

Métodos no lineales. Aprendizaje de *manifold*

Los métodos de reducción de dimensionalidad basados en el aprendizaje de *manifold* parten de la premisa de que los datos, a menudo, se distribuyen en una estructura intrínseca de menor dimensión. Al proyectarlos en este *manifold*, es posible entonces reducir la dimensionalidad mientras se preserva su estructura subyacente, lo que permite mantener la información relevante. Uno de los principios clave detrás de estos métodos es la preservación de las relaciones de vecindad entre los datos; es decir, datos que están cercanos en el espacio de alta dimensión también deberían estar próximos en el de menor dimensión (ver Fig. 1).

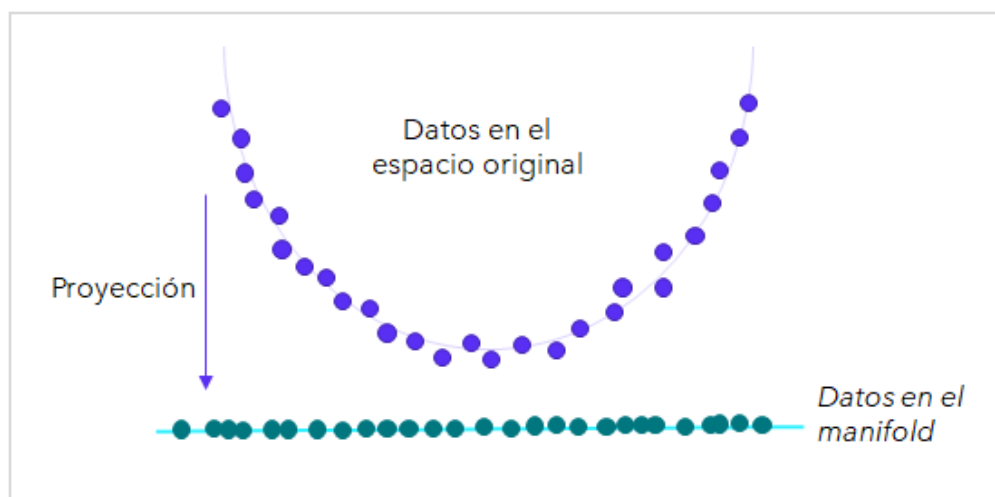


Fig. 1. Ejemplo de proyección de un conjunto de datos de dos dimensiones en un manifold de una dimensión.

Características.

- Manejo de datos no lineales. A diferencia de técnicas como el análisis de componentes principales (PCA), que asumen relaciones de linealidad entre los datos, los métodos de aprendizaje de *manifold* pueden manejar estructuras no lineales.

- Adaptación. Permiten modelar la estructura subyacente de los datos sin hacer suposiciones sobre la forma del *manifold*, por lo que pueden capturar una variedad de estructuras intrínsecas.

El algoritmo t-SNE

El algoritmo t-SNE (*t-distributed Stochastic Neighbor Embedding*) es una técnica de aprendizaje de *manifold* que resulta muy eficaz para preservar la estructura de los datos en pequeñas vecindades locales, resultando ideal para visualizar datos complejos y no lineales. Algunas de las aplicaciones comunes de t-SNE son:

- Visualización de datos. t-SNE se utiliza para visualizar datos de alta dimensionalidad en un espacio 2D o 3D, lo que permite identificar patrones, agrupaciones y relaciones entre los datos de una manera más intuitiva.
- Análisis de agrupación. En muchos casos, los datos se organizan naturalmente en grupos que siguen una estructura de *manifold*. t-SNE puede ayudar a identificar y visualizar estos clústeres de manera efectiva al preservar sus relaciones locales en la representación de baja dimensión.
- Visualización de incrustaciones de palabra. t-SNE se ha utilizado en el procesamiento del lenguaje natural para visualizar incrustaciones de palabras (*word embeddings*) en un espacio bidimensional, lo que habilita la identificación de relaciones semánticas entre palabras de manera efectiva.

Bibliografía

Hinton, G., van der Maaten, L. (2008). Visualizing Data using t-SNE. *Journal of machine learning research*. 9(86):2579–2605.

© - **Derechos Reservados:** la presente obra, y en general todos sus contenidos, se encuentran protegidos por las normas internacionales y nacionales vigentes sobre propiedad intelectual, por lo tanto su utilización parcial o total, reproducción, comunicación pública, transformación, distribución, alquiler, préstamo público e importación, total o parcial, en todo o en parte, en formato impreso o digital y en cualquier formato conocido o por conocer, se encuentran prohibidos, y solo serán lícitos en la medida en que se cuente con la autorización previa y expresa por escrito de la Universidad de los Andes.

De igual manera, la utilización de la imagen de las personas, docentes o estudiantes, sin su previa autorización está expresamente prohibida. En caso de incumplirse con lo mencionado, se procederá de conformidad con los reglamentos y políticas de la universidad, sin perjuicio de las demás acciones legales aplicables.
