DIPLOMADO
Ciencia de Datos

Módulo 7
Aprendizaje de máquina no supervisado
4. Manifold Learning

Eduardo Espinosa Avila

Universidad Nacional Autónoma de México
Dirección General de Cómputo y de Tecnologías de Información y Comunicación

DGTIC

Manifold Learning, *Kernel PCA*

Cuando tratamos con problemas de naturaleza no lineal, PCA no es una buena opción para reducir la dimensionalidad de los datos.

Para este tipo de situación existe una versión *kernelizada*, llamada KPCA.

La idea es aplicar una transformación no lineal para proyectar los datos en un espacio de dimensión superior.

Y posteriormente aplicar PCA para proyectarlo sobre un espacio de dimensión inferior, en el que los datos sean linealmente separables.

$$\mathbf{x} = [x_1, x_2]^T$$

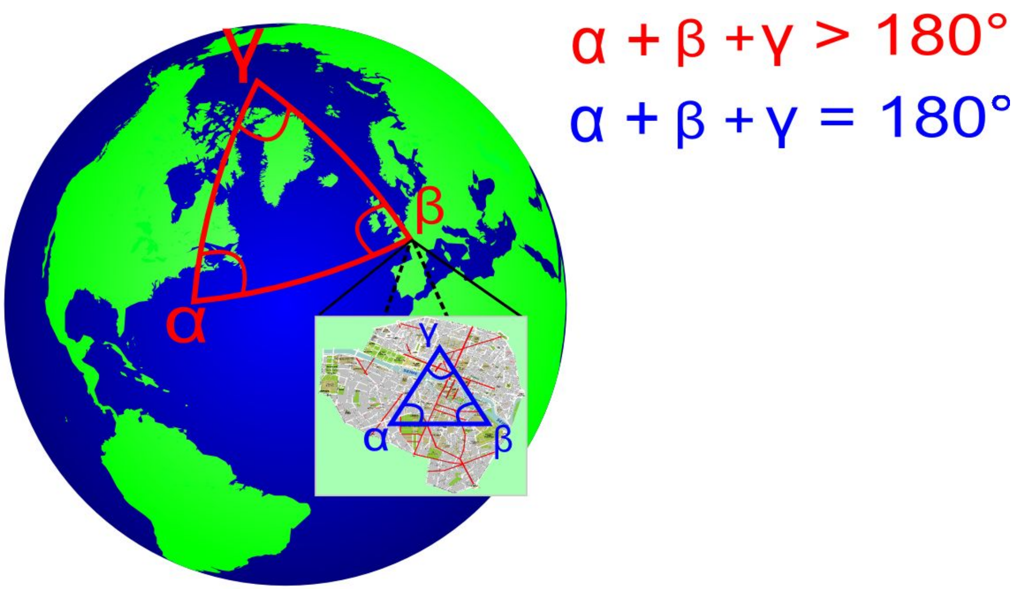
$$\downarrow \phi \downarrow$$

$$\mathbf{z} = \left[x_1^2, \sqrt{2x_1x_2}, x_2^2\right]^T$$

Manifolds

Un *manifold* es un espacio topológico, que localmente parece ser un espacio euclidiano en la cercanía de cada uno de sus puntos.

Para el caso del aprendizaje, en un *manifold* los datos son de dimensión baja *dentro* de espacios de alta dimensión.



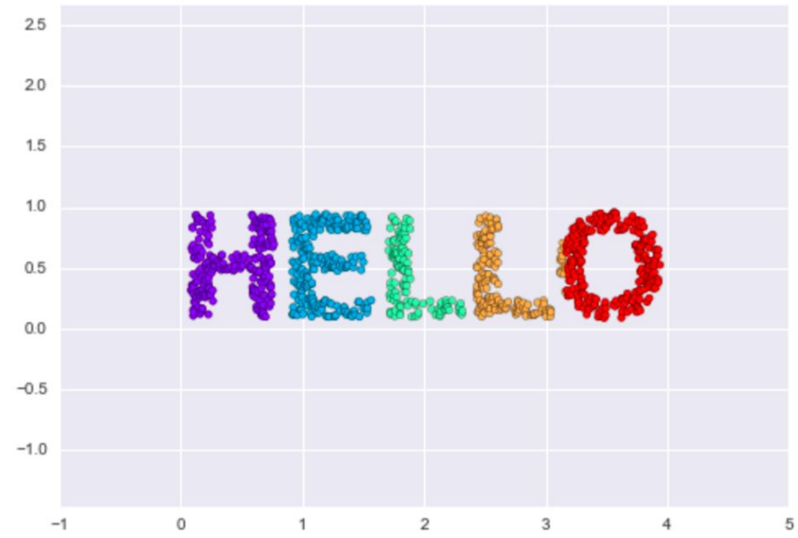
Dominio público,
https://commons.wikimedia.org/w/index.php?curid=1216752

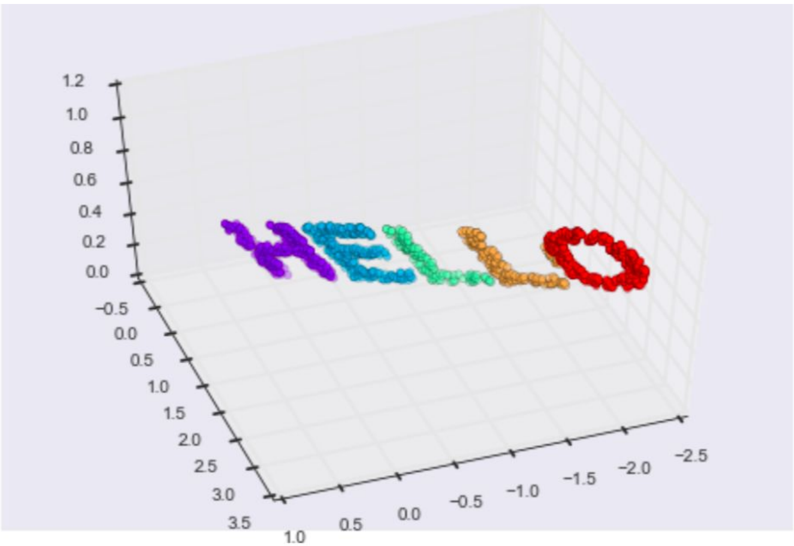
Manifolds

Por ejemplo, una hoja de papel es 2D en nuestro espacio 3D.

Reorientar o estirar la hoja dentro del espacio 3D no cambia la geometría plana de la hoja.

Estas operaciones son similares a transformaciones lineales en el espacio 3D.



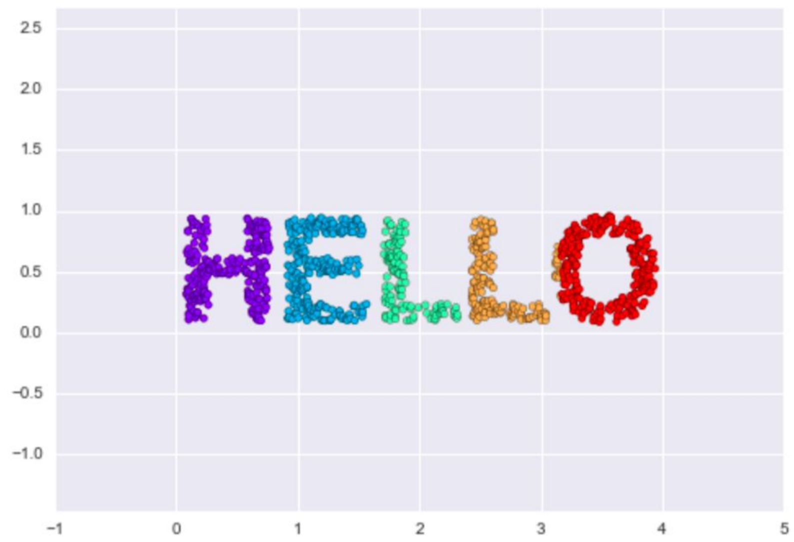


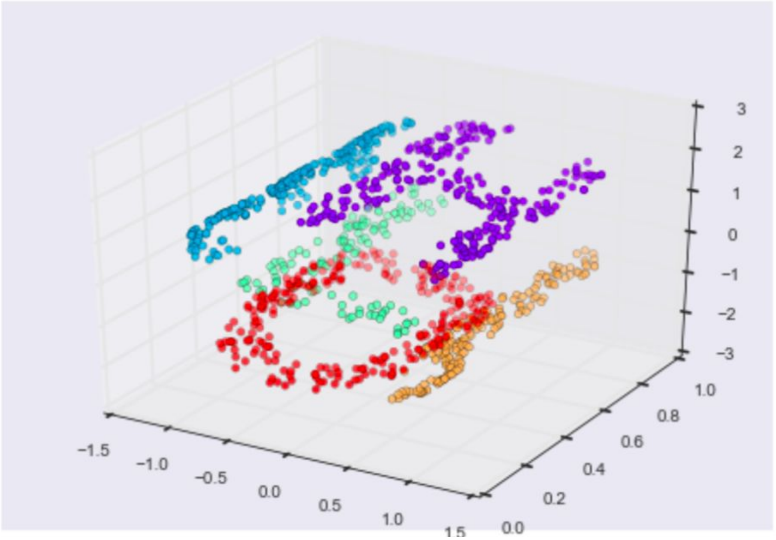
Manifolds

En cambio, doblar, enrollar o arrugar la hoja, conserva su naturaleza 2D.

Pero estas transformaciones ya no son lineales en el espacio 3D.

Los algoritmos de aprendizaje con *manifolds*, buscan encontrar la naturaleza 2D de la hoja, aún después de este tipo de operaciones.

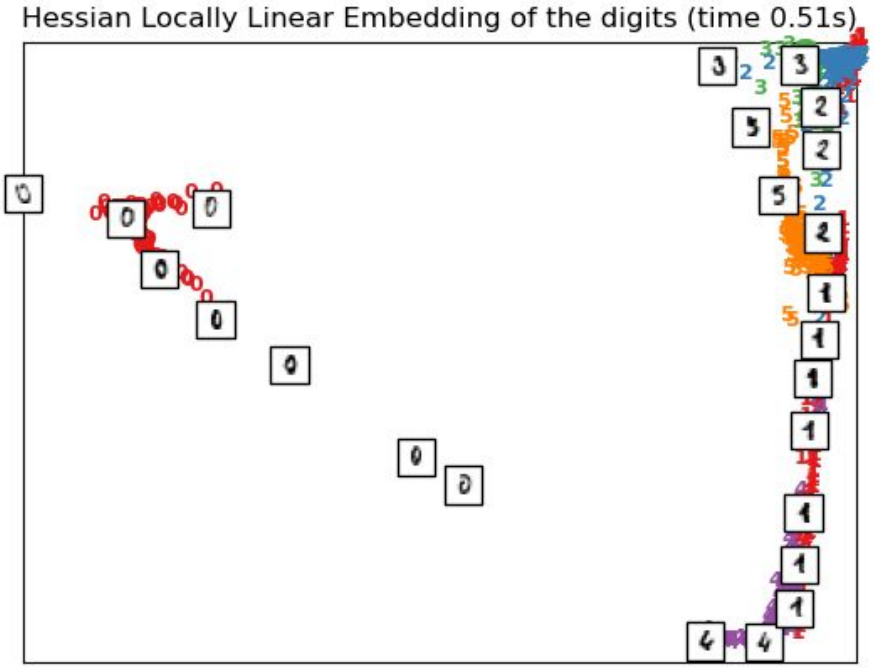




Manifold Learning, *Isomap*

Isometric Mapping (*isomap*) puede verse como una extensión de *Kernel PCA*.

Intenta encontrar la menor representación dimensional, manteniendo la [distancia geodésica]() entre los puntos.



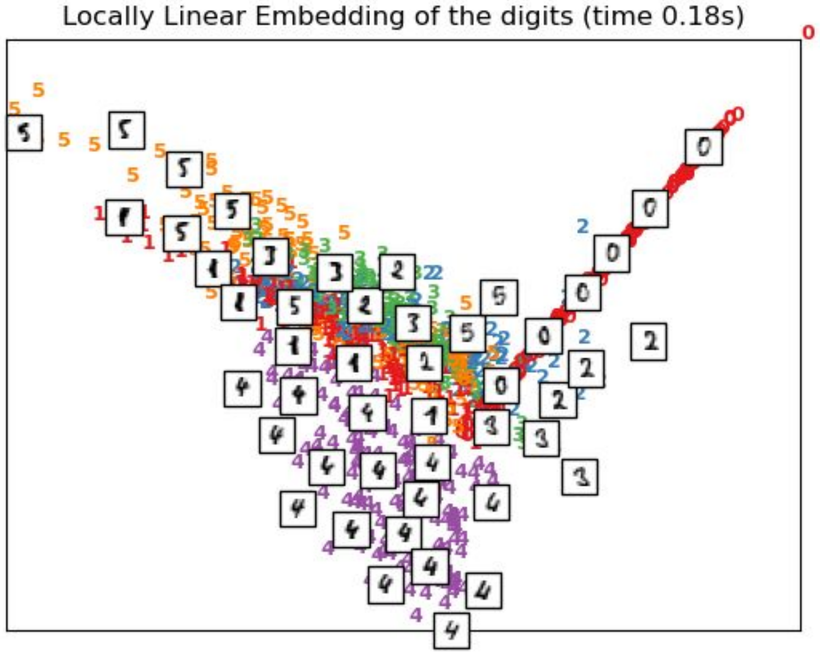
https://scikit-learn.org/stable/auto_examples/manifold/plot_lle_digits.html

Manifold Learning, *Locally Linear Embedding*

Locally Linear Embedding (LLE) busca la proyección de menor dimensión, manteniendo las distancias con el *vecindario local*.

Puede verse como la aplicación repetida de PCA a nivel local, que después se compara globalmente.

Al basarse en vecinos, es necesario proporcionar el número a considerar como parámetro.

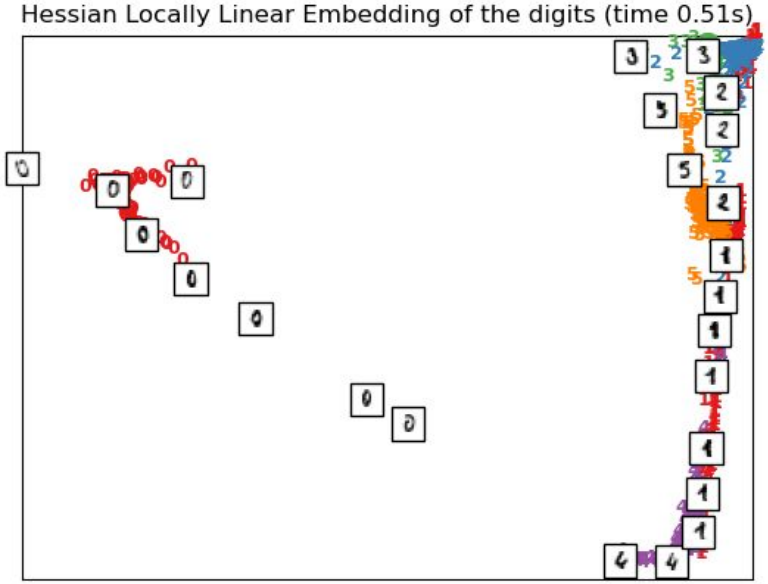


https://scikit-learn.org/stable/auto_examples/manifold/plot_lle_digits.html

Manifold Learning, *Hessian LLE*

LLE presenta un problema de regularización (sobreajuste), cuando el número de vecinos a considerar es mayor que el número de dimensiones de entrada.

Hessian LLE minimiza la [función hessiana]() del manifold, en el que residen los datos.



https://scikit-learn.org/stable/auto_examples/manifold/plot_lle_digits.html

Referencias

- VanderPlas, J.
  Python Data Science Handbook / Jake VanderPlas --
  USA : O'Reilly Media Inc., 2017 (530 páginas)

- James McQueen, Marina Meilˇa, Jacob VanderPlas and Zhongyue Zhang
  Megaman: Scalable Manifold Learning in Python
  https://jmlr.org/papers/volume17/16-109/16-109.pdf

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V. et al.
  Scikit-learn: Machine Learning in Python -- Manifold learning
  https://scikit-learn.org/stable/modules/manifold.html

Contacto

Dr. Eduardo Espinosa Avila

laloea@fisica.unam.mx

Tels: 5556225000 ext. 5003

Redes sociales:

https://twitter.com/laloea

https://www.linkedin.com/in/eduardo-espinosa-avila-84b95914a/