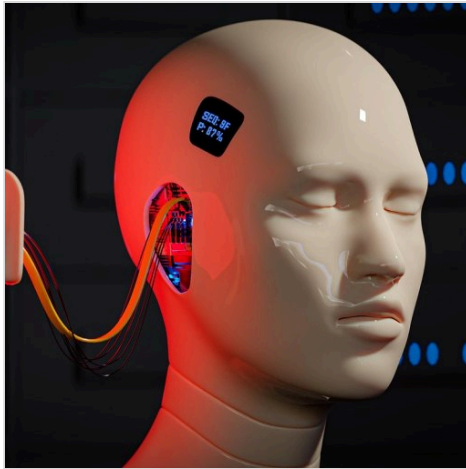


L U M B R E



PCA aplicado

Grado: 5

Materia: Aprendizaje No Supervisado · Machine Learning

11/10/2026

Antes de empezar

El análisis de componentes principales gira el espacio para mirar los datos de frente: a lo largo de las direcciones donde más varían. Reduce decenas de columnas correlacionadas a unas pocas ortogonales que conservan casi toda la información, y revela lo esencial.

QUÉ APRENDERÁS

- Proyectar datos a sus componentes principales para visualizar, descorrelacionar y comprimir, comprendiendo qué varianza se conserva.

PCA aplicado

Girar el espacio para ver la señal

Muchas variables cuentan la misma historia con distintas palabras. El PCA encuentra las direcciones donde tus datos realmente se mueven y descarta la redundancia, condensando la señal en unas pocas componentes limpias y no correlacionadas.

VIDEO

Análisis de componentes principales (PCA)

Vídeo en español que explica cómo el PCA gira el espacio hacia las direcciones de máxima varianza para resumir los datos.



Escanea para verlo en Lumbre

https://www.youtube.com/watch?v=7My_PBhxeP4

Reducir sin perder la señal

Cuando las características están muy correlacionadas, hay información duplicada. Reducir dimensiones consiste en hallar un puñado de combinaciones de las originales que capturen la mayor parte de la estructura, con menos ruido y menor coste computacional.

La varianza como información

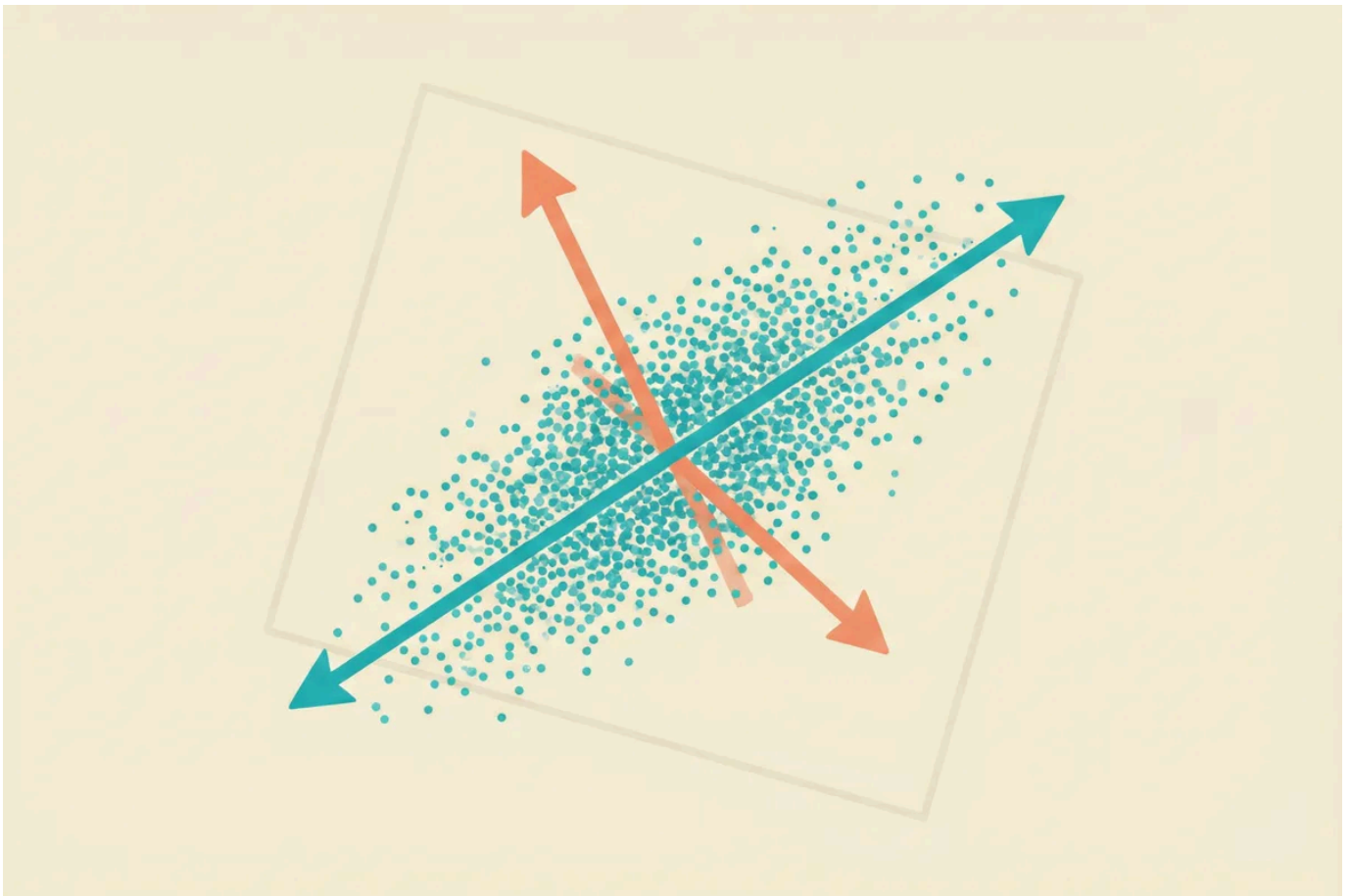
El PCA asume que donde hay más variación hay más señal. Busca, por tanto, ejes nuevos a lo largo de los cuales los datos se dispersan más. La primera componente es la dirección de máxima varianza; la segunda, la siguiente, y así decreciendo.

Componentes principales

Cada componente es una combinación lineal de todas las variables originales, elegida para capturar el máximo de varianza restante y ser ortogonal a las anteriores. Los ejes no son variables: son direcciones inventadas que resumen a las reales.

Ortogonalidad y descorrelación

Las componentes son perpendiculares entre sí, luego no correlacionadas. Esa es una de sus grandes virtudes: convierte un manojito de columnas redundantes en otras tantas independientes, ideales para modelos que sufren con la multicorrelación.



La intuición del PCA: la flecha larga es la primera componente, la dirección donde los datos más varían; la corta, ortogonal, captura la siguiente variación restante.

Escalar antes de todo

El PCA mira varianzas, y una variable con unidades enormes tendrá varianza enorme y secuestrará las componentes. Estandarizar cada columna a media cero y desviación uno es un paso casi siempre obligatorio para que la reducción refleje estructura y no escalas.

¿Cuántas componentes guardar?

La curva de varianza acumulada responde: cuántas componentes hacen falta para retener, digamos, el noventa por ciento de la varianza total. Se busca el punto donde añadir componentes ya apenas aporta, equilibrando compresión y pérdida.

Visualizar en dos o tres dimensiones

Proyectar sobre las dos primeras componentes permite pintar en un plano datos de cientos de dimensiones. No es una imagen literal sino una sombra informada: si las dos primeras capturan mucha varianza, el mapa es fiel; si no, engaña.

PCA no es agrupamiento

Las componentes ordenan por varianza, no por clases. Que dos grupos se separen en el plano proyectado depende de si esa separación es la dirección de mayor variación. A veces la señal útil vive en componentes de baja varianza que el PCA descarta.

Errores frecuentes

- No estandarizar y dejar que una columna de gran escala domine las componentes.
- Interpretar una componente como si fuera una variable original con significado directo.
- Usar las primeras componentes para clasificar sin comprobar que la separación relevante esté ahí.

Ejemplo resuelto: resumir una encuesta de cuarenta preguntas

Un estudio social aplica cuarenta ítems muy correlacionados y quiere visualizar sin perder señal. El protocolo:

1. Escalar: estandarizar cada pregunta, porque si no las de mayor varianza secuestran el resultado.
2. Centrar y descomponer: calcular las direcciones principales de máxima varianza.

3. Decidir cuántas: leer la curva de varianza acumulada y quedarse con lo que retiene, digamos, noventa por ciento.
4. Proyectar: llevar las respuestas a dos componentes y pintar el mapa.
5. Nombrar: mirar las cargas y bautizar cada componente según las preguntas que la dominan.

¿Para qué sirve en la realidad?

El PCA es pan de cada día para comprimir antes de modelar, descorrelacionar predictores, quitar ruido y visualizar datos de decenas de dimensiones en un plano honesto, siempre que las primeras componentes capturen de verdad la señal que importa.

1. ¿Qué maximiza la primera componente principal de un PCA?

- ☐ La correlación con las etiquetas de clase
- ☐ La varianza de los datos proyectados sobre ella
- ☐ El número de variables originales involucradas
- ☐ La compactez de los grupos formados

2. Tras PCA, las componentes resultantes están libres de correlación entre sí.

- ☐ Verdadero
- ☐ Falso

Interpretar las cargas

Cada componente es una mezcla ponderada de las variables originales: las cargas dicen cuánto aporta cada una. Una componente dominada por indicadores económicos quizá se lea como un eje de «riqueza». Nombrar componentes es un arte interpretativo útil.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. PCA



4. Ordena el flujo de un PCA aplicado:

Elementos:

- estandarizar cada variable a escala comparable
- centrar los datos restando la media
- hallar direcciones principales de maxima varianza
- proyectar sobre las componentes elegidas

Escalar	. . .
Centrar	. . .
Descomponer	. . .
Proyectar	. . .

PCA explicado visualmente <https://setosa.io/ev/principal-component-analysis/>
Explicación interactiva e intuitiva de componentes principales y varianza.

5. Une cada pieza del PCA con su papel.

5.	dirección de máxima varianza de los datos	-----
5.	combinación lineal de variables originales	-----
5.	propiedad que descorrelaciona los ejes	-----
5.	curva para decidir cuántas guardar	-----

Opciones: primera componente · componente principal · ortogonalidad · varianza acumulada

Condensar sin perder señal

PCA aplicado

Idea

maximizar varianza

girar ejes

Componentes

ortogonales

combinaciones

Decisión

varianza acumulada

cuántas guardar

Usos

visualizar

descorrelacionar

El mapa mental interactivo es interactivo en la versión online.



Análisis de componentes principales https://es.wikipedia.org/wiki/An%C3%A1lisis_de_componentes_principales Referencia en español de la teoría y propiedades del PCA.

6. Proyectas tus datos a dos componentes y los grupos se solapan, pero sabías que existían diferencias claras. ¿Qué te dice esto sobre dónde vive la señal separadora?

Referencia curricular

Grado:	5
Materia:	Aprendizaje No Supervisado · Machine Learning

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-nosu-004-pca-aplicado>

Material educativo generado desde Lumbre · lumbre.lat