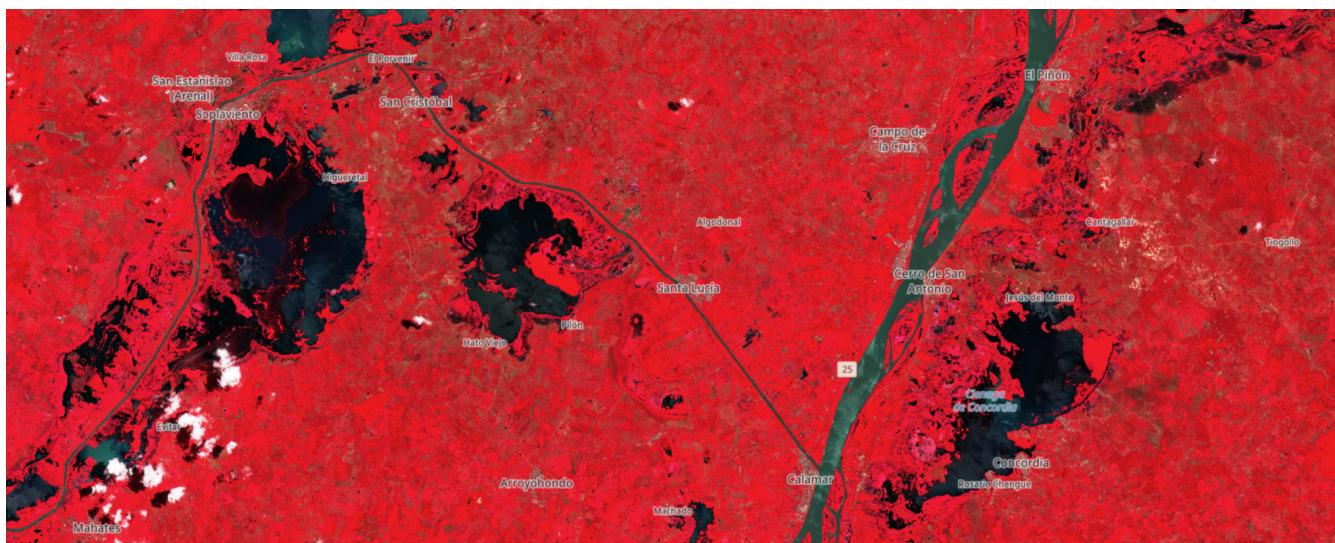

INTRODUCCIÓN A LOS MÉTODOS NO SUPERVISADOS DE APRENDIZAJE

DIPLOMADO EN MÉTODOS AVANZADOS EN PROCESAMIENTO DIGITAL DE IMÁGENES



31 de julio de 2025

Juan Sebastian Vinasco Salinas

Dirección de Investigación y Prospectiva



Índice

1. Introducción al Aprendizaje No Supervizado	3
1.1. Introducción	3
1.2. Categorización de Modelos de Aprendizaje No Supervisado	3
2. Ejemplos de modelos clásicos de agrupamiento	4
2.1. K-medias	4
3. Ejemplos de modelos clásicos de reducción de la dimensionalidad	5
3.1. Análisis de Componentes Principales (PCA)	5
3.2. Tasseled Cap	5
3.3. t-SNE	6
4. Ejemplos avanzados	6
4.1. Auto-encoder	6
4.2. Auto-encoder variacional	6
5. Ejercicio Práctico	6
5.1. Ejecución	9

Listings

Índice de cuadros

Lista de figuras

1. Ejemplo del algoritmo k-medias en un conjunto de datos de dígitos escritos a mano . . .	4
2. Ejemplo del algoritmo PCA en un conjunto de datos de dígitos Iris	5
3. Ejemplo del algoritmo Tasseled Cap	6
4. Ejemplo del algoritmo Tasseled Cap	7
5. Ejemplo del algoritmo t-SNE en una comparativa de modelos de DL	8
6. Ejemplo de la arquitectura de un auto-encoder	8

7. Ejemplo de la arquitectura de un auto-encoder variacional	9
--	---

1. Introducción al Aprendizaje No Supervizado

1.1. Introducción

Como vimos en la clase pasada, cuando disponemos de conjuntos de datos etiquetados, los modelos de *machine learning* (ML en adelante) y de *deep learning* (DL en adelante) tienen mucho que aportar. Sin embargo, en nuestro mundo digital moderno, la existencia de dichas etiquetas no puede darse por sentado, existen casos en donde podemos realizar un etiquetado manual, pero también existen otras opciones y estas son objeto del presente documento.

1.2. Categorización de Modelos de Aprendizaje No Supervisado

Es necesario mencionar que en esta sesión solamente mencionaremos dos categorías de algoritmos no supervisados (1.2.1 y 1.2.2). Otras categorías que podríamos denominar híbridas (como lo son por ejemplo: Modelos de tipo Generativo (*Generative adversarial Networks*, algunos algoritmos de optimización y control, como lo son los algoritmos genéticos, algoritmos de aprendizaje por refuerzo como lo son las cadenas de markov, o alternativas contemporáneas como aprendizaje semi-supervisado o aprendizaje débilmente supervisado)) serán dejadas por fuera.

1.2.1. Agrupamiento (Clustering)

La primera sub-categoría clásica es la de los algoritmos de agrupamiento, esta clase de algoritmos, como su nombre lo indica, intentan construir categorías, en base a la similitud entre los puntos de un conjunto de datos dado.

Este tipo de algoritmos son importantes, a la hora de explorar los conjuntos de datos a los que nos podemos ver enfrentados, y su exponente clásico más conocido es el de k-medias 2.1

1.2.2. Reducción de la dimensionalidad

La segunda sub-categoría se refiere a los algoritmos de reducción de la dimensionalidad, una forma mas intuitiva de ver este tipo de algoritmos es como una especie de algoritmos de compresión y su principal interés es la de superar la llamada «*maldición de la dimensionalidad*»

Ahora bien, ¿Qué es la llamada «*maldición de la dimensionalidad*»?

Esta se refiere a que un sistema es teóricamente imposible de resolver, cuando el numero de características supera a el numero de muestras. En estos casos, aplicar un algoritmo de reducción de la dimensionalidad es especialmente útil, por que nos permite, con una perdida «mínima» de información, aplicar nuestros algoritmos clásicos de aprendizaje supervisado.

Sin embargo, este no es el único caso en que los algoritmos de reducción de la dimensionalidad resultan utiles, a veces y cada vez mas comúnmente, este tipo de representaciones en un espacio dimensional mas bajo, resulta especialmente útiles para hacer el proceso de aprendizaje mejor.

2. Ejemplos de modelos clásicos de agrupamiento

2.1. K-medias

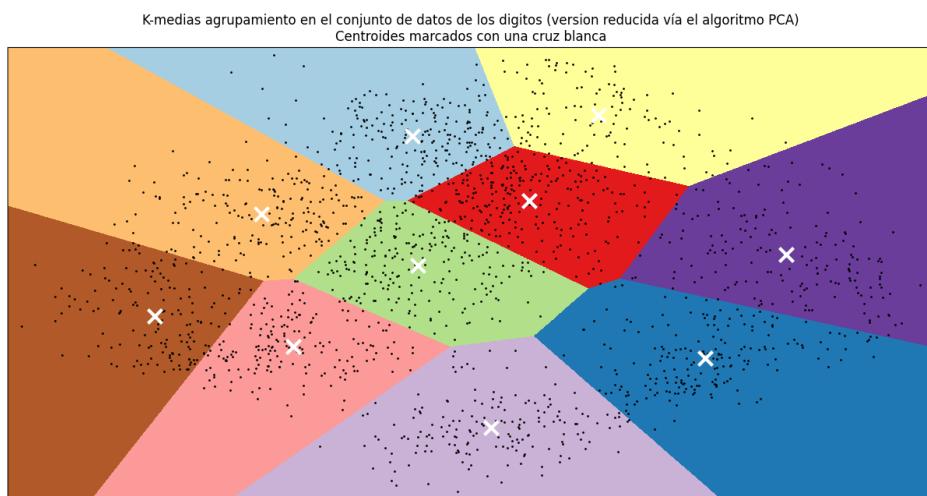


Figura 1: Ejemplo del algoritmo k-medias en un conjunto de datos de dígitos escritos a mano

La anterior imagen ha sido modificada del ejemplo ¹

¹https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_digits.html

3. Ejemplos de modelos clásicos de reducción de la dimensionalidad

3.1. Análisis de Componentes Principales (PCA)

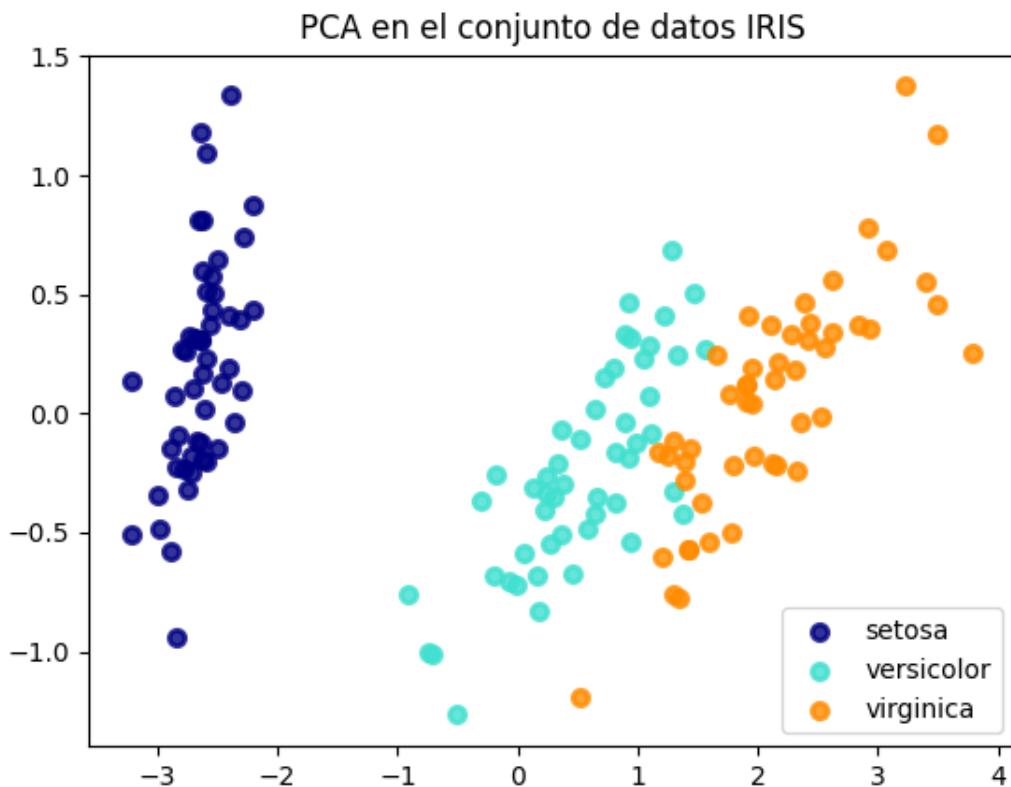


Figura 2: Ejemplo del algoritmo PCA en un conjunto de datos de dígitos Iris

La anterior imagen ha sido modificada del ejemplo ²

3.2. Tasseled Cap

La anterior imagen ha sido tomada del paper titulado «*Partitioning of spatial heterogeneity in an object-oriented riparian boundaries classification system for a South African savanna*» [1]

La anterior imagen ha sido tomada del siguiente blog³

²https://scikit-learn.org/stable/auto_examples/decomposition/plot_pca_vs_lda.html#sphx-glr-auto-examples-decomposition-plot-pca-vs-lda-py

³<https://pro.arcgis.com/es/pro-app/latest/help/analysis/raster-functions/tasseled-cap-function.htm>

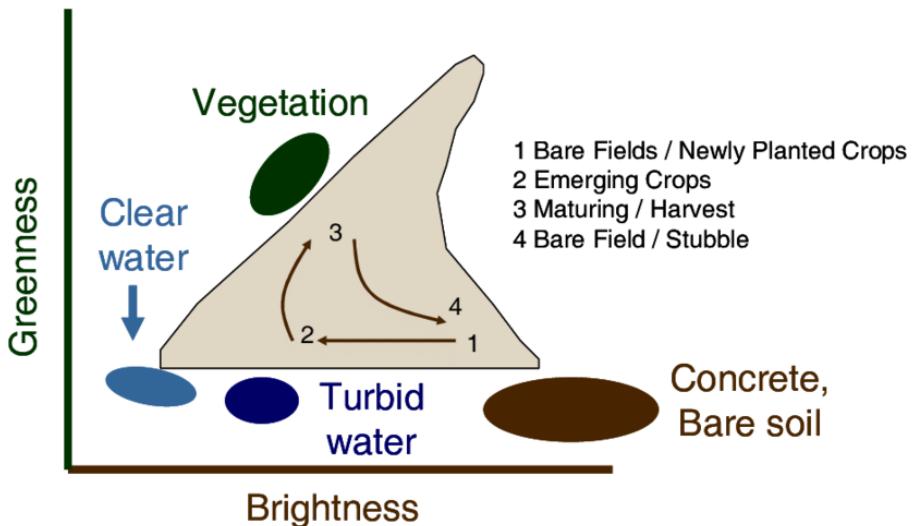


Figura 3: Ejemplo del algoritmo Tasseled Cap

3.3. t-SNE

La anterior imagen ha sido tomada del paper titulado «*A two-stage adaptation network (TSAN) for remote sensing scene classification in single-source-mixed-multiple-target domain adaptation (S\$^2 M^2 \\$T DA) scenarios*» [2]

4. Ejemplos avanzados

4.1. Auto-encoder

La anterior imagen ha sido tomada del siguiente blog⁴

4.2. Auto-encoder variacional

La anterior imagen ha sido tomada del siguiente blog⁵

5. Ejercicio Practico

Nuestro ejercicio practico seguira el tutorial de «Smart Land Use Reconstruction Pipeline»⁶

⁴<https://www.geeksforgeeks.org/numpy/types-of-autoencoders/>

⁵<https://theaisummer.com/Autoencoder/>

⁶<https://github.com/CNES/slurp>

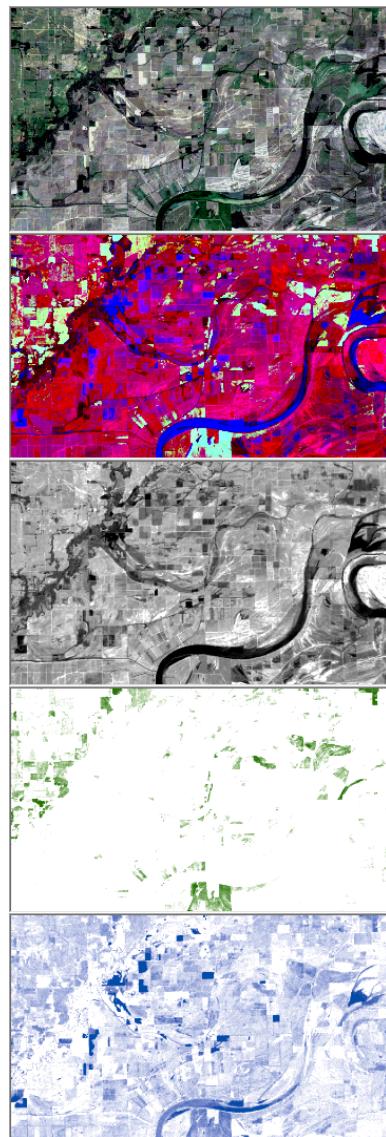


Figura 4: Ejemplo del algoritmo Tasseled Cap

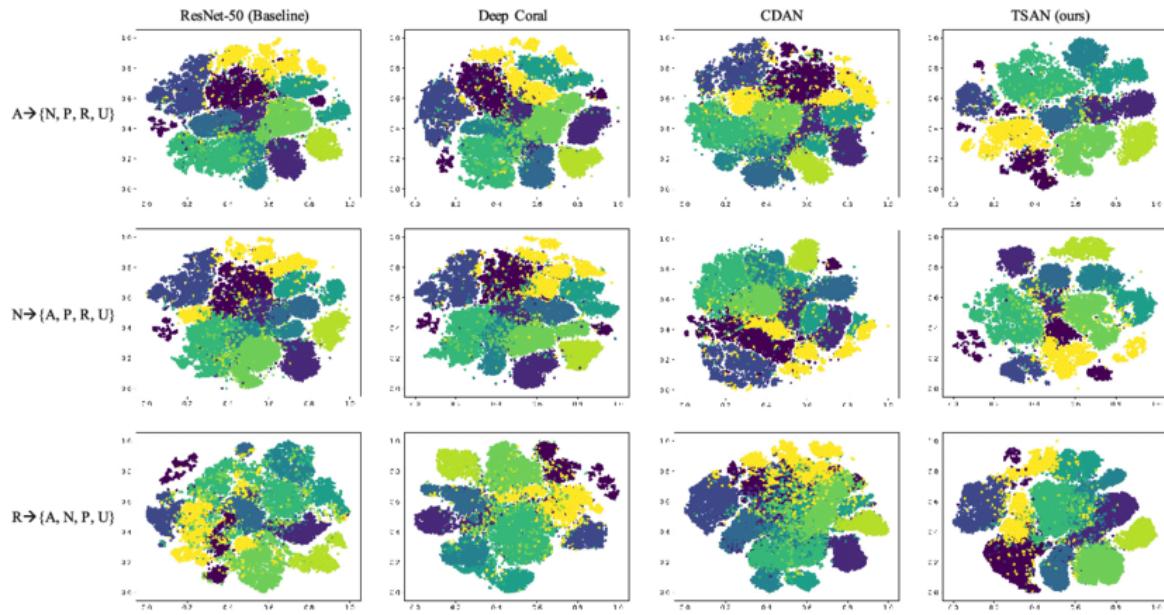


Figura 5: Ejemplo del algoritmo t-SNE en una comparativa de modelos de DL

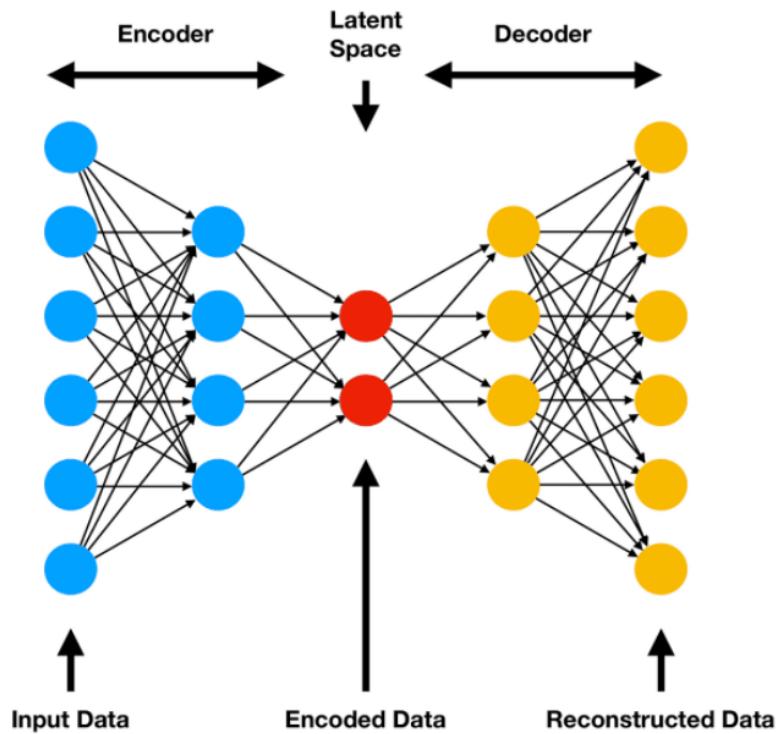


Figura 6: Ejemplo de la arquitectura de un auto-encoder

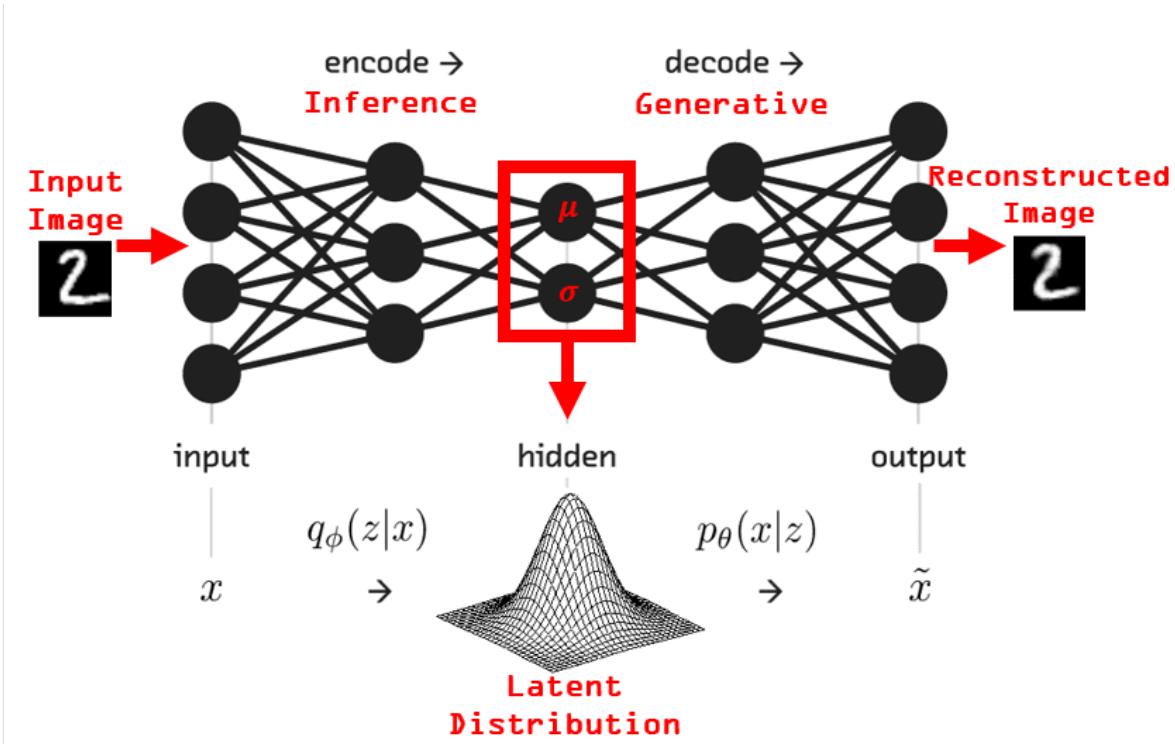


Figura 7: Ejemplo de la arquitectura de un auto-encoder variacional

La estrategia de esta cadena de procesamiento, es realizar varias clasificaciones no supervisadas sobre una imagen de muy alta resolución, para derivar unas clases comunes

Conjunto de datos:

-

Install

TODO

5.1. Ejecución

1. Generate the water mask

```
slurp_watermask prepare/effective_used_config.json
```

1. Generate the vegetation mask

```
slurp_vegetationmask prepare/effective_used_config.json
```

1. Generate the shadow mask

```
slurp_shadowmask prepare/effective_used_config.json
```

1. Generate the urban mask

```
slurp_urbanmask prepare/effective_used_config.json
```

1. Stack the masks

```
slurp_stackmasks prepare/effective_used_config.json
```

Referencias

- [1] D. S. Saah, *Partitioning of spatial heterogeneity in an object-oriented riparian boundaries classification system for a South African savanna*. University of California, Berkeley, 2004.
- [2] J. Zheng, W. Wu, S. Yuan et al., “A two-stage adaptation network (TSAN) for remote sensing scene classification in single-source-mixed-multiple-target domain adaptation (S²M²T DA) scenarios,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, págs. 1-13, 2021.