

UNIVERSIDAD  
COMPLUTENSE  
DE MADRID



Autor: María José  
Gómez Silva

# Machine Learning con Python. Semana 1.

## CAPÍTULO 1. INTRODUCCIÓN AL MACHINE LEARNING.

### 1.2 ¿Qué es un Modelo?

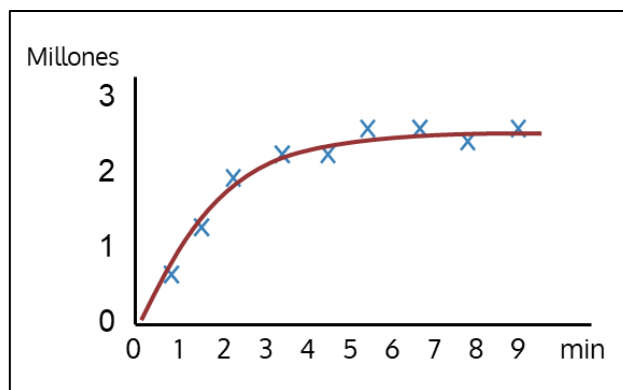
#### ¿QUÉ ES UN MODELO?

En el ámbito de la ciencia y la tecnología, un modelo es una representación abstracta, conceptual o matemática de un componente, un sistema, un proceso o una tarea.

Los modelos son interpretaciones que intentan conceptualizar o dar un sentido a los datos observados. En términos estadísticos, con los modelos se pretende generalizar el comportamiento de la población global a partir del estudio de un conjunto de muestras. Los modelos son muy útiles para realizar predicciones.

El modelo recibe datos de entrada, y ofrece un dato o varios de salida. Esa salida, también conocida como *predicción* del modelo, es la respuesta o solución que el modelo ofrece para los datos recibidos a la entrada.

Por ejemplo, imaginemos que un empresario realiza varias campañas publicitarias, cada año varía el número de minutos al día que su producto es anunciado en la radio, y cada año registra el número de ventas, obteniendo una gráfica como la de la Figura 1. A partir de los datos registrados (muestras), intenta ajustar un modelo o función (en rojo), donde la entrada es el número de minutos, y la salida el número de productos vendidos (en millones). El modelo ajustado permitirá al empresario realizar predicciones sobre el resultado de ventas que se obtendría para un número distinto de minutos en la radio, sin necesidad de probarlo, es decir permite conocer el resultado del proceso (su salida) frente a datos de entrada nuevos (desconocidos). Cuanto mejor ajustado esté el modelo, más precisa será la predicción.



*Figura 1.: Ajuste de un modelo de la cantidad de productos vendidos en función de los minutos diarios de publicidad en la radio.*

El Machine Learning como su nombre indica, permite un aprendizaje automático del modelo que mejor se ajusta a los datos suministrados. Se trata de encontrar el modelo más representativo, es decir, el que mejor rendimiento presenta en la realización de la tarea.

Cuando la tarea que la máquina debe aprender es compleja, como por ejemplo jugar al ajedrez, puede ocurrir que el modelo no sea simplemente una función, sino el conjunto de funciones, ecuaciones matemáticas, y pasos que la máquina debe realizar para resolver la tarea.

Al inicio del proceso de aprendizaje, la máquina no conoce cuál es el modelo a seguir. En el caso del ajedrez, no sabe a qué piezas o movimientos debe prestar atención, ni qué movimiento realizar para obtener ventaja en la partida. Por ello, al inicio del aprendizaje, la máquina plantea una *hipótesis* (modelo inicial) de cómo resolver la tarea. Por ejemplo, en el caso del ajedrez, mover siempre la pieza que esté más a la derecha y al frente de las que tenga disponibles. Probablemente, la hipótesis inicial sea incorrecta, o tenga poco rendimiento, y haya que hacer correcciones de la misma. Si se tiene alguna idea o intuición previa de cómo resolver la tarea, la hipótesis puede tratarse de una aproximación inicial al modelo buscado, que poco a poco se irá *ajustando* con el aprendizaje.

A medida que la máquina aprende, el modelo es modificado (ajustado), para aumentar el rendimiento con que realiza la tarea en cuestión. Para medir el rendimiento del modelo, se observa su salida frente a datos conocidos como datos de validación. Cuando se alcanza un rendimiento razonable, se dice que el aprendizaje ha terminado y se da el modelo por aprendido. Ya no hablaríamos de hipótesis, sino de *modelo*. Ese modelo aprendido es el que se guardará y se empleará para resolver la tarea de forma automática en el futuro.