

Diferencia entre algoritmos de clasificación y regresión



Ya hemos visto la diferencia de los algoritmos que tenemos que utilizar en Machine Learning si tenemos datos etiquetados o no, ahora veremos la diferencia de estos algoritmos basados en el formato de sus salidas, para este caso existen dos tipos populares de métodos de Machine Learning y es el de clasificación y regresión.

Comprender la diferencia entre la clasificación y la regresión es útil para comprender los algoritmos relacionados con ellos.

Algoritmos de Clasificación

Los algoritmos de clasificación se usan cuando el resultado deseado es una **etiqueta discreta**, en otras palabras, son útiles cuando la respuesta al problema cae dentro de un **conjunto finito de resultados posibles**.

En el caso de que el modelo entrenado es para predecir cualquiera de las dos clases objetivas, **verdadero o falso**, por ejemplo, se le conoce como **clasificación binaria**. Algunos ejemplos de esto son: predecir si un alumno **aprobará o no**, predecir si un cliente **comprará un producto nuevo o no**.

Por su parte, si tenemos que predecir **más de dos clases objetivas**, se le conoce como **clasificación multicategoría**. Un ejemplo de esto es predecir qué **asignaturas un alumno tendrá las más clasificaciones**. Este tipo de clasificación es útil para la **segmentación del cliente**, la **categorización de imágenes y audio y análisis de texto** para optimizar el sentimiento del cliente.

Algoritmos de Regresión

Por otro lado, la regresión es útil para predecir **productos que son continuos**, eso significa que la respuesta a su pregunta se presenta mediante una cantidad que puede **determinarse de manera flexible en función de las entradas del modelo** en lugar de limitarse a un conjunto de etiquetas. **En algunos casos, el valor predicho se puede usar para identificar la relación lineal entre los atributos**.

La **regresión lineal es el ejemplo más popular de un algoritmo de regresión**, aunque a menudo se subestima debido a su relativa simplicidad, es un método versátil que se puede usar para predecir los **precios de la vivienda, la probabilidad de que los clientes se desvíen o los ingresos que un cliente generará**.

Diferencia entre algoritmos de clasificación y regresión



Veamos un ejemplo de esto con un mismo conjunto de datos pero con diferentes objetivos. Supongamos que tenemos los datos de todas las películas que has visto y que a su vez te han gustado, con esta información queremos determinar lo siguiente:

Si una nueva película que está saliendo en los cines **te va a gustar o no**.

En caso de que si te guste la nueva película, **¿cuántas veces la verás?**.

¿Ya has podido determinar cuál es el problema de clasificación y cuál es el de regresión?

Para la primera pregunta, si una nueva película que está saliendo en los cines te va a gustar o no, la solución a este problema sera de tipo **clasificación**, ya que estamos clasificando las cosas según sus pertenencias, sí o no, me gusta o no. En este caso, el problema es una clasificación binaria en que la que tenemos que predecir si la salida pertenece a la clase **1 (sí) o clase 2 (no)**.

Para la segunda pregunta, en caso de si te guste la nueva película, ¿cuántas veces la verás?, este problema es de encontrar el recuento, es decir debemos predecir un valor, por lo tanto es un problema de **regresión**, ya que podría ser **5 veces, 6 veces o 10 veces**, todo depende de los datos de entrada.

Como pueden ver, elegir un algoritmo es un paso crítico en el proceso de Machine Learning, por lo que es importante que realmente se adapte al caso de uso del problema en cuestión.