**UNIVERSIDAD NACIONAL DEL ALTIPLANO**

**FACULTAD: INGENIERÍA DE MECÁNICA ELÉCTRICA,**

**ELECTRÓNICA Y SISTEMAS**

**ESCUELA PROFESIONAL DE INGENIERIA DE SISTEMAS**



**INFORME:**

ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

**PRESENTADO POR:**

QUEA MAMANI, EDDY GONZALO

**CURSO:**

ANALISIS DE DATOS

**DOCENTE ENCARGADO:**

JIMENEZ CHURA ADOLFO CARLOS

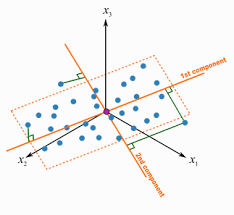
SEMESTRE IX

*PUNO, PERU*

*2025*

1. **Introducción**

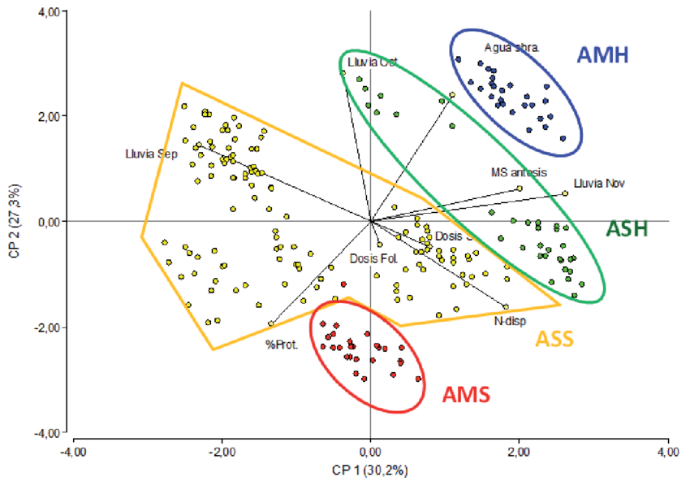
El Análisis de Componentes Principales (PCA, por sus siglas en inglés) es una técnica estadística ampliamente utilizada para la reducción de dimensionalidad en conjuntos de datos. Su objetivo es transformar un conjunto de variables posiblemente correlacionadas en un nuevo conjunto de variables linealmente no correlacionadas llamadas componentes principales. Esta transformación permite simplificar la estructura de los datos, conservar la mayor parte de su varianza y facilitar su visualización y análisis.



1. **Fundamentos del PCA**

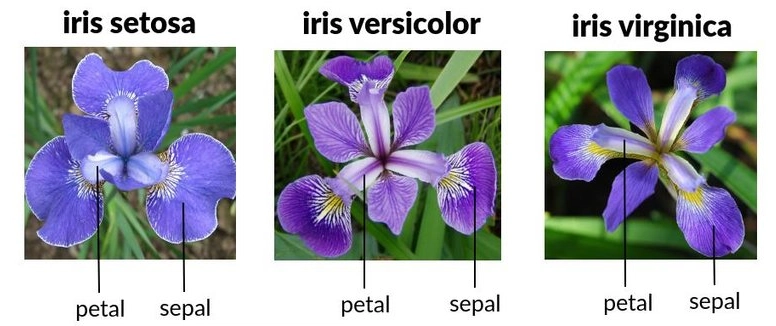
PCA se basa en los siguientes pasos principales:

* Estandarización de los datos para que cada variable tenga media cero y varianza uno.
* Cálculo de la matriz de covarianza para observar la relación entre las variables.
* Obtención de los vectores y valores propios (eigenvectors y eigenvalues) de la matriz de covarianza.
* Selección de los componentes principales basándose en los valores propios más altos.
* Proyección de los datos originales en el nuevo espacio reducido.



1. **Aplicación Práctica con el Dataset Iris**

Para la aplicación del PCA se utilizó el dataset Iris provisto por la librería sklearn.datasets. Este conjunto contiene 150 muestras de flores de tres especies de Iris (Setosa, Versicolor, Virginica), con cuatro características: longitud y ancho del sépalo y del pétalo.



**Código:**

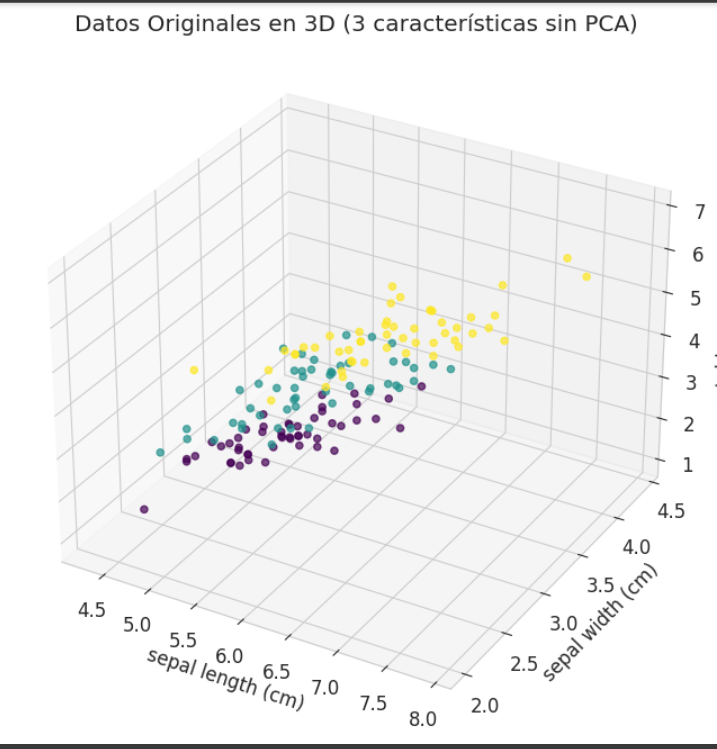
from sklearn.datasets import load\_iris

Se aplicó PCA para reducir la dimensionalidad de las 4 variables a 2 componentes principales, permitiendo visualizar claramente la separación entre especies, especialmente la clase Iris setosa, que es linealmente separable.

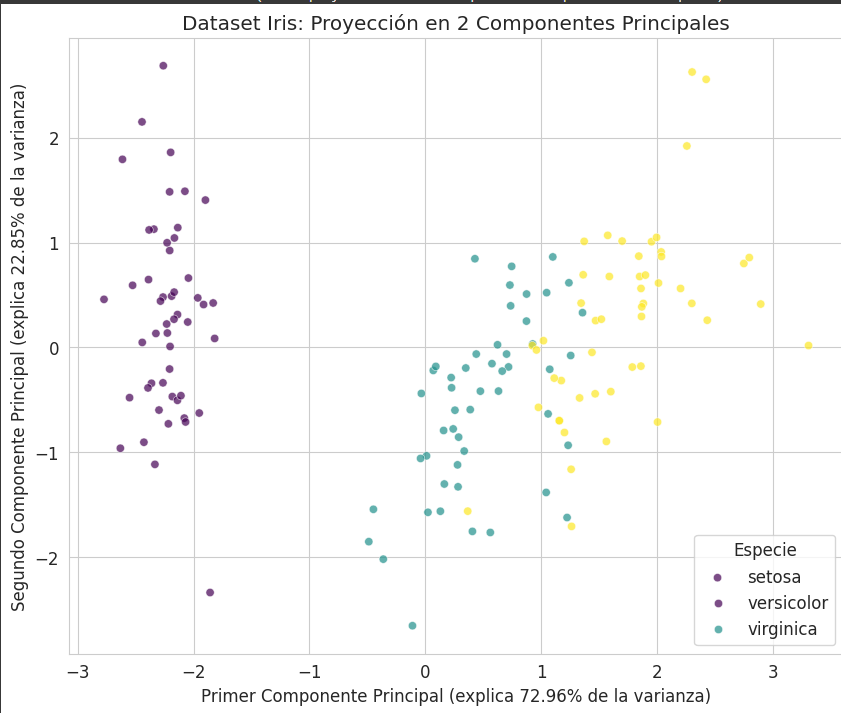
**4. Ventajas del PCA**

* + Reducción de dimensionalidad con mínima pérdida de información.
  + Mejora de la eficiencia computacional en algoritmos de machine learning.
  + Eliminación del ruido y redundancia.
  + Visualización clara de relaciones en los datos.

**Antes del PCA**



**Después del PCA**



**5. Aportes de los Artículos Científicos Consultados**

**Artículo 1:** "Principal component analysis: a review and recent developments" - J. Shlens

**Enlace:** <https://pmc.ncbi.nlm.nih.gov/articles/PMC4792409/>

**Aportes clave:**

* + Presenta una revisión profunda del trasfondo matemático del PCA.
  + Explica con claridad las conexiones entre PCA, álgebra lineal y estadística.
  + Destaca su relación con el análisis de varianza y el valor informativo de los vectores propios.
  + Resalta el uso de PCA en campos como la neurociencia y el reconocimiento de patrones.
  + Introduce mejoras y variantes modernas del PCA, como probabilistic PCA y kernel PCA.

**Artículo 2:** "An Exploration of the Application of Principal Component Analysis in Big Data Processing" - Xuedong Yu, Yan Wang

**Enlace:** <https://www.researchgate.net/publication/378625637_An_Exploration_of_the_Application_of_Principal_Component_Analysis_in_Big_Data_Processing>

**Aportes clave:**

* + Enfocado en la aplicación práctica del PCA en entornos de big data.
  + Describe cómo PCA mejora el procesamiento y almacenamiento en sistemas distribuidos.
  + Muestra casos de uso en visión por computadora, minería de texto y bioinformática.
  + Compara el rendimiento de PCA tradicional con variantes escalables para grandes volúmenes de datos.
  + Propone integrar PCA con tecnologías como Hadoop y Spark para mejorar eficiencia.

**6. Conclusiones**

El Análisis de Componentes Principales es una técnica poderosa tanto en contextos académicos como industriales. A través de la aplicación práctica con el dataset Iris y la revisión de literatura especializada, se puede concluir que:

* + PCA es esencial para la compresión de datos, la visualización y la preparación previa al modelado.
  + Su base matemática sólida permite confiar en su interpretación y resultados.
  + En la era del big data, es importante adaptar técnicas como PCA para asegurar su escalabilidad y rendimiento.