Algoritmo	UMAP	LDA	PCA	T-SNE
Nombre	Uniform Manifold Approximation and Projection	Linear Discriminant Analysis	Principal Component Analysis	
Objetivo	Preservar la estructura local y global de los datos en un espacio de dimensionalidad reducida	Maximizar la separabilidad entre clases mientras reduce la dimensionalidad de los datos	Encontrar las direcciones (o ejes) de máxima varianza en los datos	Preservar la estructura local de los datos
Tipo	No lineal Aprendizaje supervisado	Lineal Aprendizaje supervisado	Líneal	No lineal
Principios matemáticos	Suposición de manifold Conjunto simplicial difuso Geometría riemanniana **Optimización estocástica** **Topología algebraica** Teoría de grafos	- Álgebra linea - Eigenvectors/values - Combinación lineal	Algebra linea Matriz de covarianza Eigenvectors/values	- Probabilidad de vecinos - Minimizar la divergencia de Kullback-Leibler - Distribución t-student
Aplicaciones	- Visualización - Agrupamiento - Extracción de características - Detección de anomalías	- Visualización - Extracción de características - Clasificación	- Visualización - Extracción de características - Compresión de datos	- Visualización - Extracción de características - Imágenes
Ejemplo datos	Datos no lineales Biología, genómica, expresión génica	Con etiquetas de clase	- Datos numéricos continuos (finanzas) - Imágenes	- Datos no lineales como imágenes de MNIST o NLP
Implementación	https://github.com/lmcinnes/umap	https://developer.ibm.com/tutorials/awb-implementing-	https://github.com/rushter/MLAlgorithms/blob/master/mla/pca.py	https://github.com/rushter/MLAlgorithms/blob/master/mla/tsne.py

		linear-discriminant-analysis-python/		

⇨ LDA es menos flexible en comparación con UMAP, ya que asume una relación lineal entre las características y las clases.

UMAP

UMAP (Aproximación y Proyección Uniforme de Manifolds) es una técnica de reducción de dimensionalidad que tiene como objetivo preservar tanto la estructura local como global en datos de alta dimensionalidad. Está basada en un marco matemático con:

- **Suposición de manifold:** UMAP asume que los datos de alta dimensionalidad yacen en un manifold de dimensionalidad inferior incrustado en el espacio original. El objetivo es encontrar una representación de baja dimensionalidad que preserve la estructura del manifold.

- **Conjunto simplicial difuso:** UMAP construye una representación topológica difusa de los datos creando un grafo ponderado de vecinos más cercanos. Este grafo captura la estructura local y global, y se utiliza para aproximar el manifold subyacente.

- **Geometría riemanniana:** UMAP optimiza una función objetivo que mide la similitud entre el conjunto simplicial difuso en el espacio de alta dimensionalidad vs el espacio de baja dimensionalidad. Esta optimización está guiada por principios de geometría riemanniana, para encontrar una incrustación que minimice la distorsión.

- **Optimización estocástica:** UMAP emplea un algoritmo de optimización estocástica, como el descenso de gradiente estocástico (SGD). Este proceso ajusta iterativamente la incrustación para minimizar la discrepancia entre las representaciones de alta y baja dimensionalidad.

Es útil para:

- **Visualización**: UMAP puede utilizarse para visualizar datos de alta dimensionalidad en dos o tres dimensiones.

- **Agrupamiento**: Las incrustaciones de UMAP pueden utilizarse como características de entrada para algoritmos de agrupamiento, ayudando a identificar grupos en datos de alta dimensionalidad.
- **Extracción de características:** UMAP puede utilizarse como técnica de extracción de características para reducir la dimensionalidad de los datos antes de alimentarlos en modelos de aprendizaje automático (mejora el rendimiento del modelo y reduce la complejidad computacional).

- **Detección de anomalías:** Las incrustaciones de UMAP pueden utilizarse para identificar valores atípicos o anomalías en datos de alta dimensionalidad al medir la distancia entre los puntos de datos en el espacio de baja dimensionalidad.

LDA

Análisis Discriminante Lineal (LDA) es una técnica supervisada que se utiliza principalmente en problemas de clasificación. Su objetivo es proyectar los datos de alta dimensión en un espacio de menor dimensión mientras maximiza la separabilidad entre clases. Funciona encontrando las direcciones (llamadas discriminantes lineales) en las cuales las clases son más separables.

Es una técnica supervisada que requiere etiquetas de clase para aprender la proyección óptima de los datos.

Funciona encontrando una **combinación lineal de características** que caracteriza o separa dos o más clases de objetos o eventos. Los principios matemáticos detrás de LDA implican maximizar la separabilidad entre clases mientras se minimiza la varianza dentro de cada clase.

Implementación:

1. **Calcular los vectores de medias:** Se calculan los vectores de medias para cada clase en el conjunto de datos.
2. **Calcular las matrices de dispersión:** Se calcula la matriz de dispersión dentro de las clases (SW) y la matriz de dispersión entre clases (SB). SW representa la dispersión de los datos dentro de cada clase, mientras que SB representa la dispersión entre clases.
3. **Calcular los autovectores y autovalores:** Se calculan los autovectores y autovalores de la matriz
4. **Seleccionar los vectores discriminantes:** Se seleccionan los k autovectores correspondientes a los k autovalores más grandes para formar una matriz de transformación W.

5. **Proyectar los datos en el nuevo subespacio de características**: Se proyectan los datos originales en el nuevo subespacio de características definido por la matriz W.