

Aprendizaje no supervisado

Guillermo Ruiz

El aprendizaje supervisado es una rama de la Inteligencia Artificial que se enfoca en aprender de datos no etiquetados. En el mundo de la IA existen dos clases de datos, los etiquetados y los no etiquetados. En los etiquetados tenemos los datos y una columna extra que nos dice un valor que nos gustaría poder predecir mediante la computadora. Por ejemplo, los datos pueden ser información sobre casas,

Metros cuadrados	Cuartos	Baños
120	2	2
110	3	2.5

y las etiquetas serían los precios de venta. Dados los datos, queremos predecir de manera automática las etiquetas. Las etiquetas pueden ser valores continuos como el precio, temperatura o velocidad; aunque lo más común es que sean tomadas de una lista, como [**positivo**, **negativo**, **neutro**]. Una de las cosas más complicadas de obtener al crear un nuevo conjunto de datos son las etiquetas ya que por lo general se necesita de un humano que las asigne de manera individual. Los conjuntos de datos suelen tener muchos ejemplos para que puedan ser usados por técnicas como el aprendizaje profundo. Los pequeños tienen miles de ejemplos y los grandes pueden llegar a los millones. Etiquetar un millón de ejemplos de manera manual es muy tardado y costoso, se necesita de un ejército de etiquetadores.

Los datos no etiquetados también pueden ser usados para extraer información. Las dos tareas principales son la reducción de dimensión y la organización en grupos. En el aprendizaje no supervisado se aprende de los datos tal como son, mientras que en el supervisado se aprende a predecir a los etiquetadores. En esta unidad, vamos a ver algoritmos de agrupamiento cuyo objetivo es organizar los datos en grupos con la propiedad de que elementos de un mismo grupo sean *parecidos* y al mismo tiempo sean *diferentes* de los otros grupos. De manera formal, no hay una definición para decidir si dos elementos son *parecidos* pues depende del problema. Esta ambigüedad hace que el problema sea complejo y que no exista una solución definitiva. Esta complicación se verá de manera especial durante la sección de evaluación.