



TIC



Actividad 1

Introducción a Aprendizaje Automático

Introducción a Aprendizaje Automático



El Aprendizaje Automático (Machine Learning en inglés) es una disciplina de la inteligencia artificial que se centra en el estudio y desarrollo de algoritmos y modelos computacionales que permiten a las máquinas aprender patrones y realizar tareas específicas sin necesidad de ser programadas explícitamente para cada una de ellas. En esencia, el objetivo del aprendizaje automático es permitir a los sistemas informáticos mejorar su rendimiento en una tarea particular a medida que se les proporciona más datos y experiencia.



TIC



Introducción a Aprendizaje Automático

Una de las características fundamentales del aprendizaje automático es su capacidad para aprender a partir de ejemplos pasados y adaptarse a nuevas situaciones sin necesidad de intervención humana directa. En lugar de seguir instrucciones predefinidas, los algoritmos de aprendizaje automático pueden aprender automáticamente a identificar patrones y relaciones en conjuntos de datos, lo que les permite realizar predicciones o tomar decisiones.



TIC



Introducción a Aprendizaje Automático



El aprendizaje automático se aplica en una amplia variedad de campos y sectores, incluyendo la medicina, la biología, la industria, las finanzas, la robótica, el transporte, entre otros. Algunos ejemplos de aplicaciones prácticas del aprendizaje automático incluyen el reconocimiento de voz, la detección de fraudes, la recomendación de productos, la conducción autónoma, el diagnóstico médico asistido por ordenador, entre otros.

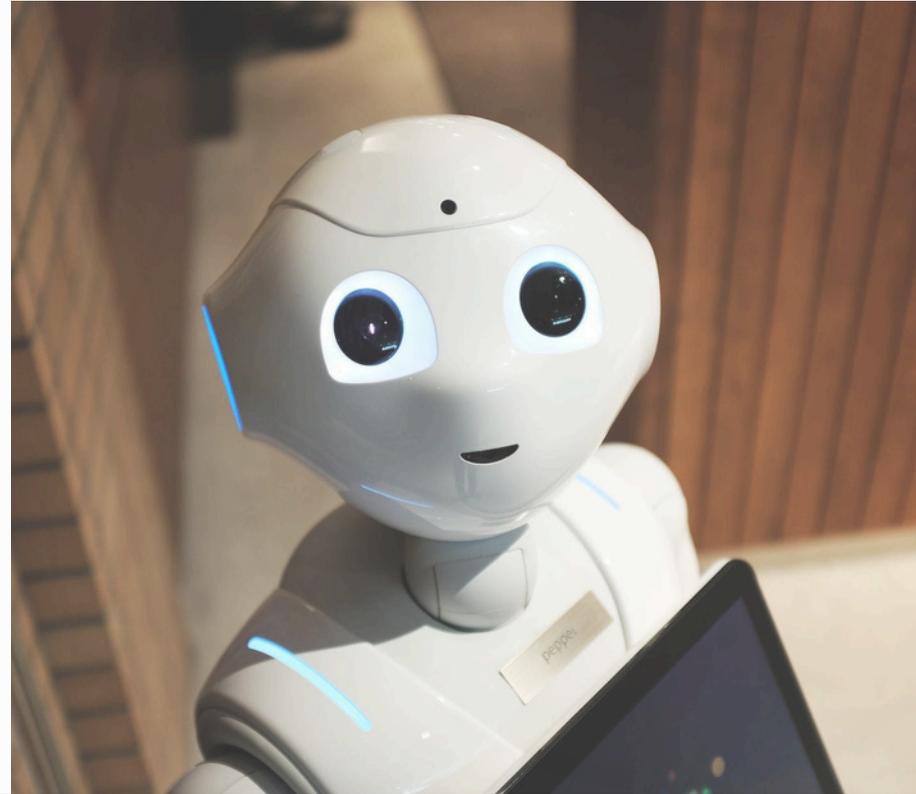


TIC



Introducción a Aprendizaje Automático

El aprendizaje automático es una herramienta poderosa que permite a las máquinas aprender y mejorar su rendimiento en tareas específicas a partir de datos y experiencia pasada, lo que tiene el potencial de impulsar la innovación y transformar numerosos aspectos de nuestras vidas y sociedades.



TIC



Introducción a Aprendizaje Automático



El aprendizaje automático es una forma de estadística aplicada con el uso de computadoras para estimar funciones complejas.

Los algoritmos de aprendizaje automático se pueden dividir en las categorías de aprendizaje supervisado y no supervisado.

Un algoritmo de aprendizaje automático se construye combinando un algoritmo de optimización, una función de costo, un modelo y un conjunto de datos.



TIC



Introducción a Aprendizaje Automático



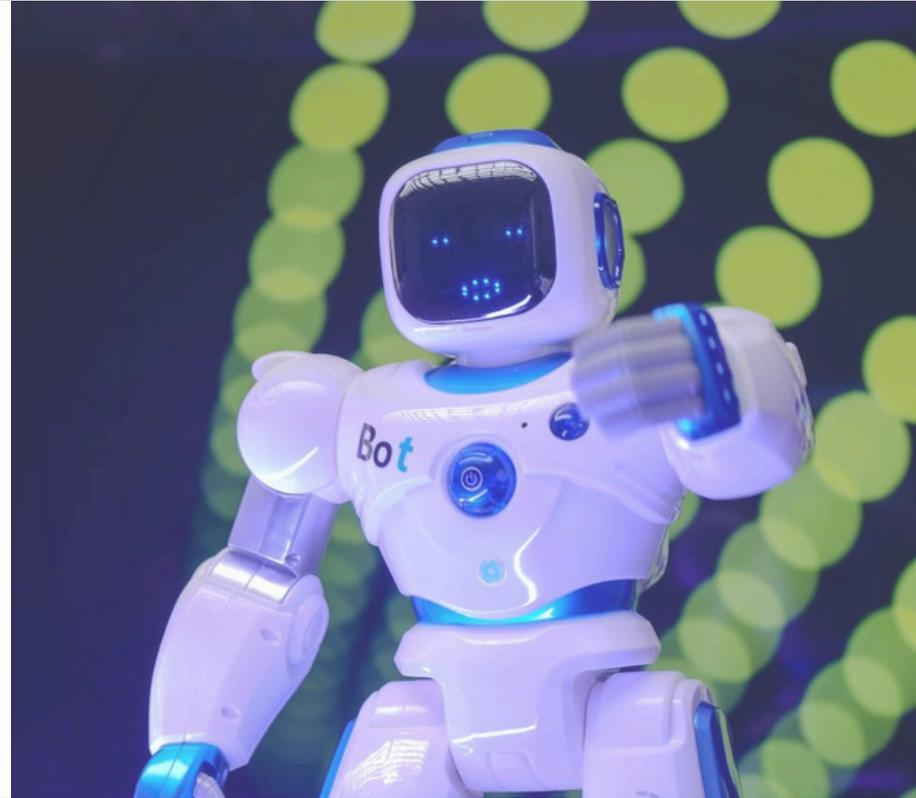
TIC



Un algoritmo de aprendizaje es un algoritmo que puede aprender de los datos.

¿Pero qué se entiende por aprender?

Mitchell (1997) proporciona la definición "Se dice que un programa de computadora aprende de la experiencia E con respecto a algunas tareas T y la medida de rendimiento R , si su Rendimiento en Tareas mejora con la Experiencia".



Tareas en Aprendizaje Automático



TIC



En el contexto del Aprendizaje Automático, una "tarea" se refiere a una acción o objetivo específico que se busca lograr mediante la aplicación de algoritmos y modelos de aprendizaje automático a conjuntos de datos relevantes.

Las tareas en el aprendizaje automático pueden ser muy diversas y abarcar una amplia gama de aplicaciones y problemas.

El Aprendizaje Automático se enfrenta con tareas difíciles de resolver para programas escritos y diseñados por seres humanos.

El aprendizaje es el medio para obtener la capacidad de realizar la tarea.



Tareas en Aprendizaje Automático



TIC



Ejemplos comunes de tareas en aprendizaje automático incluyen:



Estos son solo algunos ejemplos de las muchas tareas que se pueden abordar utilizando técnicas de aprendizaje automático. Cada tarea tiene sus propias características y desafíos únicos, y se pueden emplear diferentes algoritmos y enfoques para resolverlas de manera efectiva. La elección del algoritmo y la metodología adecuados depende en gran medida de la naturaleza específica de la tarea y de las características de los datos disponibles.

Una tarea es el proceso que se lleva a cabo sobre la entrada, para obtener una salida determinada.



Detección de anomalías

Esta tarea implica identificar patrones o instancias inusuales en un conjunto de datos, que pueden indicar comportamientos anómalos o potencialmente peligrosos. Por ejemplo, detectar transacciones fraudulentas en una base de datos de transacciones bancarias.



Recomendación

En la recomendación, el objetivo es predecir qué productos, servicios o contenido serán de interés para un usuario en particular, basándose en sus preferencias y comportamientos pasados, así como en las de usuarios similares. Por ejemplo, sistemas de recomendación de películas o música en plataformas de streaming.



Agrupamiento

En esta tarea, el objetivo es dividir un conjunto de datos en grupos o clústeres, donde los miembros de cada grupo son más similares entre sí que con los miembros de otros grupos. Por ejemplo, agrupar clientes en segmentos de mercado con características similares para campañas de marketing personalizadas.



Regresión

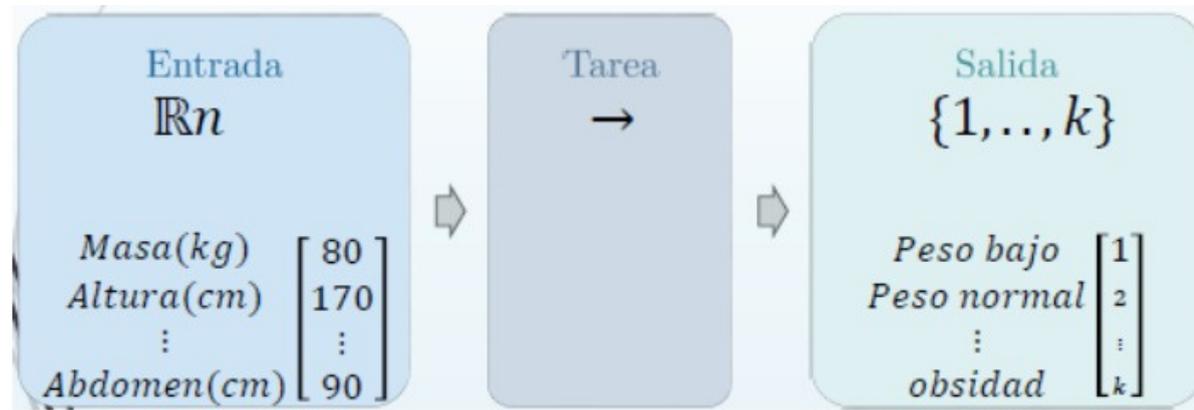
En la regresión, el objetivo es predecir un valor numérico continuo en función de un conjunto de variables de entrada. Por ejemplo, predecir el precio de una casa en función de su tamaño, ubicación y otras características.





Clasificación

En esta tarea, el objetivo es asignar una etiqueta o categoría a un conjunto de datos, con base en las características observadas. Por ejemplo, clasificar correos electrónicos como spam o no spam, o clasificar imágenes digitales en categorías como perros, gatos o pájaros.



Conjuntos de Datos y Divisiones



TIC



Los conjuntos de datos y sus divisiones, incluyendo el conjunto de entrenamiento, conjunto de prueba y conjunto de validación, son componentes fundamentales en el proceso de entrenamiento y evaluación de modelos de aprendizaje automático.



Conjunto de
datos



Divisiones



Consideraciones
generales

La correcta gestión de conjuntos de datos y divisiones es esencial para el desarrollo de modelos precisos y generalizables. Cada conjunto cumple un papel único en el proceso de aprendizaje supervisado, desde el entrenamiento inicial hasta la evaluación final. La representatividad y la imparcialidad son claves para obtener resultados confiables y aplicables en entornos del mundo real.



Consideraciones Generales:

El tamaño adecuado de los conjuntos de entrenamiento, prueba y validación depende del tamaño total del conjunto de datos y la complejidad del problema.

Es crucial estrati car los conjuntos para mantener la proporción de clases en problemas de clasi cación, evitando así desequilibrios.



En la práctica del aprendizaje automático, es común dividir el conjunto de datos en tres partes principales:

Conjunto de Entrenamiento:

El conjunto de entrenamiento es una parte del conjunto de datos utilizado para entrenar modelos de aprendizaje supervisado.

Su función principal es permitir que el modelo aprenda patrones y relaciones entre las características de entrada y las salidas esperadas.

La calidad del modelo depende en gran medida de la representatividad del conjunto de entrenamiento. Debe abarcar la diversidad de casos que el modelo encontrará en situaciones reales para generalizar de manera efectiva.

Conjunto de Prueba:

El conjunto de prueba se utiliza para evaluar el rendimiento del modelo después de que ha sido entrenado en el conjunto de entrenamiento.

Proporciona una evaluación imparcial del modelo, permitiendo medir su capacidad para generalizar a datos no vistos.

Es esencial que el conjunto de prueba sea independiente y no esté sesgado hacia ninguna característica específica. De lo contrario, las métricas de rendimiento pueden no reflejar la verdadera capacidad del modelo.

Conjunto de Validación:

El conjunto de validación se utiliza durante el proceso de entrenamiento para ajustar hiperparámetros y evitar el sobreajuste.

Permite evaluar cómo se desempeña el modelo en datos no utilizados durante el entrenamiento y ajustar la configuración del modelo para obtener un rendimiento óptimo.



Clasificación

Un conjunto de datos es simplemente una colección de ejemplos o instancias que se utilizan para entrenar, probar y validar modelos de aprendizaje automático. Cada ejemplo en un conjunto de datos consiste en características o variables que describen las entradas del modelo, así como las etiquetas correspondientes que representan las salidas deseadas o el resultado que se intenta predecir.

Por ejemplo, en un problema de clasificación de imágenes de animales, cada ejemplo puede consistir en una imagen de un animal junto con una etiqueta que indica la especie a la que pertenece.



TIC

