

# Fundamentos de Machine Learning

## Introducción a machine Learning

Machine Learning (ML) es una rama de la inteligencia artificial (IA) que se centra en el desarrollo de algoritmos y técnicas que permiten a las computadoras aprender de los datos y realizar tareas específicas **sin instrucciones explícitas**. Los modelos ML identifican patrones en los datos y hacen predicciones o toman decisiones basadas en esos patrones.

## Aplicaciones de Machine Learning

- **Reconocimiento de Imágenes:** Identificación de objetos, personas o texto en imágenes.
  - *Ejemplo:* Los sistemas de reconocimiento facial utilizados en dispositivos móviles.
- **Procesamiento de Lenguaje Natural (NLP):** Traducción automática, análisis de sentimientos, chatbots.
  - *Ejemplo:* Google Translate y asistentes virtuales como Siri y Alexa.
- **Sistemas de Recomendación:** Sugerencias de productos, películas o música basadas en el comportamiento del usuario.
  - *Ejemplo:* Recomendaciones de películas en Netflix y productos en Amazon.
- **Detección de Fraude:** Identificación de transacciones fraudulentas en tiempo real
  - *Ejemplo:* Sistemas de detección de fraude en transacciones con tarjetas de crédito.
- **Automatización Industrial:** Optimización de procesos de manufactura y mantenimiento predictivo.
  - *Ejemplo:* Robots industriales en líneas de ensamblaje y mantenimiento predictivo en maquinaria pesada.

# Tipos de Aprendizaje

## Aprendizaje Supervisado:

En el aprendizaje supervisado, el modelo se entrena utilizando un conjunto de datos etiquetados, donde cada ejemplo de entrenamiento se compone de una entrada y una salida deseada. El objetivo es aprender una función que mapee las entradas a las salidas con alta precisión.

**Ejemplo:** Clasificación de correos electrónicos como "spam" o "no spam". Cada correo electrónico es una entrada, y la etiqueta ("spam" o "no spam") es la salida deseada.

## Algoritmos comunes:

- **Regresión Lineal:** Predice un valor continuo basado en la relación lineal entre las variables de entrada y la salida.
  - *Ejemplo: Predecir el precio de una casa en función del tamaño, ubicación y número de habitaciones.*
- **Regresión Logística:** Predice una probabilidad que puede usarse para clasificar ejemplos en categorías binarias.
  - *Ejemplo: Clasificar si un paciente tiene o no una enfermedad basándose en sus síntomas.*
- **Máquinas de Soporte Vectorial (SVM):** Encuentra un hiperplano que mejor separa las clases en el espacio de características.
  - *Ejemplo: Clasificar correos electrónicos como "spam" o "no spam".*
- **Árboles de Decisión:** Divide los datos en subconjuntos basados en características que resultan en las mayores ganancias de información.
  - *Ejemplo: Decidir si conceder un préstamo basándose en el historial crediticio, ingresos y otros factores.*
- **K-Vecinos Más Cercanos (k-NN):** Clasifica ejemplos basándose en la clase mayoritaria de sus vecinos más cercanos.
  - *Ejemplo: Clasificar una nueva imagen de un dígito manuscrito basándose en las imágenes más cercanas en el conjunto de entrenamiento.*

## Aprendizaje No Supervisado

En el aprendizaje no supervisado, se trabaja con datos sin etiquetas y el objetivo es encontrar estructuras o patrones ocultos en los datos. Este tipo de aprendizaje es útil para descubrir relaciones y agrupaciones dentro de los datos.

*Ejemplo: Agrupación de clientes en diferentes segmentos de mercado basados en sus comportamientos de compra.*

### Algoritmos comunes:

- **k-Means:** Agrupa datos en k clusters, minimizando la varianza dentro de cada clúster.
  - *Ejemplo: Agrupar clientes en segmentos basados en sus comportamientos de compra para campañas de marketing dirigidas.*
- **Análisis de Componentes Principales (PCA):** Reduce la dimensionalidad de los datos manteniendo la mayor parte de la variabilidad.
  - *Ejemplo: Reducir el número de características en un conjunto de datos de imágenes para facilitar el procesamiento y análisis.*
- **Clustering Jerárquico:** Construye una jerarquía de clústeres de manera aglomerativa (de abajo hacia arriba) o divisiva (de arriba hacia abajo).
  - *Ejemplo: Crear un dendrograma que muestra la relación jerárquica entre diferentes especies de plantas basadas en sus características morfológicas.*

## Aprendizaje por Refuerzo:

El aprendizaje por refuerzo es un paradigma donde un agente aprende a tomar decisiones a través de la interacción con un entorno. El agente recibe recompensas o castigos basados en sus acciones y su objetivo es maximizar la recompensa acumulada a lo largo del tiempo.

***Ejemplo:** Un agente de IA que juega un videojuego y aprende a maximizar su puntuación mediante ensayo y error.*

## Algoritmos comunes:

- **Q-Learning:** Aprende una política para tomar acciones que maximizan la recompensa acumulada.
  - *Ejemplo:* Un robot está “aprendiendo” a navegar por un laberinto para encontrar la salida más rápidamente.
- **Deep Q-Networks (DQN):** Utiliza redes neuronales profundas para aproximar la función de valor en problemas de alta dimensionalidad.
  - *Ejemplo:* Un agente de IA jugando a videojuegos complejos como Atari, aprendiendo estrategias óptimas a través de la experiencia.

# Conceptos Básicos

## Función de Costo (o Pérdida):

La función de costo es una métrica que mide el error entre las predicciones del modelo y los valores reales. El objetivo del entrenamiento del modelo es minimizar esta función de costo.

## Ejemplos de funciones de costo:

- **Error Cuadrático Medio (MSE):** Utilizado en problemas de regresión, mide el promedio de los cuadrados de los errores.

- *Fórmula:*

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- *Ejemplo:* Minimizar la MSE para predecir el precio de una casa.

- **Entropía Cruzada:** Utilizado en problemas de clasificación, mide la diferencia entre las distribuciones de probabilidad predichas y las verdaderas.

- *Fórmula:*

$$H(p, q) = - \sum_i p(x_i) \log q(x_i)$$

- *Ejemplo:* Minimizar la entropía cruzada para clasificar imágenes de dígitos manuscritos.

## Optimización:

La optimización es el proceso de ajustar los parámetros del modelo para minimizar la función de costo. Esto se logra mediante algoritmos de optimización que iterativamente actualizan los parámetros.

## Métodos comunes de optimización:

- **Descenso de Gradiente:** Actualiza los parámetros en la dirección opuesta al gradiente de la función de costo respecto a los parámetros.

- *Fórmula:*

$$\theta = \theta - \alpha \nabla J(\theta)$$

- *Ejemplo:* Ajustar los coeficientes en una regresión lineal para minimizar el error cuadrático medio.
- **Descenso de Gradiente Estocástico (SGD):** Una variante del descenso de gradiente que actualiza los parámetros utilizando un subconjunto aleatorio de datos en cada iteración.
  - *Ventaja:* Convergencia más rápida en grandes conjuntos de datos.
  - *Ejemplo:* Utilizar SGD para entrenar una red neuronal con un gran conjunto de datos de imágenes.

## Overfitting vs. Underfitting:

### Overfitting:

Ocurre cuando un modelo se ajusta demasiado a los datos de entrenamiento, capturando el ruido y los detalles específicos del conjunto de entrenamiento en lugar de generalizar a nuevos datos.

- **Indicadores de Overfitting:** Alta precisión en datos de entrenamiento pero baja precisión en datos de prueba.
- **Técnicas para prevenir Overfitting:** Regularización, validación cruzada (cross-validation), poda (pruning), dropout.
- *Ejemplo:* Un árbol de decisión que es muy profundo y se ajusta a cada punto de datos en el conjunto de entrenamiento.

### Underfitting:

Ocurre cuando un modelo es demasiado simple para capturar la complejidad de los datos de entrenamiento, resultando en un rendimiento pobre tanto en los datos de entrenamiento como en los de prueba.

- **Indicadores de Underfitting:** Baja precisión en ambos conjuntos de datos.
- **Soluciones:** Aumentar la complejidad del modelo, agregar más características, reducir la regularización.
- *Ejemplo:* Utilizar un modelo de regresión lineal para datos que siguen una relación no lineal compleja.

## Preprocesamiento de datos

El pre-procesamiento de datos es una etapa crucial en el pipeline de machine learning. Consiste en preparar y transformar los datos brutos para que sean adecuados para el entrenamiento de modelos.

### Limpieza de Datos:

La limpieza de datos implica eliminar o corregir datos erróneos, duplicados o incompletos. Es un paso esencial para garantizar la calidad de los datos.

### Tareas comunes de limpieza de datos:

- **Eliminación de duplicados:** Remover registros duplicados que pueden sesgar el entrenamiento.
  - *Ejemplo:* Eliminar registros duplicados de transacciones bancarias.
- **Manejo de valores faltantes:** Imputar valores faltantes con la media, mediana o modo, o eliminar registros incompletos.
  - *Ejemplo:* Rellenar los valores faltantes en una columna de ingresos con la media de los ingresos.
- **Corrección de errores:** Identificar y corregir errores en los datos, como entradas malformadas o inconsistencias.
  - *Ejemplo:* Corregir errores tipográficos en entradas de texto.

### Normalización y Escalado:

La normalización y el escalado de características aseguran que todas las características contribuyan por igual al modelo. Esto es especialmente importante para algoritmos que son sensibles a las escalas de las características, como la regresión lineal y las SVM.



## Técnicas comunes:

- **Escalado Mín-Máx:** Escala los datos para que estén dentro de un rango específico, generalmente  $[0, 1]$ .
  - *Fórmula:*

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

- *Ejemplo:* Escalar las alturas de personas para que estén en el rango  $[0, 1]$ .
- **Estandarización:** Transforma los datos para que tengan una media de 0 y una desviación estándar de 1.
  - *Fórmula:*

$$x' = \frac{x - \mu}{\sigma}$$

- *Ejemplo:* Estandarizar los puntajes de exámenes para que tengan una media de 0 y una desviación estándar de 1.

## Transformación de Datos:

La transformación de datos incluye convertir datos categóricos en numéricos y manejar características derivadas.

- **Codificación One-Hot:** Convierte variables categóricas en un formato que puede ser proporcionado a los algoritmos de machine learning. Cada categoría se convierte en una columna binaria (0 o 1).
  - *Ejemplo:* La característica "Color" con valores ["rojo", "verde", "azul"] se transforma en tres columnas: Color\_rojo, Color\_verde y Color\_azul.
- **Binning:** Agrupa valores continuos en intervalos discretos.
  - *Ejemplo:* Agrupar edades en intervalos como 0-10, 11-20, 21-30, etc.
- **Transformaciones Matemáticas:** Aplicar transformaciones logarítmicas, cuadráticas, etc., para manejar distribuciones sesgadas o relaciones no lineales.
  - *Ejemplo:* Aplicar una transformación logarítmica a los ingresos para manejar una distribución sesgada positivamente.