

Machine Learning

Proceso: Conjunto de actividades secuenciales que transforman insumos en resultados.

Procedimiento: Secuencia detallada de acciones para llevar a cabo una tarea específica.

Modelo probabilístico

Es una técnica estadística usado para tomar en cuenta el impacto de eventos aleatorios o acciones en predecir el potencial de resultados futuros

¿Qué es Machine Learning?

Es el campo de investigación enfocado a un estudio formal de sistemas de aprendizaje, se construye de ideas tales como estadística, ciencia de computación, ingeniería, ciencia cognitiva, teoría de optimización y muchas otras disciplinas

Aprendizaje supervisado y aprendizaje no supervisado

Aprendizaje supervisado: Modelo de aprendizaje automático donde se entrena a un algoritmo usando un conjunto de datos con entradas y salidas conocidas.

- Menos complejo
- predice salidas de nuevos datos
- Mas resultados certeros
- Mas fácil de entender
- Variables de entrada y salida son dados
- Este aprendizaje se puede hacer fuera de línea
- Usado para Sentiment análisis, predicción de precio, detección de spam, etc

Aprendizaje no supervisado: Modelo de aprendizaje automático donde se entrena a un algoritmo usando un conjunto de datos sin salidas etiquetadas, buscando patrones o estructuras en los datos.

- Mas complejo
- Obtener información a partir de enormes cantidades de nuevos datos.
- Menos resultados precisos

- Ambigüedad de datos requiere más poder de computo
- Solo variables de entrada son dadas
- El aprendizaje es en línea

Aprendizaje supervisado:

El aprendizaje supervisado es una técnica de aprendizaje automático en la cual un algoritmo se entrena utilizando un conjunto de datos que contiene entradas y sus correspondientes salidas o respuestas. El objetivo es que, una vez entrenado, el algoritmo pueda predecir la salida para nuevas entradas que no estuvieron presentes en el conjunto de entrenamiento.

Las principales características del aprendizaje supervisado incluyen:

1. **Datos etiquetados:** Se utiliza un conjunto de datos donde cada ejemplo viene con una etiqueta o respuesta correcta.
2. **Predicción:** Una vez que el modelo está entrenado, puede predecir la etiqueta o respuesta para datos desconocidos.
3. **Tipos de problemas:** Se aplica principalmente a problemas de regresión (predicción de valores continuos) y clasificación (asignación de categorías a entradas).

Ejemplo: Supongamos que queremos predecir si va a llover o no basándonos en la temperatura y la humedad. Con un conjunto de datos histórico donde sabemos las temperaturas, humedades y si llovió o no en días pasados, podemos entrenar un modelo de aprendizaje supervisado. Una vez entrenado, si le proporcionamos una nueva temperatura y humedad, el modelo puede predecir si lloverá o no en base a lo que aprendió.

Aprendizaje No supervisado:

El aprendizaje no supervisado es una técnica de aprendizaje automático en la que un algoritmo se entrena utilizando un conjunto de datos que no tiene etiquetas o respuestas previamente definidas. A diferencia del aprendizaje supervisado, donde el objetivo es aprender a predecir una salida basada en la entrada, en el aprendizaje no supervisado el objetivo es identificar patrones, estructuras o relaciones subyacentes en los datos.

Las principales características del aprendizaje no supervisado incluyen:

1. **Datos sin etiquetar:** Se utiliza un conjunto de datos que no tiene respuestas o etiquetas predefinidas para cada ejemplo.
2. **Descubrimiento de patrones:** El algoritmo busca encontrar estructuras, relaciones o patrones ocultos en los datos.

3. **Tipos de problemas:** Los problemas más comunes abordados por el aprendizaje no supervisado son el clustering (agrupación) y la reducción de dimensionalidad.

Ejemplos:

- **Clustering:** Supongamos que tenemos datos sobre las compras de clientes en un supermercado, pero no sabemos nada sobre los tipos de clientes. El aprendizaje no supervisado puede identificar grupos (o clusters) de clientes basados en sus patrones de compra, como clientes que compran productos saludables, clientes que compran en grandes cantidades, etc.
- **Reducción de dimensionalidad:** Si se tiene un conjunto de datos con muchas características (o dimensiones), el aprendizaje no supervisado puede ayudar a identificar las características más significativas y reducir la complejidad del conjunto de datos sin perder mucha información, facilitando así su análisis y visualización.

En resumen, el aprendizaje no supervisado se centra en descubrir estructuras ocultas en datos sin etiquetar, en lugar de predecir salidas específicas basadas en entradas dadas.

Clasificación vs regresión

Clasificación y regresión son dos tipos de problemas en el aprendizaje supervisado. Ambos involucran predecir una salida basada en una o más entradas, pero difieren en el tipo de salida que se predice.

1. Clasificación:

- **Objetivo:** Asignar una entrada dada a una de varias categorías o clases predefinidas.
- **Tipo de salida:** Categórico o discreto. Por ejemplo, clasificar correos electrónicos como "spam" o "no spam".
- **Ejemplo:** Dada una imagen, clasificarla como "gato", "perro" o "ave".

2. Regresión:

- **Objetivo:** Predecir un valor continuo basado en la entrada.
- **Tipo de salida:** Continuo o numérico. Por ejemplo, predecir el precio de una casa basado en su tamaño y ubicación.
- **Ejemplo:** Dadas las horas de estudio, predecir el puntaje que obtendrá un estudiante en un examen.

OverFitting vs Underfitting

No está suficientemente entrenado

Esta demasiado entrenado, no generaliza y la función toca todos los puntos de entrenamiento

Para poder reducir el Overfitting, nosotros debemos reducir el espacio de parámetro, incrementar el espacio de muestreo e identificar que feature es la más fundamental, y después eliminarla