

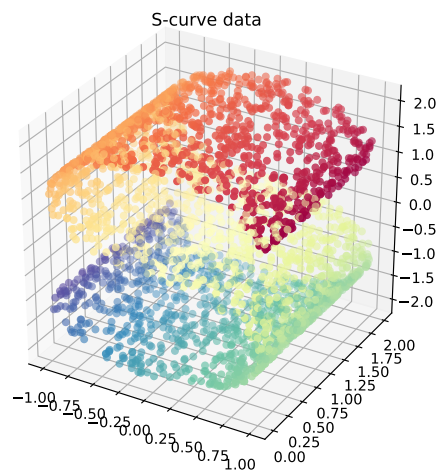
Ciencia de Datos 2021

Lista 04

11.marzo.2021

1. Haz 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.

Aplícalo para datos en 3D con la siguiente estructura de la Figura (toma 3000 observaciones). Muestra las proyecciones sobre los primeros dos componentes para cada método. Ilustra 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)
```

2. Usa el método *t*-SNE, para hacer una proyección del conjunto de datos *Digits*, el cual consiste de 1797 imágenes de dígitos de tamaño 8×8 (similar a MNIST, pero más simple).

Compara las vistas de las proyecciones a 2D o 3D usando diferentes valores de perplejidad: $\lambda = 0.01, 0.1, 1., 10., 100$. Agrega etiquetas en las gráficas para ver cómo se distribuyen las nubes de puntos en función de cada dígito.

Explica cuál es tu intuición sobre qué está haciendo este parámetro de perplejidad.

(Sugerencia: Usar el mismo parámetro `random_state` para todos los experimentos, para que el resultado sólo dependa de λ).

El conjunto *Digits* puede llamarse en Python como:

```
from sklearn import datasets
digits = datasets.load_digits()
X = digits['data']
y = digits['target']
```

3. 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`

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