

Investigación PCA

El PCA es un algoritmo que selecciona características concretas para convertir un conjunto de variables a uno más reducido y se conocen como componentes principales. El algoritmo ayuda a encontrar la estructura subyacente la cual se define por la dirección donde la varianza de datos es mayor. Al saber cuales son las variables que aportan más información se descartan las menos importantes y así se reduce el conjunto de datos. Además de reducir el conjunto, este ayuda a detectar anomalías.

Aunque es un gran algoritmo que ayuda en bastantes cosas, tiene limitaciones por lo que se recomienda hacer una normalización o estandarización antes de aplicar el algoritmo. Además, al usar PCA, este algoritmo asume que los datos siguen una distribución normal o gaussiana.

Referencias

Santos, P. R. D. L. (2021, 16 junio). *Python para todos: ¿Qué es el análisis de Componentes Principales o PCA?* - Think Big Empresas. Think Big.
<https://empresas.blogthinkbig.com/python-para-todos-que-es-el-pca/>

Pregunta: ¿Cómo podría ayudarnos a mejorar la calidad de nuestros clusters?

Al reducir el tamaño del conjunto y reducir anomalías, se puede mejorar el clustering al identificar los patrones importantes de los datos.