

## La arquitectura codificador-decodificador y los autocodificadores

Las arquitecturas neuronales codificador – decodificador (*encoder-decoder*) han demostrado ser muy efectivas en muchas tareas de machine learning. Constan de dos partes: el codificador y el decodificador (Fig. 1). Veamos cuál es la función de cada uno:

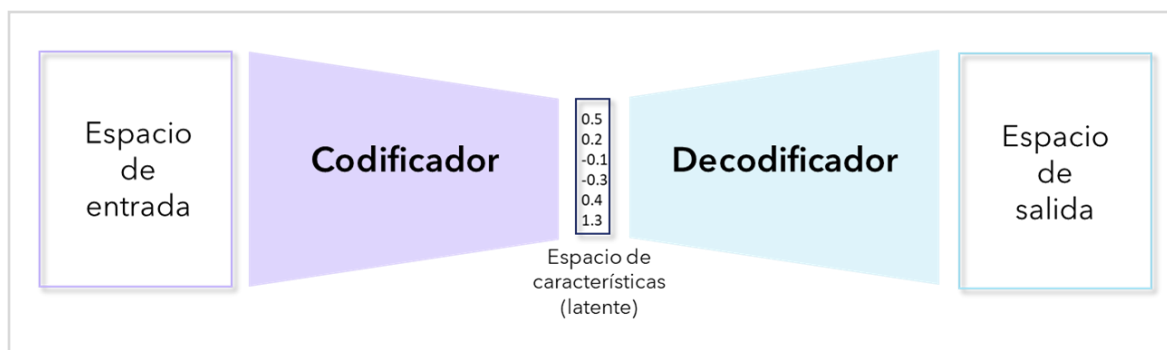


Fig. 1. Arquitectura codificador – decodificador.

- Codificador (encoder). Toma una entrada, que puede ser una secuencia de palabras, una imagen o cualquier otro tipo de objeto, y la convierte en una representación vectorial de longitud fija o embedding en un espacio de características. Esta nueva representación, definida por un conjunto de particularidades “latentes”, captura información semántica y contextual presente en la entrada.
- Decoder (decodificador). Recibe como entrada el embedding generado por el encoder y genera una salida coherente para esta representación vectorial.

Estas arquitecturas son utilizadas en una variedad de aplicaciones, desde la traducción automática, la generación de texto y el resumen de documentos hasta la conversión de voz a texto y la generación de imágenes, entre otras. La naturaleza, tanto del codificador como el decodificador, dependerá de la tarea a resolver. Veamos algunos ejemplos en detalle:

**Descripción de imágenes.** Un descriptor (titulador) de imágenes utiliza la parte del encoder para extraer características visuales relevantes de una imagen, generando una representación vectorial que representa una codificación de la imagen en un espacio de características de dimensionalidad reducida. El decoder entonces, a partir de este embedding, genera una frase que describe la imagen (Fig. 2). Es posible también generar descripciones a partir de videos. El modelo se entrena con pares (imagen, descripción).

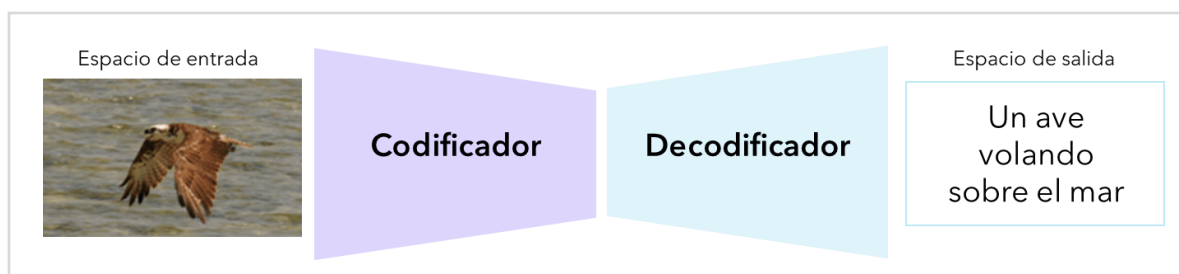


Fig. 2. Arquitectura codificador - decodificador para la titulación de imágenes. El codificador puede ser una red neuronal convolucional y el decodificador una red recurrente o Transformer.

**Segmentación semántica.** La segmentación semántica consiste en asignar una etiqueta de clase a cada píxel de una imagen para identificar los tipos de objetos presentes en esta. Es decir, divide una imagen en regiones significativas y asigna a cada región una etiqueta que indica a qué clase pertenece. La parte del encoder de la arquitectura toma la imagen de entrada y la procesa generando características semánticas que recogen información de contexto. El decoder toma la representación codificada y la utiliza para producir una salida que tenga la misma resolución espacial que la imagen original, asignando una etiqueta semántica a cada píxel de la imagen. El modelo se entrena con pares (imagen, máscara de etiquetas).

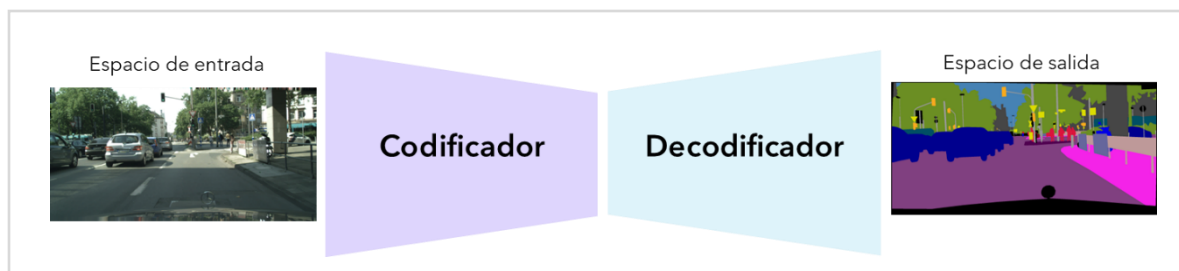


Fig. 3. Arquitectura codificador - decodificador para segmentación semántica. Ambos componentes pueden ser redes neuronales convolucionales.

**Traducción automática.** La arquitectura encoder-decoder ha demostrado ser muy efectiva para realizar traducción automática. En esta, el encoder toma como entrada una secuencia de palabras en el idioma de origen y genera un contexto latente, el cual encapsula el significado de la oración en este idioma. El decoder toma este vector de contexto latente y lo utiliza para generar una secuencia de palabras en el idioma de destino (Fig. 4). El modelo es entrenado con pares (texto-idioma origen, texto-idioma destino).

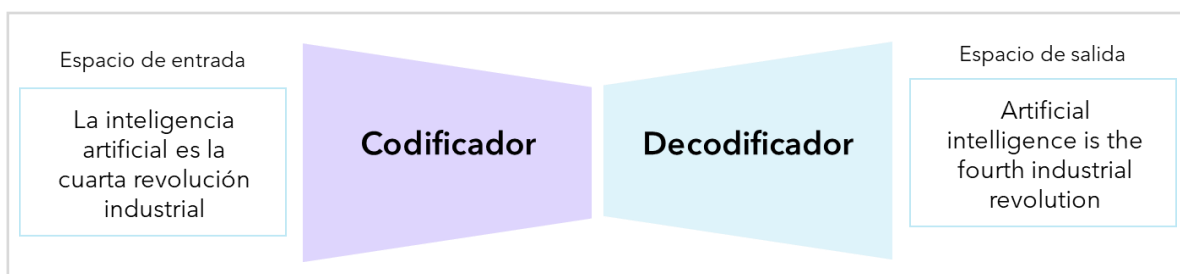


Fig. 4. Arquitectura codificador - decodificador para traducción. Ambos componentes pueden ser redes neuronales recurrentes.

## El autocodificador

Cuando en la arquitectura codificador - decodificador el espacio de entrada es el mismo para la salida, se tiene un autocodificador (*autoencoder*). Es un tipo de red neuronal utilizada para aprender representaciones eficientes de los datos, sin necesidad de etiquetas (Fig. 5). Funciona de la siguiente manera:

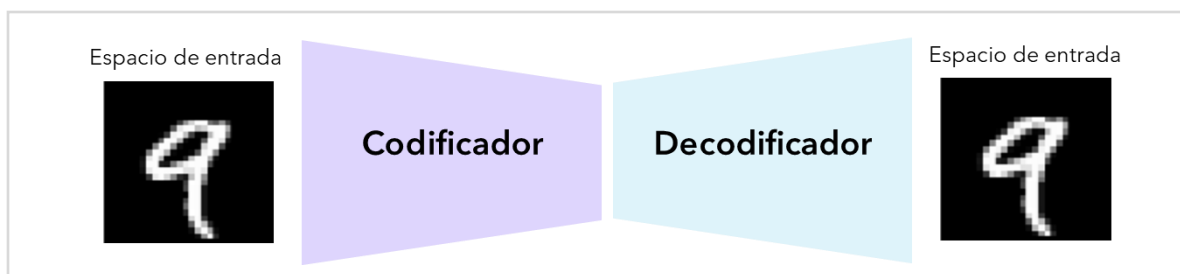


Fig. 5. Arquitectura de un autocodificador.

- **Codificador.** En la fase de codificación, la red neuronal toma la entrada y la comprime en una representación de dimensionalidad inferior, también conocida como "código". El encoder reduce la dimensionalidad de los datos al pasar por capas ocultas, donde cada capa extrae características relevantes de la entrada. El objetivo de esta fase es capturar las particularidades más importantes de los datos de entrada en una forma compacta.
- **Decodificador.** En la fase de decodificación, la representación codificada (embedding) se utiliza para reconstruir la entrada original. El decodificador toma este vector y la "descomprime" en una forma que se asemeje lo más posible a la entrada original. Al igual que el codificador, el decoder consiste en capas ocultas que, en este caso, expanden la representación codificada de nuevo a la dimensión original.

Durante el entrenamiento del autocodificador se utiliza una función de costo que mide la discrepancia entre la entrada original y la reconstrucción generada por el decoder. El objetivo del autocodificador es minimizar esta función de costo, lo que lleva a que la representación codificada capture las características más importantes de los datos de entrada de manera que la reconstrucción sea lo más precisa posible.

Una vez que el autocodificador ha sido entrenado, la representación aprendida en la capa codificada puede ser utilizada para diferentes propósitos, como reducción de dimensionalidad, visualización de datos, eliminación de ruido y detección de anomalías. Además, el autocodificador puede ser empleado como base para la generación de datos similares a los de entrenamiento cuando se entrena como modelo generativo.

## Bibliografía

Foster, D. (2023). Generative Deep Learning. Capítulo 3: Variational autoencoders. O'Reilly Media, Inc.

Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition*. O'Reilly Media, Inc.

---

© - **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.

---