

# Aprendizaje supervisado versus no supervisado: ¿cuál es la diferencia?

## ¿Qué es el aprendizaje supervisado?

El aprendizaje supervisado es un enfoque de aprendizaje automático que se define por el uso de conjuntos de datos etiquetados. Estos conjuntos de datos están diseñados para entrenar o "supervisar" algoritmos para clasificar datos o predecir resultados con precisión. Utilizando entradas y salidas etiquetadas, el modelo puede medir su precisión y aprender con el tiempo.

El aprendizaje supervisado se puede dividir en dos tipos de problemas al extraer datos : clasificación y regresión:

- **Los problemas de clasificación** utilizan un algoritmo para asignar con precisión datos de prueba en categorías específicas, como separar manzanas de naranjas. O, en el mundo real, se pueden utilizar algoritmos de aprendizaje supervisado para clasificar el spam en una carpeta separada de su bandeja de entrada. Los clasificadores lineales, las máquinas de vectores de soporte, los árboles de decisión y los bosques aleatorios son tipos comunes de algoritmos de clasificación.
- **La regresión** es otro tipo de método de aprendizaje supervisado que utiliza un algoritmo para comprender la relación entre variables dependientes e independientes. Los modelos de regresión son útiles para predecir valores numéricos basados en diferentes puntos de datos, como las proyecciones de ingresos por ventas para un negocio determinado. Algunos algoritmos de regresión populares son la regresión lineal, la regresión logística y la regresión polinómica.

## ¿Qué es el aprendizaje no supervisado?

El aprendizaje no supervisado utiliza algoritmos de aprendizaje automático para analizar y agrupar conjuntos de datos sin etiquetar. Estos algoritmos descubren patrones ocultos en los datos sin necesidad de intervención humana (por lo tanto, no están "supervisados").

Los modelos de aprendizaje no supervisados se utilizan para tres tareas principales: agrupación, asociación y reducción de dimensionalidad:

- **La agrupación** es una técnica de minería de datos para agrupar datos sin etiquetar en función de sus similitudes o diferencias. Por ejemplo, los algoritmos de agrupación de K-medias asignan puntos de datos similares en grupos, donde el valor de K representa el tamaño de la agrupación y la granularidad. Esta técnica es útil para la segmentación del mercado, la compresión de imágenes, etc.
- **La asociación** es otro tipo de método de aprendizaje no supervisado que utiliza diferentes reglas para encontrar relaciones entre variables en un conjunto de datos determinado. Estos métodos se utilizan con frecuencia para el análisis de la cesta de la compra y los motores de recomendación, en la línea de las recomendaciones de "Los clientes que compraron este artículo también compraron".

- **La reducción de dimensionalidad** es una técnica de aprendizaje que se utiliza cuando la cantidad de características (o dimensiones) en un conjunto de datos determinado es demasiado alta. Reduce la cantidad de entradas de datos a un tamaño manejable y al mismo tiempo preserva la integridad de los datos. A menudo, esta técnica se utiliza en la etapa de preprocesamiento de datos, como cuando los codificadores automáticos eliminan el ruido de los datos visuales para mejorar la calidad de la imagen.

### La principal diferencia entre aprendizaje supervisado y no supervisado: datos etiquetados

La principal distinción entre los dos enfoques es el uso de conjuntos de datos etiquetados. En pocas palabras, el aprendizaje supervisado utiliza datos de entrada y salida etiquetados, mientras que un algoritmo de aprendizaje no supervisado no.

En el aprendizaje supervisado, el algoritmo "aprende" del conjunto de datos de entrenamiento haciendo predicciones iterativas sobre los datos y ajustándolos para obtener la respuesta correcta. Si bien los modelos de aprendizaje supervisado tienden a ser más precisos que los modelos de aprendizaje no supervisados, requieren una intervención humana inicial para etiquetar los datos de manera adecuada. Por ejemplo, un modelo de aprendizaje supervisado puede predecir cuánto durará su viaje en función de la hora del día, las condiciones climáticas, etc. Pero primero tendrás que entrenarlo para que sepa que el tiempo lluvioso prolonga el tiempo de conducción.

Los modelos de aprendizaje no supervisados, por el contrario, funcionan por sí solos para descubrir la estructura inherente de los datos no etiquetados. Tenga en cuenta que todavía requieren cierta intervención humana para validar las variables de salida. Por ejemplo, un modelo de aprendizaje no supervisado puede identificar que los compradores en línea suelen comprar grupos de productos al mismo tiempo. Sin embargo, un analista de datos necesitaría validar que tiene sentido que un motor de recomendación agrupe la ropa de bebé con un pedido de pañales, puré de manzana y vasitos con sorbete.