



**AWAKELAB**

**BASECAMP**

Ciencia de Datos

## Módulo: Fundamentos de deep learning.

---

### Aprendizaje Esperado

---

1. Describir los conceptos fundamentales las redes neuronales y su utilidad para la resolución de problemas de aprendizaje de máquina
- 

### Deep Learning

*El aprendizaje profundo (deep learning) intenta imitar el cerebro humano, aunque lejos de igualar su capacidad, lo que permite que los sistemas agrupen datos y hagan predicciones con una precisión increíble.*

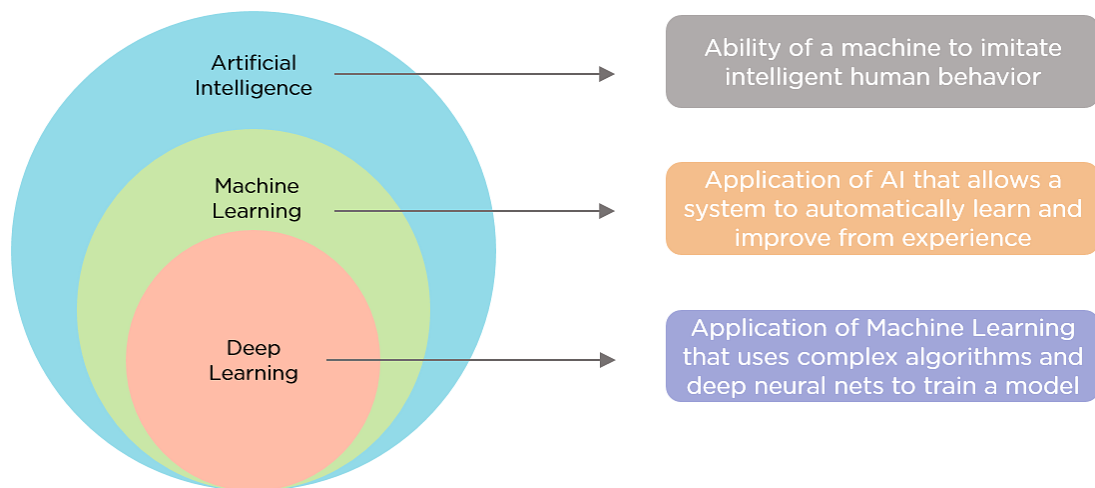
### ¿Qué es el deep learning?

El deep learning es un subconjunto del aprendizaje automático , que es esencialmente una red neuronal con tres o más capas. Si bien una red neuronal con una sola capa aún puede hacer predicciones aproximadas, las capas ocultas adicionales pueden ayudar a optimizar y refinar la precisión.

El aprendizaje profundo impulsa muchas aplicaciones y servicios de inteligencia artificial (IA) que mejoran la automatización, realizando tareas analíticas y físicas sin intervención humana. La tecnología de aprendizaje profundo se encuentra detrás de los productos y servicios cotidianos (como los asistentes digitales, los controles remotos de TV habilitados para voz y la detección de fraudes con tarjetas de crédito), así como de las tecnologías emergentes (como los automóviles autónomos).

## Diferencias con la Inteligencia Artificial, Machine Learning y Deep Learning

Aquí hay una ilustración diseñada para ayudarnos a comprender las diferencias fundamentales entre la inteligencia artificial, el aprendizaje automático y el aprendizaje profundo.



La inteligencia artificial es el concepto de crear máquinas inteligentes inteligentes.

Machine Learning es un subconjunto de inteligencia artificial que ayuda a crear aplicaciones impulsadas por IA.

El aprendizaje profundo o deep learning es un subconjunto del aprendizaje automático que utiliza grandes volúmenes de datos y algoritmos complejos para entrenar un modelo.

**¿Qué resolver con aprendizaje profundo?**

Las aplicaciones de aprendizaje profundo del mundo real son parte de nuestra vida diaria, pero en la mayoría de los casos, están tan bien integradas en los productos y servicios que los usuarios no son conscientes del complejo procesamiento de datos que se lleva a cabo en segundo plano. Algunos de estos ejemplos incluyen lo siguiente:

- **Cumplimiento de la ley**

Los algoritmos de aprendizaje profundo pueden analizar y aprender de los datos transaccionales para identificar patrones peligrosos que indiquen una posible actividad fraudulenta o delictiva. El reconocimiento de voz, la visión por computadora y otras aplicaciones de aprendizaje profundo pueden mejorar la eficiencia y la eficacia del análisis de investigación al extraer patrones y evidencia de grabaciones de sonido y video, imágenes y documentos, lo que ayuda a las fuerzas del orden público a analizar grandes cantidades de datos de manera más rápida y precisa.

- **Servicios financieros**

Las instituciones financieras utilizan regularmente análisis predictivos para impulsar el comercio algorítmico de acciones, evaluar los riesgos comerciales para la aprobación de préstamos, detectar fraudes y ayudar a administrar las carteras de crédito e inversión de los clientes.

- **Servicio al Cliente**

Muchas organizaciones incorporan tecnología de aprendizaje profundo en sus procesos de servicio al cliente. Los chatbots, utilizados en una variedad de aplicaciones, servicios y portales de atención al cliente, son una forma sencilla de IA. Los chatbots tradicionales usan lenguaje natural e incluso reconocimiento visual, que se encuentran comúnmente en los menús de los centros de llamadas. Sin embargo, las soluciones de chatbot más sofisticadas intentan determinar, a través del aprendizaje, si hay múltiples respuestas a preguntas ambiguas. Según las respuestas que

recibe, el chatbot intenta responder estas preguntas directamente o enrutar la conversación a un usuario humano.

Los asistentes virtuales como Siri de Apple, Amazon Alexa o Google Assistant amplían la idea de un chatbot al habilitar la funcionalidad de reconocimiento de voz. Esto crea un nuevo método para involucrar a los usuarios de forma personalizada.

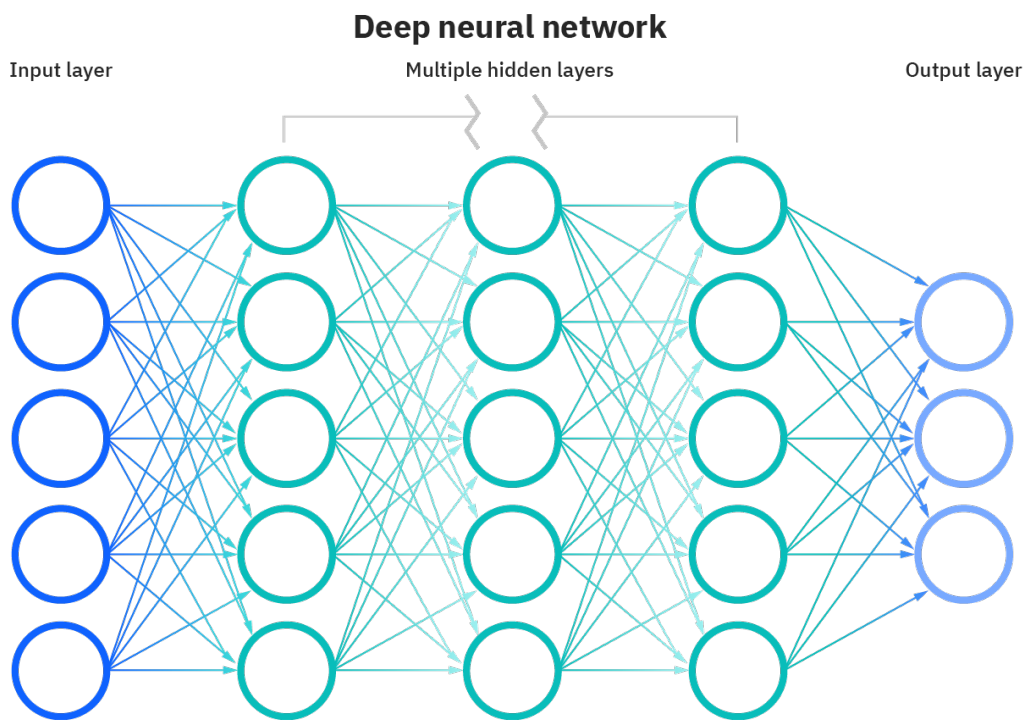
- **Cuidado de la salud**

La industria de la salud se ha beneficiado enormemente de las capacidades de aprendizaje profundo desde la digitalización de los registros e imágenes de los hospitales. Las aplicaciones de reconocimiento de imágenes pueden ayudar a los radiólogos y especialistas en imágenes médicas, ayudándolos a analizar y evaluar más imágenes en menos tiempo.

### **¿Qué son las redes neuronales artificiales?**

Las redes neuronales, también conocidas como redes neuronales artificiales (ANN) o redes neuronales simuladas (SNN), son un subconjunto del aprendizaje automático y están en el corazón de los algoritmos de deep learning. Su nombre y estructura están inspirados en el cerebro humano, imitando la forma en que las neuronas biológicas envían señales entre sí.

Las redes neuronales artificiales (ANN) se componen de capas de nodos, que contienen una capa de entrada, una o más capas ocultas y una capa de salida. Cada nodo, o neurona artificial, se conecta con otro y tiene asociado un peso y un umbral. Si la salida de cualquier nodo individual está por encima del valor de umbral especificado, ese nodo se activa y envía datos a la siguiente capa de la red. De lo contrario, no se pasan datos a la siguiente capa de la red.



Las redes neuronales se basan en datos de entrenamiento para aprender y mejorar su precisión con el tiempo. Sin embargo, una vez que estos algoritmos de aprendizaje se ajustan con precisión, se convierten en herramientas poderosas en informática e inteligencia artificial, lo que nos permite clasificar y agrupar datos a alta velocidad. Las tareas de reconocimiento de voz o reconocimiento de imágenes pueden llevar minutos en lugar de horas en comparación con la identificación manual por parte de expertos humanos. Una de las redes neuronales más conocidas es el algoritmo de búsqueda de Google.

### **Para qué se usa una red neuronal**

Estos sistemas de algoritmia que nos ayudan a resolver problemas tienen múltiples aplicaciones que podemos englobar en:

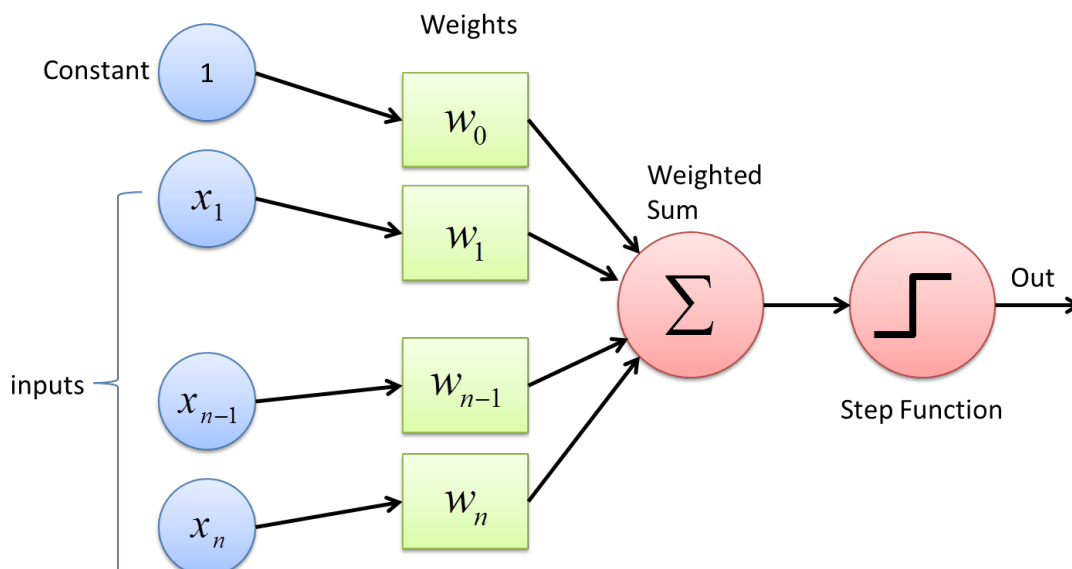
- **Predicción de sucesos y simulaciones:** Producción de los valores de salida esperados en función de los datos entrantes.
- **Reconocimiento y clasificación:** Asociación de patrones y organización de conjuntos de datos en clases predefinidas. Incluso identificando características únicas sin datos previos.

- Procesamiento de datos y modelización: Validación, agregación y análisis de datos. Diseño y búsqueda de fallos en sistemas de software complejos.
- Ingeniería de control: Monitorización de sistemas informáticos y manipulación de robots. Incluida la creación de sistemas y robots autónomos.
- Inteligencia Artificial: Formando parte de las tecnologías de deep learning y machine learning que son partes fundamentales de la inteligencia artificial

## El perceptrón

Un Perceptrón es un algoritmo utilizado para el aprendizaje supervisado de clasificadores binarios. Los clasificadores binarios deciden si una entrada, generalmente representada por una serie de vectores, pertenece a una clase específica.

En resumen, un perceptrón es una red neuronal de una sola capa. Constan de cuatro partes principales que incluyen valores de entrada, pesos y sesgos, suma neta y una función de activación.



El perceptrón está formado por una serie de componentes como:  
 Entrada. Las entradas en el algoritmo del perceptrón se entienden como  $x_1$ ,  $x_2$ ,  $x_3$ ,  $x_4$  y así sucesivamente. Todas estas entradas denotan los

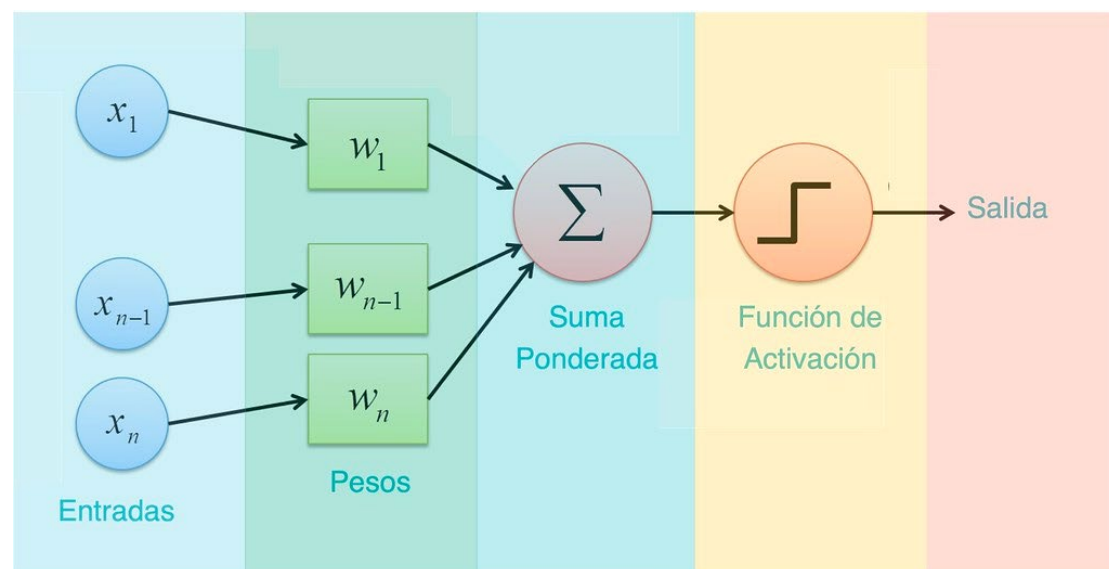
valores del perceptrón de características y la ocurrencia total de las características.

**Pesos.** Se observan como valores que se planifican a lo largo de la sesión de preparación del perceptrón. Los pesos ofrecen un valor preliminar en el inicio del aprendizaje del algoritmo. Con la ocurrencia de cada inexactitud de entrenamiento, los valores de los pesos se actualizan. Estos se representan principalmente como  $w_1$ ,  $w_2$ ,  $w_3$ ,  $w_4$  y así sucesivamente.

**Suma ponderada.** Es la proliferación de cada valor de entrada o característica asociada con el valor de paso correspondiente.

**Salida.** La suma ponderada se pasa a la función de activación y cualquier valor que obtengamos después del cálculo es nuestra salida predicha.

**Función de activación.** Cada función de activación, o no lineal, toma un único número y realiza una determinada operación matemática fija sobre él. Hay varias funciones de activación que se pueden encontrar en la práctica, las más comunes son la Sigmoide o la ReLU o unidad lineal rectificada.



**Topología de una red neuronal**



La topología de una red neuronal se refiere a la forma en que se conectan las neuronas y es un factor importante en el funcionamiento y el aprendizaje de la red. Una topología común en el aprendizaje no supervisado es un mapeo directo de entradas a una colección de unidades que representan categorías (por ejemplo, mapas autoorganizados ).

La topología más común en el aprendizaje supervisado es la red feedforward de tres capas totalmente conectada (ver Backpropagation , Radial Basis Function Networks). Todos los valores de entrada a la red están conectados a todas las neuronas en la capa oculta (ocultos porque no son visibles en la entrada o la salida), las salidas de las neuronas ocultas están conectadas a todas las neuronas en la capa de salida y las activaciones de las neuronas de salida constituyen la salida de toda la red. Tales redes son populares en parte porque teóricamente se sabe que son aproximadores de funciones universales (con, por ejemplo, un sigmoide o gaussiano...).

### Cómo aprende una red neuronal

El algoritmo de gradiente descendente es un **mecanismo de entrenamiento** de sistemas de aprendizaje automático como los basados en redes neuronales. Es el más extendido y utilizado para el aprendizaje o entrenamiento de redes neuronales debido a su sencillez y facilidad de implementación.



## Descenso de gradiente estocástico

**Gradient Descent** es un algoritmo de optimización que encuentra el conjunto de variables de entrada para una función de destino que da como resultado un valor mínimo de la función de destino, llamado el mínimo de la función.

Como sugiere su nombre, el descenso de gradiente implica calcular el gradiente de la función objetivo.

Puede recordar del cálculo que la derivada de primer orden de una función calcula la pendiente o la curvatura de una función en un punto dado. Leída de izquierda a derecha, una derivada positiva sugiere que la función objetivo tiene una pendiente ascendente y una derivada negativa sugiere que la función objetivo tiene una pendiente descendente.

- **Derivada** : *Pendiente o curvatura de una función objetivo con respecto a valores de entrada específicos para la función.*

Si la función de destino toma múltiples variables de entrada, se pueden tomar juntas como un vector de variables. Trabajar con vectores y matrices se denomina álgebra lineal y hacer cálculos con estructuras de álgebra lineal se denomina cálculo matricial o cálculo vectorial. En cálculo vectorial, el vector de derivadas de primer orden (derivadas parciales) generalmente se conoce como el gradiente de la función objetivo.

- **Gradiente** : *Vector de derivadas parciales de una función objetivo con respecto a las variables de entrada.*

El algoritmo de descenso de gradiente requiere el cálculo del gradiente de la función objetivo con respecto a los valores específicos de los valores de entrada. El gradiente apunta hacia arriba, por lo tanto, el negativo del gradiente de cada variable de entrada se sigue hacia abajo para dar como resultado nuevos valores para cada variable que resultan en una evaluación más baja de la función objetivo.

Se utiliza un tamaño de paso para escalar el gradiente y controlar cuánto cambia cada variable de entrada con respecto al gradiente.

- **Step Size** : *Tasa de aprendizaje o alfa, un hiper parámetro utilizado para controlar cuánto cambiar cada variable de entrada con respecto al gradiente.*

Este proceso se repite hasta que se localiza el mínimo de la función objetivo, se evalúa un número máximo de soluciones candidatas o alguna otra condición de parada.

El descenso de gradiente se puede adaptar para minimizar la función de pérdida de un modelo predictivo en un conjunto de datos de entrenamiento, como un modelo de clasificación o regresión. Esta adaptación se llama [descenso de gradiente estocástico](#).

- **Descenso de gradiente estocástico** : *extensión del algoritmo de optimización de descenso de gradiente para minimizar una función de pérdida de un modelo predictivo en un conjunto de datos de entrenamiento.*

La función de destino se toma como la función de pérdida o error en el conjunto de datos, como el error cuadrático medio para la regresión o la entropía cruzada para la clasificación. Los parámetros del modelo se toman como variables de entrada para la función objetivo.

- **Función de pérdida** : *función objetivo que se minimiza.*
- **Parámetros del modelo** : *parámetros de entrada a la función de pérdida que se están optimizando.*

El algoritmo se denomina " estocástico " porque los gradientes de la función objetivo con respecto a las variables de entrada son ruidosos (por ejemplo, una aproximación probabilística). Esto significa que la evaluación del gradiente puede tener ruido estadístico que puede oscurecer la verdadera señal del gradiente subyacente, causado por la escasez y el ruido en el conjunto de datos de entrenamiento.

## Stochastic Gradient Descent With Back-propagation

Stochastic Gradient Descent es un algoritmo de optimización que se puede utilizar para entrenar modelos de redes neuronales. El algoritmo Stochastic Gradient Descent requiere que se calculen gradientes para cada variable en el modelo para que se puedan calcular nuevos valores para las variables.

La retropropagación es un algoritmo de diferenciación automática que se puede utilizar para calcular los gradientes de los parámetros en las redes neuronales.

Juntos, el algoritmo de propagación hacia atrás y el algoritmo de descenso de gradiente estocástico se pueden usar para entrenar una red neuronal. Podríamos llamar a esto "*descenso de gradiente estocástico con propagación hacia atrás*".

- ***Descenso de gradiente estocástico con propagación inversa*** : una descripción más completa del algoritmo general utilizado para entrenar una red neuronal, haciendo referencia al algoritmo de optimización y al algoritmo de cálculo de gradiente.

Es común que los practicantes digan que entrenan su modelo usando retropropagación. Técnicamente, esto es incorrecto. Incluso como abreviatura, esto sería incorrecto. La retropropagación no es un algoritmo de optimización y no se puede usar para entrenar un modelo.

## Referencias

[1] ¿Qué es el Deep Learning?

[https://www.sas.com/es\\_ar/insights/analytics/deep-learning.html](https://www.sas.com/es_ar/insights/analytics/deep-learning.html)

[2] Inteligencia artificial

<https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/ai-vs-machine-learning-vs-deep-learning#:~:text=Artificial%20Intelligence%20is%20the%20concept,algorithms%20to%20train%20a%20model.>

[3] Neural networks

<https://www.ibm.com/cloud/learn/neural-networks>

[4] Aplicaciones de Redes neuronales

<https://www.datacentric.es/blog/insight/red-neuronal-artificial-aplicaciones/>

[5] Perceptron Model

<https://www.analyticssteps.com/blogs/introduction-perceptron-model-machine-learning>

[6] Gradient descent

<https://www.unir.net/ingenieria/revista/gradient-descent/>

[7] Diferencias Gradiente estocástico

<https://machinelearningmastery.com/difference-between-backpropagation-and-stochastic-gradient-descent/>

## Material Complementario

[1] Qué es el Deep Learning

<https://www.youtube.com/watch?v=s0SbvGiG28w>

[2] Diferencias: Machine learning, Deep learning e Inteligencia artificial

<https://www.youtube.com/watch?v=THrgrb-3lsM>