2 = 0,085 \quad (\text{medio} - 5.2 \% \text{ varianza})$$

$$\lambda_3 = 0,008 \quad (\text{menor} - 0.4 \% \text{ varianza})$$

Verificación: $\lambda_1 + \lambda_2 + \lambda_3 = 1,623 = \text{traza}(\mathbf{S})$



Paso 5a: Sistema de ecuaciones para eigenvector

Para cada eigenvalor λ_i , el eigenvector \mathbf{v}_i satisface:

$$(\mathbf{S} - \lambda_i \mathbf{I})\mathbf{v}_i = \mathbf{0}$$

Para $\lambda_1 = 1,530$, formamos el sistema:

$$\begin{pmatrix} 0,616 - 1,530 & 0,615 & 0,410 \\ 0,615 & 0,717 - 1,530 & 0,516 \\ 0,410 & 0,516 & 0,290 - 1,530 \end{pmatrix} \begin{pmatrix} v_{11} \\ v_{12} \\ v_{13} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} -0,914 & 0,615 & 0,410 \\ 0,615 & -0,813 & 0,516 \\ 0,410 & 0,516 & -1,240 \end{pmatrix} \begin{pmatrix} v_{11} \\ v_{12} \\ v_{13} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$



Paso 5b: Expandir el sistema

El sistema de ecuaciones es:

$$-0,914v_{11} + 0,615v_{12} + 0,410v_{13} = 0 \quad (\text{ec. 1})$$

$$0,615v_{11} - 0,813v_{12} + 0,516v_{13} = 0 \quad (\text{ec. 2})$$

$$0,410v_{11} + 0,516v_{12} - 1,240v_{13} = 0 \quad (\text{ec. 3})$$

Este sistema tiene infinitas soluciones (cualquier múltiplo de v_1).

Estrategia: Asignar $v_{13} = 1$ y resolver para v_{11} y v_{12} .



Paso 5c: Resolver el sistema

Asignando $v_{13} = 1$ y usando las dos primeras ecuaciones:

$$-0,914v_{11} + 0,615v_{12} = -0,410$$

$$0,615v_{11} - 0,813v_{12} = -0,516$$

Multiplicando ec. 1 por 0,615 y ec. 2 por 0,914:

$$-0,562v_{11} + 0,378v_{12} = -0,252$$

$$0,562v_{11} - 0,743v_{12} = -0,472$$

Sumando ambas: $-0,365v_{12} = -0,724 \Rightarrow v_{12} = 1,253$

Sustituyendo: $v_{11} = 1,077$

Vector sin normalizar: $(1,077, 1,253, 1,000)^T$



Paso 5d: Normalizar el eigenvector

Para normalizar, dividimos por la norma:

$$\|\mathbf{v}\| = \sqrt{v_{11}^2 + v_{12}^2 + v_{13}^2}$$

$$\|\mathbf{v}\| = \sqrt{(1,077)^2 + (1,253)^2 + (1,000)^2} = \sqrt{1,160 + 1,570 + 1,000} = \sqrt{3,730} = 1,931$$

Eigenvector normalizado:

$$\mathbf{v}_1 = \frac{1}{1,931} \begin{pmatrix} 1,077 \\ 1,253 \\ 1,000 \end{pmatrix} = \begin{pmatrix} 0,558 \\ 0,649 \\ 0,518 \end{pmatrix}$$

$$\text{Verificación: } \|\mathbf{v}_1\| = \sqrt{0,558^2 + 0,649^2 + 0,518^2} = 1,000$$

Paso 5e: Eigenvectores completos

Repetiendo el proceso para λ_2 y λ_3 :

Eigenvectores normalizados:

$$\mathbf{v}_1 = \begin{pmatrix} 0,558 \\ 0,649 \\ 0,518 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} -0,735 \\ 0,052 \\ 0,676 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 0,384 \\ -0,759 \\ 0,525 \end{pmatrix}$$

Propiedades importantes:

- ▶ Todos tienen norma = 1
- ▶ Son ortogonales entre sí: $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, $\mathbf{v}_1 \cdot \mathbf{v}_3 = 0$, $\mathbf{v}_2 \cdot \mathbf{v}_3 = 0$
- ▶ Forman una base ortonormal del espacio 3D



Paso 6: Varianza explicada

PC	Eigenvalor	Var. Explicada	Acumulada
PC1	1.530	94.4 %	94.4 %
PC2	0.085	5.2 %	99.6 %
PC3	0.008	0.4 %	100.0 %

Con 2 componentes retenemos **99.6 %** de la varianza!

Interpretación:

- ▶ PC1 captura la mayor parte de la variabilidad (94.4 %)
- ▶ PC2 captura solo 5.2 % adicional
- ▶ PC3 es casi despreciable (0.4 %)



Paso 7a: Proyección a 2D - Fórmulas

Notación matricial:

$$\mathbf{Z} = \tilde{\mathbf{X}}\mathbf{V}_{[1:2]}$$

donde $\mathbf{V}_{[1:2]} = [\mathbf{v}_1, \mathbf{v}_2]$ es la matriz 3×2 de los primeros 2 eigenvectores.

Notación escalar (para cada punto):

$$z_{i1} = \tilde{x}_{i1}v_{11} + \tilde{x}_{i2}v_{12} + \tilde{x}_{i3}v_{13}$$

$$z_{i2} = \tilde{x}_{i1}v_{21} + \tilde{x}_{i2}v_{22} + \tilde{x}_{i3}v_{23}$$

Interpretación:

- ▶ Cada punto 3D se proyecta a 2D usando los eigenvectores como base
- ▶ z_{i1} es la coordenada en PC1, z_{i2} en PC2

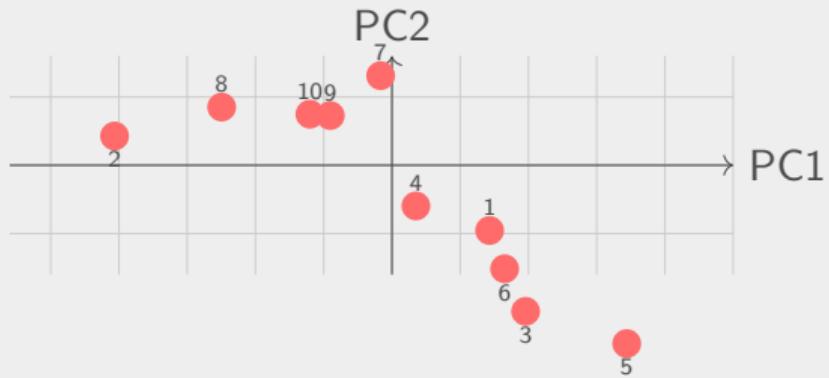


Paso 7b: Proyección a 2D - Resultados

ID	PC1	PC2
1	0.795	-0.239
2	-2.258	0.107
3	1.088	-0.535
4	0.195	-0.149
5	1.911	-0.653
6	0.917	-0.379
7	-0.092	0.328
8	-1.386	0.213
9	-0.502	0.182
10	-0.668	0.187

Los 10 puntos originales de 3D ahora representados en 2D, reteniendo 99.6 % de la varianza

Visualización 2D: Datos proyectados

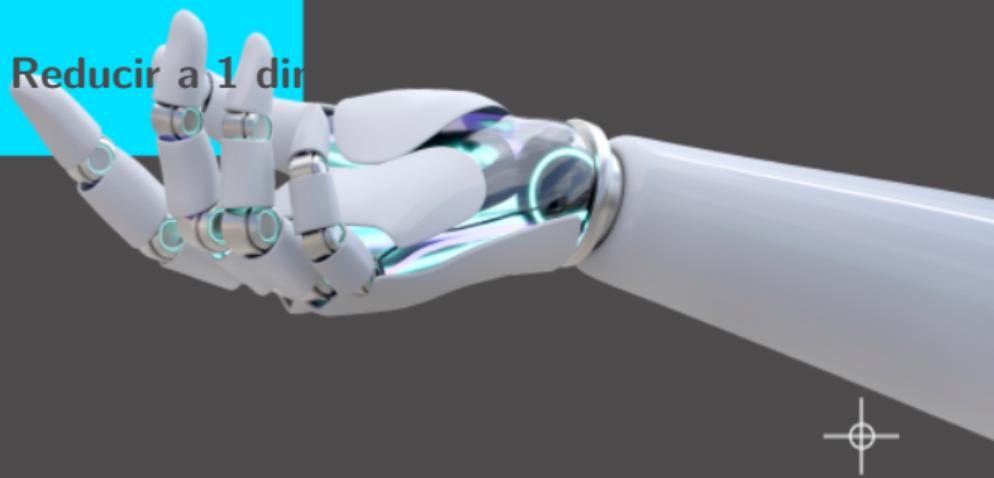


Datos originales de 3D ahora representados en 2D, reteniendo 99.6 % de la varianza.



Ejercicio 2

Dataset: 10 registros, 3 dimensiones **Reducir a 1 dim**



Tenemos 10 observaciones en 3 dimensiones ($D = 3$). Objetivo: reducir a $M = 1$.

ID	x_1	x_2	x_3
1	3.0	2.5	1.5
2	1.2	1.0	0.6
3	2.8	2.3	1.4
4	2.5	2.1	1.2
5	3.3	2.8	1.7
6	2.7	2.2	1.3
7	2.3	1.9	1.1
8	1.5	1.3	0.7
9	1.8	1.5	0.9
10	1.4	1.2	0.8

Paso 1: Calcular la media

Complete los cálculos:

$$\bar{x}_1 = \frac{1}{10}(3,0 + 1,2 + \dots) = \frac{\text{_____}}{10} = \text{_____}$$

$$\bar{x}_2 = \frac{1}{10}(\text{_____}) = \frac{\text{_____}}{10} = \text{_____}$$

$$\bar{x}_3 = \frac{1}{10}(\text{_____}) = \frac{\text{_____}}{10} = \text{_____}$$

Vector de medias: $\bar{x} = (\text{_____}, \text{_____}, \text{_____})^T$



Paso 2: Centrar los datos

Complete la tabla de datos centrados:

ID	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3
1	_____	_____	_____
2	_____	_____	_____
3	_____	_____	_____
:	:	:	:
10	_____	_____	_____



Paso 3: Matriz de covarianza

Complete la matriz de covarianza:

$$\mathbf{S} = \begin{pmatrix} \text{_____} & \text{_____} & \text{_____} \\ \text{_____} & \text{_____} & \text{_____} \\ \text{_____} & \text{_____} & \text{_____} \end{pmatrix}$$



Paso 4: Eigenvalores y eigenvectores

Complete los eigenvalores:

$$\lambda_1 = \text{_____} (\text{_____}\%)$$

$$\lambda_2 = \text{_____} (\text{_____}\%)$$

$$\lambda_3 = \text{_____} (\text{_____}\%)$$

Complete el primer eigenvector (dirección de máxima varianza):

$$\mathbf{v}_1 = \begin{pmatrix} \text{_____} \\ \text{_____} \\ \text{_____} \end{pmatrix}$$

Con 1 componente retenemos _____ de la varianza total.

Paso 5: Proyección a 1D

Complete la tabla de proyecciones:

ID	PC1
1	_____
2	_____
3	_____
4	_____
5	_____
6	_____
7	_____
8	_____
9	_____
10	_____



Visualización 1D: Datos proyectados

Grafique aquí los 10 puntos proyectados
según sus cálculos del paso anterior



Conclusiones del Ejercicio 2

Complete sus conclusiones:





UTEC Posgrado



UTECH Posgrado