

Actividad 1

Módulo: Aprendizaje de Máquina Supervisado

Intención del aprendizaje esperado:

1. Describir los conceptos fundamentales de aprendizaje de máquina para resolver un problema

Ejercicios Planteados

Pregunta 1: Investigue respecto a los siguientes conceptos.

- Aprendizaje supervisado:

En el aprendizaje supervisado, la máquina se enseña con el ejemplo. De este modo, el operador proporciona al algoritmo de aprendizaje automático un conjunto de datos conocidos que incluye las entradas y salidas deseadas, y el algoritmo debe encontrar un método para determinar cómo llegar a esas entradas y salidas.

Mientras el operador conoce las respuestas correctas al problema, el algoritmo identifica patrones en los datos, aprende de las observaciones y hace predicciones. El algoritmo realiza predicciones y es corregido por el operador, y este proceso sigue hasta que el algoritmo alcanza un alto nivel de precisión y rendimiento.

En el aprendizaje supervisado, los algoritmos trabajan con datos “etiquetados” (*labeled data*), intentando encontrar una función que, dadas las variables de entrada (*input data*), les asigne la etiqueta de salida adecuada. El algoritmo se entrena con un “histórico” de datos y así “aprende” a asignar la etiqueta de salida adecuada a un nuevo valor, es decir, predice el valor de salida.

El aprendizaje supervisado se suele usar en: Problemas de *clasificación* (identificación de dígitos, diagnósticos, o detección de fraude de identidad) y Problemas de *regresión* (predicciones meteorológicas, de expectativa de vida, de crecimiento etc).

Estos dos tipos principales de aprendizaje supervisado, clasificación y regresión, se distinguen por el tipo de variable objetivo. En los casos de clasificación, *es de tipo categórico*, mientras que, en los casos de regresión, la variable objetivo *es de tipo numérico*.

Los algoritmos más habituales que aplican para el aprendizaje supervisado son:

1. Árboles de decisión.
2. Clasificación de Naïve Bayes.
3. Regresión por mínimos cuadrados.
4. Regresión Logística.
5. Support Vector Machines (SVM).
6. Métodos “Ensemble” (Conjuntos de clasificadores).

- Aprendizaje sin supervisión:

Aquí, el algoritmo de aprendizaje automático estudia los datos para identificar patrones. No hay una clave de respuesta o un operador humano para proporcionar instrucción. En cambio, la máquina determina las correlaciones y las relaciones mediante el análisis de los datos disponibles.

En un proceso de aprendizaje no supervisado, se deja que el algoritmo de aprendizaje automático interprete grandes conjuntos de datos y dirija esos datos en consecuencia. Así, el

algoritmo intenta organizar esos datos de alguna manera para describir su estructura. Esto podría significar la necesidad de agrupar los datos en grupos u organizarlos de manera que se vean más organizados.

A medida que evalúa más datos, su capacidad para tomar decisiones sobre los mismos mejora gradualmente y se vuelve más refinada.

El aprendizaje no supervisado tiene lugar cuando no se dispone de datos “etiquetados” para el entrenamiento. Sólo conocemos los datos de entrada, pero no existen datos de salida que correspondan a un determinado *input*. Por tanto, sólo podemos describir la estructura de los datos, para intentar encontrar algún tipo de organización que simplifique el análisis. Por ello, tienen un carácter exploratorio. Por ejemplo, las tareas *de clustering*, buscan agrupamientos basados en similitudes, pero nada garantiza que éstas tengan algún significado o utilidad. En ocasiones, al explorar los datos sin un objetivo definido, se pueden encontrar correlaciones espúreas curiosas, pero poco prácticas. El aprendizaje no supervisado se suele usar en: Problemas de *clustering*, Agrupamientos de co-ocurrencias y Perfilado o *profiling*. Sin embargo, los problemas que implican tareas de encontrar similitud, predicción de enlaces o reducción de datos, pueden ser supervisados o no.

Los tipos de algoritmo más habituales en aprendizaje no supervisado son:

1. Algoritmos de clustering
2. Análisis de componentes principales
3. Descomposición en valores singulares (singular value decomposition)
4. Análisis de componentes principales (Independent Component Analysis)

- Aprendizaje por refuerzo:

El aprendizaje por refuerzo se centra en los procesos de aprendizajes reglamentados, en los que se proporcionan algoritmos de aprendizaje automáticos con un conjunto de acciones, parámetros y valores finales.

Al definir las reglas, el algoritmo de aprendizaje automático intenta explorar diferentes opciones y posibilidades, monitorizando y evaluando cada resultado para determinar cuál es el óptimo.

En consecuencia, este sistema enseña la máquina a través del proceso de ensayo y error. Aprende de experiencias pasadas y comienza a adaptar su enfoque en respuesta a la situación para lograr el mejor resultado posible.

Este tipo aprendizaje se basa en mejorar la respuesta del modelo usando un proceso de retroalimentación. El algoritmo aprende observando el mundo que le rodea. Su información de entrada es el *feedback* o retroalimentación que obtiene del mundo exterior como respuesta a sus acciones. Por lo tanto, el sistema aprende a base de ensayo-error.

No es un tipo de aprendizaje supervisado, porque no se basa estrictamente en un conjunto de datos etiquetados, sino en la monitorización de la respuesta a las acciones tomadas. Tampoco es un aprendizaje no supervisado, ya que, cuando modelamos a nuestro "aprendiz" sabemos de antemano cuál es la recompensa esperada.

- **Algoritmos de regresión:** En las tareas de regresión, el programa de aprendizaje automático debe estimar y comprender las relaciones entre las variables. El análisis de regresión se enfoca en una variable dependiente y una serie de otras variables cambiantes, lo que lo hace particularmente útil para la predicción y el pronóstico.

- **Algoritmos de agrupación:** Se utilizan en el aprendizaje no supervisado, y sirven para categorizar datos no etiquetados, es decir, datos sin categorías o grupos definidos. El algoritmo funciona mediante la búsqueda de grupos dentro de los datos, con el número de grupos representados por la variable K. A continuación, funciona de manera iterativa para asignar cada punto de datos a uno de los K grupos según las características proporcionadas.

- **Algoritmos de Aprendizaje Profundo:** Los algoritmos de aprendizaje profundo ejecutan datos a través de varias capas de algoritmos de redes neuronales, las cuales pasan a una representación simplificada de los datos a la siguiente capa.

La mayoría funciona bien en conjuntos de datos que tienen hasta unos cientos de características o columnas. Sin embargo, un conjunto de datos no estructurado, como el de una imagen, tiene una cantidad tan grande de características que este proceso se vuelve engorroso o completamente inviable. Los algoritmos de aprendizaje profundo aprenden progresivamente más sobre la imagen a medida que pasa por cada capa de red neuronal. Las primeras capas aprenden a detectar características de bajo nivel como los bordes, y las capas posteriores combinan las características de las capas anteriores en una representación holística.

Pregunte 2: Busque una imagen que una los conceptos anteriores o cree usted un diagrama que los relacione (Puede agregar otros conceptos si lo cree necesario).



