

Entender el aprendizaje automático

Entonces, ¿qué es el aprendizaje automático (*machine learning*)? En realidad, esta es una pregunta muy difícil de responder si han intentado buscarla en Google, porque de hecho, hay mucho debate sobre lo que es AA. La Ciencia de datos, estadística, inteligencia artificial (IA), el aprendizaje automático, todos ellos, cubren temas similares, y las personas no se han puesto de acuerdo cómo se debe definir perfectamente el aprendizaje automático.

Algunas personas piensan que el aprendizaje automático es solo automatización, y eso tiene sentido, porque a menudo las aplicaciones de AA son, hasta cierto punto, automatización. Por ejemplo, ustedes pueden usar un computador para transcribir una entrevista en lugar de usar a una persona, o pueden usar algo como Google Translate para traducir el texto que una persona hubiera tenido que traducir en otra situación.

Pero, de hecho, desde que surgió la programación, los programadores han automatizado las cosas. Por ejemplo, tal vez ustedes tienen un robot de limpieza de pisos como Roomba. Bueno, ustedes no necesitan aprendizaje automático para que funcione. Podrían escribir un código que diga que cuando se presiona el botón de encendido, el robot debe barrer el piso. Y eso no es aprendizaje automático. Además, el aprendizaje automático puede hacer más que simplemente automatizar. Por ejemplo, puede hacer cosas que los humanos no pueden hacer, como hacer una recomendación sobre lo que un lector podría querer leer con base en lo que leyó anteriormente. O predecir, no sé, las tendencias climáticas o el precio de bitcoin. O hacer obras de arte locas como esa.

¿Cómo se ajusta esto al patrón de lo que es el resto del aprendizaje automático? Bueno, la forma en que me gusta describir el AA es que es solo encontrar patrones en los datos y, generalmente, usar esos patrones para hacer predicciones sobre datos no vistos. Pueden pensarlo como una especie de nuevo paradigma de programación. Entonces, ¿qué significa esto? Bueno, echemos un vistazo a la clasificación. Esta es una de las tareas en las que las personas usan el aprendizaje automático. Y es una de las más comunes, así que hablemos de clasificación.

La clasificación es solo etiquetar cosas. Entonces, tal vez ustedes tengan fotos de perros y gatos, y deseen crear un algoritmo que, cuando reciba nuevas fotos, pueda etiquetar automáticamente a los perros y gatos por ustedes. La forma en que tradicionalmente haríamos esto con la programación, antes del aprendizaje automático, es que un programador se sentaría y pensaría en qué hace que un gato sea un gato y un perro sea un perro, y haría muchas reglas diferentes en su cabeza. Y luego pondría eso en su algoritmo. Entonces, si un animal tiene una cola larga y le gustan los peces y no le gustan las personas, entonces es un gato. De lo contrario es un perro. Este es el enfoque de la programación tradicional.

Lo que es diferente con el aprendizaje automático es que, en lugar de que un programador establezca unas reglas, un algoritmo de aprendizaje automático aprende con el ejemplo y crea

las reglas él mismo. La forma en que esto sucede es que construimos un modelo de máquina. Entonces, primero tenemos que recopilar muchos datos etiquetados para el entrenamiento, como cientos o miles de fotos de gatos etiquetados como gatos y perros etiquetados como perros. Y preferiblemente de muchas razas diferentes y desde diferentes ángulos, y con una iluminación diferente. Le mostramos estos datos etiquetados a un algoritmo de aprendizaje automático el cual devuelve un modelo. Entonces, si su algoritmo hizo un buen trabajo, cuando el modelo ve una nueva imagen de un perro, puede etiquetarlo como un perro. Eso, en pocas palabras, es aprendizaje automático. O al menos eso es un algoritmo de clasificación, que les dije es una gran parte del aprendizaje automático. Pero, de nuevo, es un poco difícil conceptualizar realmente todas las diferentes formas en que se puede aplicar AA, por lo que lo que quiero hacer es, en las siguientes secciones, mostrarles varias aplicaciones diferentes de aprendizaje automático utilizadas por redacciones en diferentes tipos de datos. Y con suerte, ustedes, como un algoritmo de aprendizaje automático, pueden aprender con el ejemplo lo que cubre este campo.