

PCA

Bustamante Callisaya Hansel Alain

PCA (Análisis de Componentes Principales) es una técnica utilizada para simplificar un conjunto de datos, especialmente cuando tenemos muchas variables. La idea principal es encontrar las direcciones en las que los datos tienen más variabilidad, y luego reducir el número de variables manteniendo la mayor parte de esa información.

¿Cómo funciona PCA?

1. **Centro de datos:** Primero, restamos la media de cada característica (o columna) de nuestros datos. Esto se hace para que nuestros datos estén centrados en el origen (cero), lo que facilita el análisis posterior.
2. **Matriz de covarianza:** Luego calculamos la matriz de covarianza, que nos da una idea de cómo las diferentes características varían juntas. Si dos características están muy relacionadas, su covarianza será alta.
3. **Vectores propios y valores propios:** A continuación, buscamos los **vectores propios** (direcciones principales de variabilidad) y los **valores propios** (que nos dicen cuánta variabilidad hay en esas direcciones). Los vectores propios más importantes son aquellos que tienen los valores propios más altos, ya que muestran las direcciones con mayor dispersión en los datos.
4. **Proyección sobre los componentes principales:** Una vez que tenemos los vectores propios, proyectamos nuestros datos originales sobre estos nuevos ejes (que son las componentes principales). Esto nos permite reducir el número de características, ya que las primeras componentes principales capturan la mayor parte de la variabilidad.
5. **Reducción de dimensionalidad:** Finalmente, si seleccionamos solo los primeros **k** componentes principales (los más importantes), podemos reducir el número de dimensiones en nuestros datos sin perder mucha información.