

Elements of Machine Learning 2024

Tarea 03

08.marzo.2024

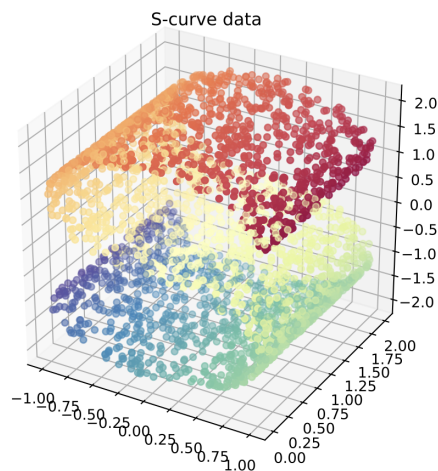
En esta tarea se requiere realizar visualizaciones de *manifold learning* sobre varios conjunto de datos.

1. Realizar el ejercicio del sistema de recomendación, con los datos de `recommender.ipynb`, `movies.csv`, `ratings.csv` usados en el aula del 16 de febrero.

Mostrar ejemplos o visualizaciones de algunos de los resultados obtenidos para diferentes usuarios. Tratar de interpretar en cada ejemplo cuál es el tipo de categoría o género de películas asociada a las diferentes variables latentes.

2. Hacer una comparación de distintos métodos de proyección local o manifold learning: Linear local embedding LLE, t-SNE, Spectral Embedding, ISOMAP y Escalamiento multidimensional MDS, para comparar diferentes visualizaciones de un conjunto de datos.

Aplicarlo para datos en 3D con la siguiente estructura de la Figura 2 (tomar 3000 observaciones). Muestre las proyecciones sobre los primeros dos componentes para cada método. Ilustre todas las proyecciones en un mismo plot, usando subplot, para facilitar la comparación visual.



La distribución de puntos de la Figura anterior en Python puede llamarse como:

```
from sklearn.datasets import make_s_curve
X, color = make_s_curve(n samples=3000).
```

3. El conjunto de datos **wines.csv**, contiene información sobre varios componentes químicos asociados a tres diferentes cepas de vino.

Elija 3 métodos de *Manifold Learning* (t-SNE debe ser uno de estos), y realice visualizaciones en 2D ó 3D para este conjunto. A partir de las visualizaciones, tratar de:

- Identificar alguna explicación para los ejes.
- Identificar posibles grupos o clústers.
- Verificar si dichos conglomerados separan bien a las tres diferentes cepas de vino.

A partir de sus visualizaciones, explicar las principales diferencias entre las tres cepas.

Más información sobre el origen de estos datos puede encontrarse en <http://archive.ics.uci.edu/ml/machine-learning-databases/wine>.

4. El índice de felicidad *Happy Planet Index* es un intento de medir el bienestar sostenible para todos, el cual puede encontrarse en el sitio <http://happyplanetindex.org/>.

Puedes descargar los datos del índice de felicidad 2016 en <http://happyplanetindex.org/resources> o también están disponibles en el archivo `hpi-data-2016.xlsx`.

Usar el método SOM para encontrar visualizaciones útiles para el conjunto de datos del índice de felicidad y de sus variables. Complementar estas visualizaciones con otros métodos, y hacer un análisis completo a partir de los datos.

Construir un "mapa de calor" que permita visualizar el índice de felicidad de cada país mediante tonos de color. Muestre su visualización y explique sus hallazgos y conclusiones sobre la relación entre este índice y las regiones geográficas.

No olviden redactar su análisis de datos, destacando sus conclusiones e *insights* más importantes.