

Capítulo 4

Principales algoritmos de Machine Learning

1. Lo que vamos a descubrir y requisitos

En el capítulo anterior hemos aprendido o repasado los fundamentos del análisis estadístico descriptivo que, como veremos con la práctica, nos permitirán comprender y preparar nuestros datos antes del aprendizaje. Ahora vamos a conocer los algoritmos principales de Machine Learning que nos permitirán hacer este aprendizaje.

Recuerde que nuestro objetivo al escribir este libro es difundir los conceptos de la inteligencia artificial. Por lo tanto, **no incluiremos las explicaciones teóricas ni matemáticas de todos los algoritmos de aprendizaje.**

Nos quedaremos con una explicación lo más explícita posible ilustrada con uno o varios ejemplos si es necesario. Haciendo una comparación con el mundo del bricolaje, vamos a presentarle las distintas herramientas que hay que utilizar en función del trabajo que vamos a hacer, pero no le explicaremos cómo se han fabricado.

Le recomendamos que considere este capítulo como un cuaderno de notas que puede consultar conforme vaya leyendo para entender por qué utilizamos uno u otro algoritmo y comprender su funcionamiento a grandes rasgos.

■ Observación

Requisitos necesarios para abordar este capítulo correctamente: haber leído el capítulo Estadísticas para comprender los datos.

2. ¿Supervisado o no supervisado? ¿Regresión o clasificación?

Realizar un aprendizaje supervisado consiste en proporcionarle a la máquina datos etiquetados y propicios para el aprendizaje. Es decir, que vamos a analizar y preparar los datos y a darles un significado. La máquina realizará su aprendizaje a partir de este significado. El objetivo es indicarle a la máquina que, para una serie de datos y para una observación precisa, el valor que se ha de predecir es un gato, un perro u otro valor.

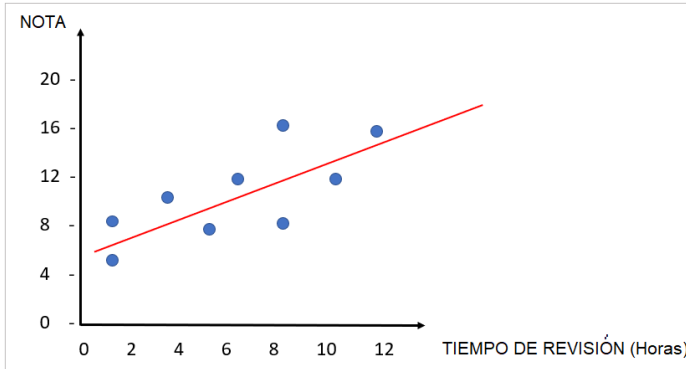
Cuando se trata de predecir un valor, hablaremos de **regresión**; en caso contrario hablaremos de **clasificación**. Predecir el porcentaje de éxito de un equipo de fútbol durante un partido es una regresión; predecir que la foto mostrada es un gato o un perro es una clasificación.

3. Los algoritmos de aprendizaje supervisados por la regresión (predicción de valores)

3.1 Regresión lineal simple (linear regression)

Este algoritmo busca establecer, en la forma de una recta, una relación entre una variable explicada y una variable explicativa. Por ejemplo, predecir una nota de un examen (variable explicada) en función del número de horas de revisiones (variable explicativa).

En otras palabras, los datos de una serie de observaciones se representan en la forma de una nube de puntos y se busca encontrar una recta que pase lo más cerca posible de estos puntos.



Regresión lineal simple

De esta manera, conociendo la cantidad de horas de estudio, podemos predecir de manera aproximada la nota que se obtendrá en el próximo examen.

3.2 Regresión lineal múltiple (Multiple Linear Regression-MLR)

Allí donde utilizamos una sola variable explicativa para explicar otra variable (una nota en función del tiempo de estudio), en la regresión lineal vamos a utilizar varias variables explicativas.

Por ejemplo, queremos predecir el tiempo que va a invertir un ciclista para ganar una etapa del Tour de Francia, en función de su edad, del tiempo que ha invertido en hacer la etapa anterior, de su clasificación en el pelotón y otros datos.

Una fase importante durante el uso de múltiples variables explicativas es su **normalización** (adaptación de la escala). En nuestro ejemplo, es posible que el tiempo invertido en minutos durante la etapa precedente varíe de 160 a 200, y la posición en el pelotón, entre 1 y 80 en función de la cantidad de participantes en el Tour de Francia.

Por lo tanto, no todas las variables explicativas están en la misma escala (de 160 a 200 frente a de 1 a 80).

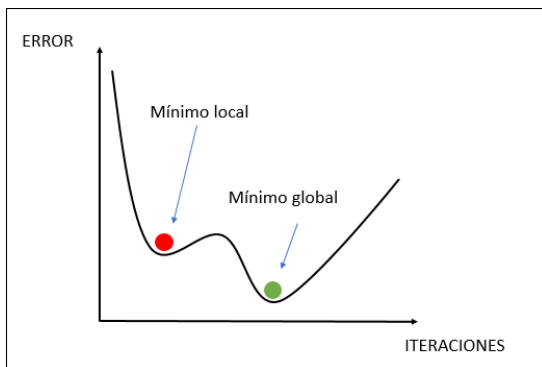
La adaptación a la escala (*scaling*) consistirá en conseguir que la media de cada serie de observaciones sea igual a 0, y que la varianza y la desviación típica sean iguales a 1. Este método también se llama centrar una variable.

Después de esta fase podemos pasar a la predicción gracias al método de **descenso del gradiente** o también al **método de los mínimos cuadrados**. Estos dos métodos toman en cuenta las distintas variables explicativas que se ponen a escala con objeto de predecir la variable explicada.

3.3 Método de descenso del gradiente

Este concepto es fundamental porque se aplica en varios algoritmos de aprendizaje de Machine Learning y de Deep Learning que veremos un poco más adelante en este libro.

Cuando un sistema está en fase de aprendizaje, comete errores. El índice de error disminuye conforme avanza el aprendizaje. Pero es posible que, en un momento dado, el error aumente para volver a disminuir y alcanzar un nivel de error más bajo que el anterior, que es el nivel óptimo de aprendizaje.



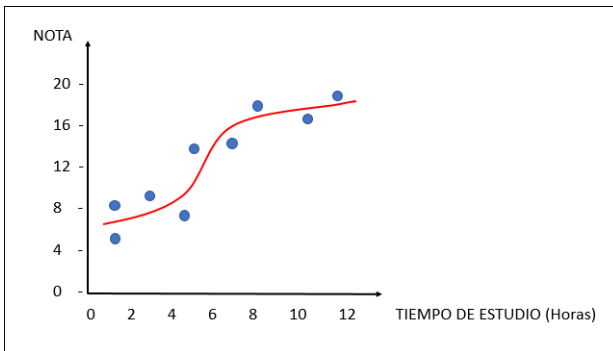
El descenso del gradiente

En la figura anterior podemos comprobar que al principio del aprendizaje el error disminuye de manera progresiva para remontar a continuación. Entonces diríamos que se ha alcanzado el nivel óptimo de aprendizaje, puesto que aparecen errores nuevos. Sin embargo, se puede notar que, después de nuevas iteraciones de aprendizaje, el error sigue disminuyendo para alcanzar un nivel inferior al precedente, llamado mínimo global. Por lo tanto, no se había alcanzado el nivel óptimo de aprendizaje.

El algoritmo del gradiente consiste en encontrar mediante sucesivas iteraciones el mínimo global de la función de coste (error). Utilizando una analogía a menudo empleada en la literatura, imagínese en lo alto de una montaña con el objetivo de llegar a la llanura más abajo. Con cada paso analiza su situación y decide avanzar algunos pasos, sale para volver a subir y tomar el camino que lleva a la meta. El gradiente corresponde a la pendiente del suelo sobre el que camina. De la misma manera, el paso lleva el índice de aprendizaje. Veremos su aplicación práctica en el capítulo La predicción con neuronas.

3.4 Regresión polinomial (polynomial regression)

A veces es difícil encontrar una recta que pueda pasar entre los puntos de la serie de observaciones de manera óptima. Sin embargo, en ocasiones es posible encontrar un vínculo entre las variables con ayuda de una curva. Es lo que permite la regresión polinomial añadiendo pliegues a la curva usando elementos llamados polinomios.



Regresión polinomial

3.4.1 Monomio y polinomio

Un monomio es una expresión matemática que se expresa con esta forma:

$$\propto x^n$$

Donde:

- α (alfa) es un número real o complejo llamado coeficiente del monomio.
- n es un número entero natural que representa el grado del monomio.

Entonces, $5x^2$ es un monomio de coeficiente 5 y de grado 2.

Un polinomio es una suma de monomios.

Por lo tanto, se puede decir que $5x^2 + 2x$ es un polinomio.

3.5 Regresión logística

Como acabamos de ver, cuando los datos no se pueden separar de manera lineal, se pueden utilizar polinomios para darle a nuestra recta la posibilidad de hacer curvas con la finalidad de separar nuestras observaciones.

La regresión logística utiliza una función logística llamada sigmoide o curva S. Este tipo de algoritmo está destinado a aplicarse en los problemas de clasificación.

Hablaremos más detalladamente de la función sigmoide cuando expliquemos las aplicaciones prácticas de las redes neuronales.