
Universidad
Tecmilenio®

Universidad Tecmilenio

# Inteligencia artificial a través del aprendizaje profundo

Aprendizaje profundo a la resolución de tareas de inteligencia artificial

Aprendizaje generativo

# Introducción

Las redes neuronales se utilizan ampliamente en clasificación o etiquetado de imágenes, detección de señales y traducción de textos, ya sea que se trate de la detección de una señal, biometría falsa o algún tipo de predicción o pronóstico, es posible realizarlo con aprendizaje profundo. Las aplicaciones son variadas y se ubican en cualquiera de estos tres dominios: imágenes, señales o lenguaje.

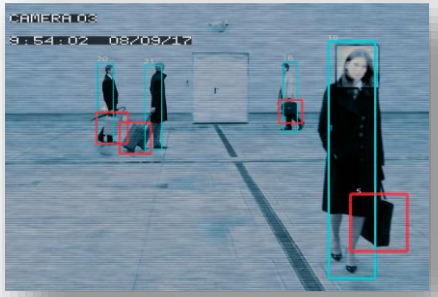
Para lograrlo, dentro del aprendizaje automático existen diferentes tipos de aprendizaje: supervisado, no supervisado, por refuerzo y semi supervisado. El aprendizaje supervisado requiere de un conjunto de datos etiquetados o categorizados para indicarle al modelo lo que se desea que aprenda. En el aprendizaje no supervisado los datos no están etiquetados, por lo que los algoritmos los analizan para obtener nuevo conocimiento o agrupar elementos en función de alguna característica.

Compara, por ejemplo, el funcionamiento del algoritmo que sugiere recomendaciones de contenido digital en Netflix o en Amazon contra el algoritmo de Alexa o Siri, cuando se trata de tener alguna conversación con alguno de estos agentes. ¿Notas la diferencia?

# Explicación

## Clasificación de imágenes

Google Imágenes se implementó en el año 2001 y es una aplicación especializada del buscador principal de Google. El tipo de tarea que resuelve esta herramienta se conoce como **clasificación o etiquetado de imágenes**.




La clasificación de imágenes es un proceso en el que se toman datos de entrada (por ejemplo, una imagen de una manzana) y se genera una categoría o clase como salida ("manzana") o la probabilidad de que la entrada corresponde a una categoría o clase en particular ("existe un 90% de probabilidad que la imagen corresponda a una manzana").

Para resolver este tipo de problemas, generalmente se utilizan redes neuronales convolucionales (CNN, por sus siglas en inglés) y redes neuronales multicapa con retro propagación, aunque no son las únicas.

# Explicación

El proceso de clasificación de imágenes se puede descomponer en las siguientes etapas:

**Preprocesamiento de imagen**: el objetivo de esta etapa es mejorar las características de la imagen mediante la supresión de distorsiones no deseadas y la mejora de algunas características relevantes.

**Detección de objetos**: la detección se refiere a la ubicación de objetos, lo que implica segmentar la imagen e identificar la posición del objeto de interés.

**Extracción de características y entrenamiento**: se identifican los patrones más interesantes de la imagen o características que pueden ser exclusivas de una clase o categoría en particular y que pueden ayudar al modelo a diferenciar entre clases.

**Clasificación del objeto**: se categorizan los objetos detectados en clases predefinidas.

Una red neuronal convolucional procesa la información en sus capas, identificando características que le permiten reconocer diferentes elementos.

| | |
|---|---|
| **Capa de convolución.** | Aplica filtros a la imagen de entrada con el fin de capturar información local y a detalle, obteniendo una extracción de características de la imagen de entrada. |
| **Capa de agrupación.** | Sustituye áreas específicas de la imagen para reducir las dimensiones de las características de la misma. Este proceso de agrupación posibilita la detección de objetos independientemente de su ubicación. |
| **Capa ReLu.** | El objetivo de esta capa es introducir no linealidades para manejar información "complicada" de mejor manera. |
| **Capa completamente conectada.** | Es la capa que funciona como clasificador y mapea la representación de los atributos aprendidos al espacio de categorías o clases. |

# Explicación

A continuación, se presenta el proceso de clasificación de imágenes utilizando una **red neuronal convolucional:**

- Se ingresa una imagen en la entrada de la red.
- Se aplican diferentes filtros para obtener un mapa de características de la imagen.
- Se aplican funciones ReLu para incrementar las no linealidades.
- Se aplica una capa de agrupamiento a cada vector de características.
- Se unen las subimágenes obtenidas en un vector grande.
- El vector se usa como entrada de una red neuronal completamente conectada.
- Se procesan las características en la red, la última capa de esta da el "voto" de la clase o categoría buscada.
- Se entrena por medio de propagación y retro propagación durante muchas épocas hasta obtener una red neuronal bien definida.

# Explicación

**Detección de objetos y segmentación**

El **reconocimiento de objetos** es un término que se utiliza para describir a un conjunto de tareas en visión computacional, cuyo objetivo es identificar objetos en las imágenes o fotografías digitales. Las tareas son la clasificación de imágenes, la localización de objetos y la detección de objetos.

Mientras que la clasificación de imágenes es una tarea que implica asignar una etiqueta de clase a una imagen; la localización de objetos trata de identificar la ubicación de objetos en una imagen, dibujando un recuadro alrededor de cada uno de ellos. Actualmente, los modelos más utilizados en la detección de objetos son YOLO (You Only Look Once) y SSD (Single Shot Object Detectors).

**Redes neuronales convolucionales con base en regiones**

Girshick (2015), desarrolló una familia de modelos para dar solución a las tareas de localización y detección de objetos al que llamó **R-CNN.** Este modelo propuesto se compone de tres módulos:

**Propuesta de regiones**: genera y extrae propuestas de regiones independientemente de la categoría, por ejemplo, recuadros delimitadores candidatos.

**Extractor de características**: se extraen características de cada región candidata utilizando una red neuronal convolucional.

**Clasificador**: clasifica las características en alguna de las clases conocidas con un clasificador SVM lineal.

# Explicación

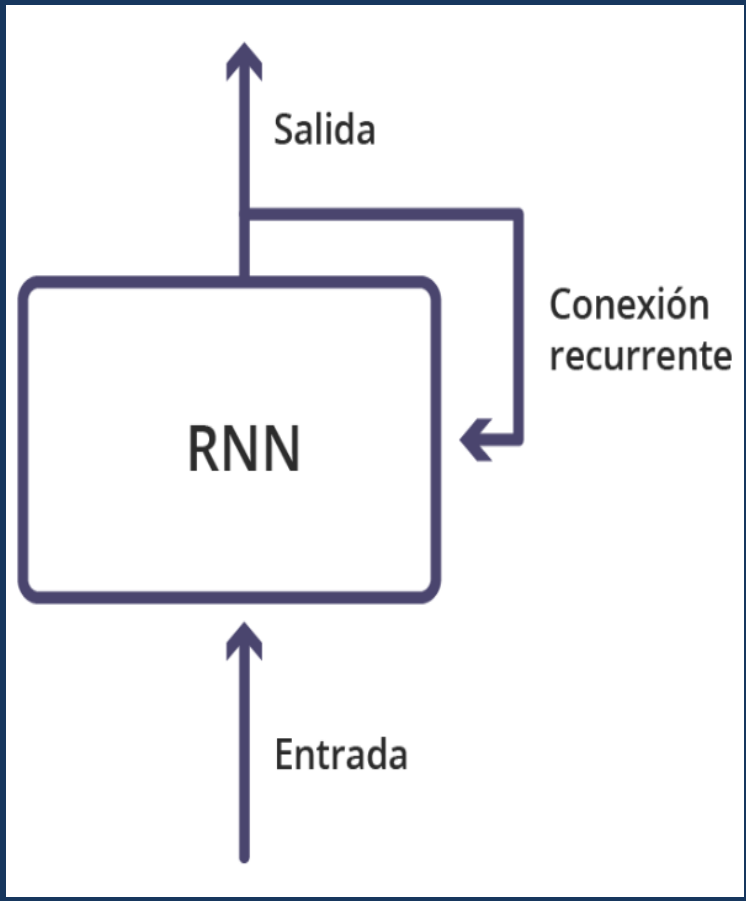
**Texto y secuencias.**

La clasificación y la categorización de textos son tareas esenciales en la búsqueda de documentos, filtrado, búsquedas en Internet, identificación de idiomas y análisis de sentimientos.

Una red neuronal recurrente (RNN, por sus siglas en inglés) es un tipo de red adecuada para el reconocimiento de patrones en secuencias de datos como texto, video, habla, lenguaje, genomas y series de tiempo.

Cuando un humano lee un libro, lo hace procesando palabra por palabra, almacenando temporalmente en su memoria información que ha leído y eso le permite obtener una representación fluida del significado de las oraciones que lee. Entonces, la inteligencia biológica procesa la información de forma incremental, mientras almacena un modelo interno de lo que está procesando, construido con información del "pasado" que se actualiza en la medida que llega información nueva al modelo.

Las RNN adoptan el principio descrito previamente en una versión más simple, procesan una secuencia iterando a través de los elementos de la secuencia y manteniendo un estado que contiene información relativa a lo que ha visto hasta el momento presente. En la figura 1 se muestra un diagrama conceptual de la idea descrita.



## Autocodificación

Los modelos en el aprendizaje no supervisado pueden categorizarse en dos tipos: discriminativos y generativos. Cada uno de estos modelos tiene un principio de operación distinto.

**Ejemplo.**

Juan es padre de dos niños, el niño A, cuya característica principal es aprender todo a detalle, y el niño B, quien sólo aprende a partir de las diferencias de lo que ve. Un día, Juan decide llevar a sus hijos a un zoológico que tiene sólo dos tipos de animales: caballos y elefantes.

Después de salir del zoológico, se topan con un animal en la calle y Juan les pregunta a sus hijos si el animal visto es un caballo o un elefante. El niño A toma un pedazo de papel y dibuja una imagen de cada uno, las compara con el animal que está frente a ellos y contesta en función de la similitud entre las imágenes y el animal, y contesta: "es un caballo". El niño B solo conoce las diferencias entre los animales gracias a las distintas propiedades que identificó en ellos, por lo que contestó: "es un caballo".

Ambos niños identificaron correctamente al animal que se encontraron después del zoológico, sin embargo, la manera en la que aprendieron y encontraron la respuesta es muy distinta; el niño A es un ejemplo de modelo **generativo** y el niño B de un modelo **discriminativo**.

## Modelo generativo

Un modelo **discriminativo** hace predicciones sobre instancias de datos no conocidos con base en la probabilidad condicional y puede utilizarse en tareas de clasificación y regresión. Un modelo **generativo** se enfoca en la distribución de un conjunto de datos para devolver la probabilidad de un ejemplo dado.

Por lo tanto, un modelo generativo está constituido por algoritmos de aprendizaje, cuyo objetivo es aprender la distribución de los datos del conjunto de entrenamiento y es una arquitectura de aprendizaje profundo, así que utiliza redes neuronales que intentan modelar una distribución de datos lo más parecida posible a una real.

Los autocodificadores son un tipo de red neuronal multicapa completamente conectada que se entrena para copiar la entrada a su salida. Comprimen la entrada en una codificación de menor dimensión para después reconstruirla en la salida, a partir de esta representación codificada. La codificación es un resumen o representación compacta de los datos de entrada y se conoce como **espacio latente**. La compresión de datos, entonces, es un proceso de codificación.

<u>Un autocodificador tiene tres componentes:</u>

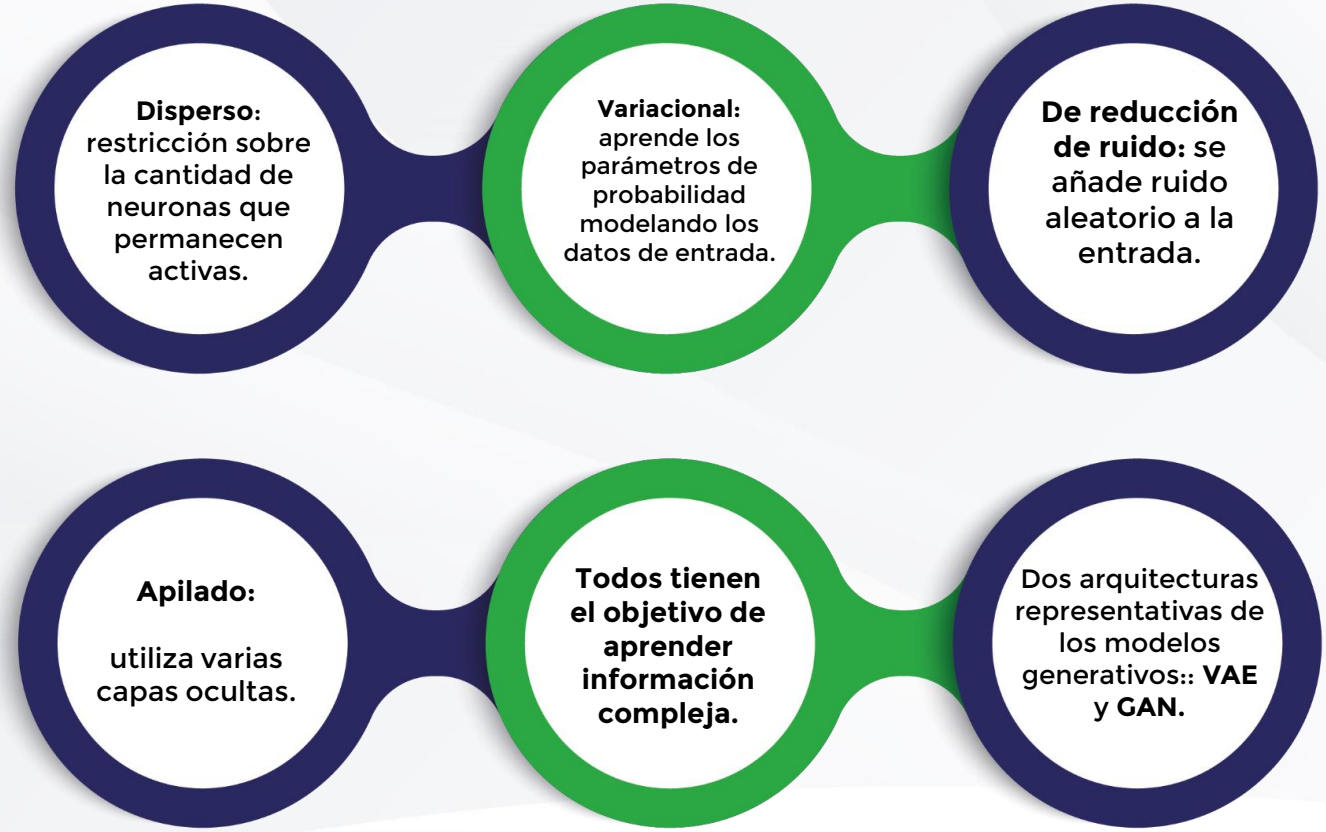
1. Un codificador: comprime la entrada y produce un código, es decir, produce nuevos atributos a partir de los viejos atributos.

2. Un código.

3. Un decodificador: reconstruye la entrada utilizando solamente el código.

# Explicación

Hay diferentes tipos de **autocodificadores** y son específicos para ciertos tipos de tareas:

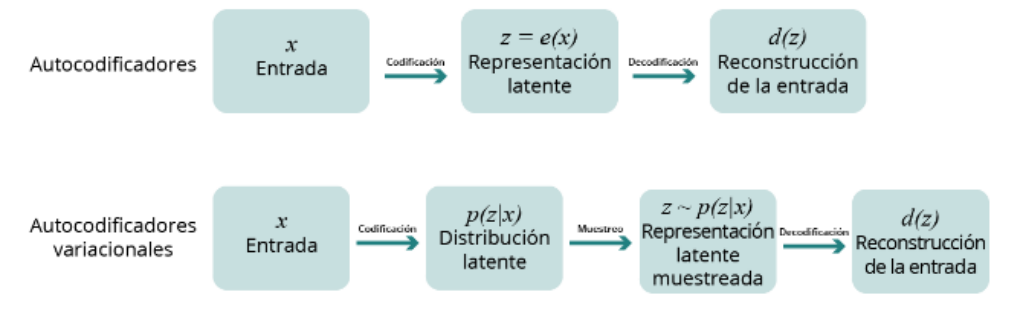


# Explicación

## Autocodificación variacional

Un VAE es un autocodificador, cuya distribución de códigos se regulariza durante la etapa de entrenamiento para garantizar que su espacio latente adquiera buenas propiedades que le permitan generar datos nuevos. Este tipo de arquitectura usa las redes neuronales con distribuciones de probabilidad.

Su uso va desde la generación de imágenes hasta la generación de audio. En el campo de las imágenes, por ejemplo, el VAE trata de maximizar la semejanza entre las imágenes que ha generado y las imágenes con las que se entrenó.




## Red generativa antagónica (GAN)

tipo de estructuras en aprendizaje automático que se forman por dos tipos de modelos entrenados de forma simultánea: el generador, cuyo objetivo es desarrollar datos falsos, y el discriminador, entrenado para diferenciar entre datos falsos y reales. Este tipo de estructura es capaz de producir o generar información o contenido nuevo.

# Actividad

**¿Cuál es la diferencia entre una red neuronal convolucional y una recurrente?**

**Enlista el proceso de clasificación de imágenes**

**¿Cuál es la diferencia entre los modelos discriminativos y los generativos?**

**¿A qué se le llama autocodificación dentro del aprendizaje profundo?**

# Cierre

El aprendizaje profundo (AP) es un área del aprendizaje automático donde los modelos son largas cadenas de funciones geométricas aplicadas una tras otra. Estas operaciones están estructuradas en módulos conocidos como capas, y los modelos de AP se pueden ver como una pila de capas o un grafo de capas.

Por otro lado, los autocodificadores aprenden cómo comprimir datos con base en ciertos atributos descubiertos en estos, por ejemplo, correlaciones en los datos de entrada durante la etapa de entrenamiento, por ello son capaces de reconstruir datos similares a los observados durante su entrenamiento; y las GAN tienen una gran habilidad para modelar datos de alta dimensión y tratar con problemas en donde existe información incompleta y generar salidas multimodales.

# Bibliografía


- Girshick, R. (2015). *Fast R-CNN. IEEE International Conference on Computer Vision (ICCV).* Recuperado de https://ieeexplore.ieee.org/document/7410526

- Google Developers. (s.f.*). Overview of GAN Structure.* Recuperado de https://developers.google.com/machine-learning/gan/gan_structure