# **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.

Aquí hay un resumen comparativo entre el LDA para reducción de dimensionalidad y el LDA para modelización de temas en documentos de texto:

Su objetivo principal es reducir la dimensionalidad de los datos mientras mantiene la información relevante para mejorar la separabilidad entre clases en problemas de clasificación.

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algoritmo | UMAP | LDA | PCA | T-SNE |
| Nombre | Uniform Manifold Approximation and Projection | Linear Discriminant Analysis | Principal Component Analysis |  |
| Tipo | Aprendizaje no supervisado | Aprendizaje supervisado |  |  |
| Principios matemáticos | Suposición de manifold  Conjunto simplicial difuso  Geometría riemanniana  **Optimización estocástica**  **Topología algebraica**  Teoría de grafos |  | Algebra linea  Eigenvectors/values |  |
| Aplicaciones | - Visualización  - Agrupamiento  - Extracción de características  - Detección de anomalías |  | - Extracción de características | - Extracción de características  - Imágenes |
| Ejemplo datos | Biología, Genómica |  |  |  |
| Implementación | <https://github.com/lmcinnes/umap> | <https://developer.ibm.com/tutorials/awb-implementing-linear-discriminant-analysis-python/> | <https://github.com/rushter/MLAlgorithms/blob/master/mla/pca.py> | <https://github.com/rushter/MLAlgorithms/blob/master/mla/tsne.py> |
| Enlaces |  |  |  |  |

<https://plotly.com/python/t-sne-and-umap-projections/>