¿Por qué deberías utilizar las técnicas de reducción de dimensionalidad para transformar tus datos antes de entrenarlos?

Porque este reduce el costo computacional de entrenar un modelo de aprendizaje automatico. tambien ayuda a la visualizacion ya que es mas facil representarlos en un espacio de menor dimension.

¿Por qué deberías evitar las técnicas de reducción de dimensionalidad para transformar tus datos antes de entrenarlos?

Al reducir la dimensionalidad de los datos es posible que se pierda informacion importante, si la reduccion elimina caracteristicas que son criticas para el modelo, la precision puede verse comprometida.

Nombra un algoritmo de reducción de dimensionalidad popular y descríbelo brevemente.

PCA, analisis de componentes principales, este busca encontrar una combinacion lineal de las caracteristicas originales que maximice la varianza de los datos. que resulta en una reduccion de dimensionalidad.

Antes de realizar una reducción de dimensionalidad, ¿puedes transformar los datos de nuevo en el espacio de características original? ¿Cómo?

Dependera de la tecnica realizada. Si es PCA entonces se pueden multiplicar los datos reducidos por la matriz transpuesta de los componentes principales, matriz de vectores propios, que se obtuvieron durante el ajuste de PCA.

¿Cómo se seleccionan los números de los componentes principales necesarios para ACP?

Se seleccionan mediante pca(n\_stimators=x) donde x puede ser el porcentaje de variance buscado o la cantidad de componentes a utilizar

¿Cuándo utilizarías ACP?

Cuando se tiene un conjunto de datos muy grande, que pueda impactar en tiempo y computo o cuando se requiera visualizar un conjunto de datos con muchas caracteristicas.

¿Por qué realizarías ACP incluso si no tienes muchas características?

Para visualizar datos de alta dimensionalidad

¿En qué casos NO utilizarías ACP?

Si los datos son categoricos, el acp esta diseñado para trabajar con datos numericos. o si los datos tienen una distribucion no normal o valores atipicos extremos.

¿Cuál es la interpretación geométrica de un vector propio y valor propio?

El vector propio es un vector que no cambia de direccion cuando se transforma mediante la matriz de covarianza de los datos. Es decir, es una direccion en la cual los datos mantienen una estructura determinada despues de la transformacion.En ACP los vectors propios representan las direcciones en las cuales los datos tienen la maxima variacion.

El valor propio es un numero que representa la cantidad de varianza explicada por el vector propio correspondiente. Mide la importancia de un vector propio en la estructura de datos. En ACP los valores propios se utilizan para seleccionar el numero adecuado de componentes principales a retener. los valores propios mas grandes corresponden a las direcciones con la mayor cantidad de varianza explicada.

En terminos geometricos, representan la direccion y la magnitud de la variacion en los datos. Los vectores propios representan las direcciones principales de variacion, mientras que los valores propios representan la cantidad de variacion explicada en esas direcciones.

¿Cuál es la interpretación algebraica de un vector propio y valor propio?

Un vector propio es un vector que satisface la siguiente ecuacion: A \* v = λ \* v

Donde A es la matriz de covarianza de los datos, v es el vector propio y λ es el valor propio correspondiente. Esta ecuacion indica que cuando se multiplica la matriz de covarianza por un vector propio, el resultado es un multiplo del mismo vector propio. es decir el vector propio no cambia su direccion bajo la transformacion lineal dada por la matriz de covarianza.