

UNIVERSIDAD  
COMPLUTENSE  
DE MADRID



Autor: María José  
Gómez Silva

# Machine Learning con Python. Semana 1.

## CAPÍTULO 1. INTRODUCCIÓN AL MACHINE LEARNING.

### 1.6 Convergencia y Learning Rate

#### CONVERGENCIA

La secuencia de pasos descrita en la sección anterior se repite cíclicamente hasta cumplir una determinada condición que define el final del proceso de aprendizaje. Esta condición podría consistir simplemente en haber realizado un número predeterminado de iteraciones, o puede estar basada en la convergencia del modelo.

Para estudiar la convergencia del modelo, es necesario introducir el concepto de curva de aprendizaje. La *curva de aprendizaje* representa el valor de la función de pérdidas en las sucesivas iteraciones del entrenamiento. El comportamiento deseable es que el valor de la función de pérdidas disminuya a medida que avanza el proceso, como en la Figura 1. Eso indicaría que las predicciones del modelo cada vez se parecen más al objetivo marcado, y por lo tanto, se puede decir que el modelo está aprendiendo.

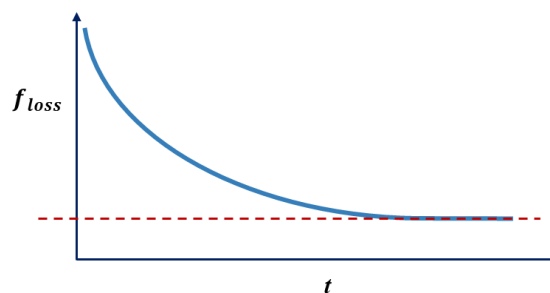


Figura 1. Ejemplo de curva de aprendizaje en la que el modelo converge.

Se dice que el proceso de aprendizaje converge, cuando el valor de las pérdidas varía poco (menos de un cierto umbral), entre sucesivas iteraciones. En la curva de aprendizaje puede identificarse la convergencia como la estabilización del valor de pérdidas en torno a un determinado valor. Si el valor de la función de pérdidas se estabiliza y deja de variar, o varía poco, también lo harán los valores de los parámetros del modelo. Por lo tanto, puede decirse que el aprendizaje ha convergido hacia unos valores concretos de los parámetros o lo que es lo mismo se ha aprendido un determinado modelo.

## LEARNING RATE

La tasa de aprendizaje (learning rate),  $\alpha$ , regula la magnitud de las variaciones (incrementos o decrementos) de los valores de los parámetros del modelo en el momento de su actualización, al final de cada iteración del entrenamiento, tal y como define la siguiente ecuación.

$$p_{t+1} = p_t - \alpha \nabla f_{loss}(p_t)$$

La tasa de aprendizaje es un hiperparámetro. Llamamos *hiperparámetros* a todos aquellos parámetros, que no son los parámetros propios del modelo, pero que regulan su proceso de aprendizaje.

La tasa de aprendizaje determina la magnitud de cada paso dado en la dirección del gradiente descendiente, es decir determina en qué medida o con qué velocidad se avanza en el proceso de aprendizaje.

Si el valor de  $\alpha$  es grande, el avance en cada iteración, hacia un valor mínimo de las pérdidas será grande, y consecuentemente el aprendizaje será rápido (se necesitan pocas iteraciones). Sin embargo, si escogemos un valor de  $\alpha$  demasiado grande, el avance puede adquirir un comportamiento divergente y no llegar a alcanzar el mínimo buscado, como se puede observar en la Figura 2. Si el valor de  $\alpha$  es pequeño, el avance, en cada iteración, hacia un valor mínimo de la función de pérdidas será pequeño, y consecuentemente el aprendizaje será lento (se necesitan muchas iteraciones). Por eso, este hiperparámetro se denomina tasa o ratio de aprendizaje, ya que determina la velocidad con que el modelo aprende.

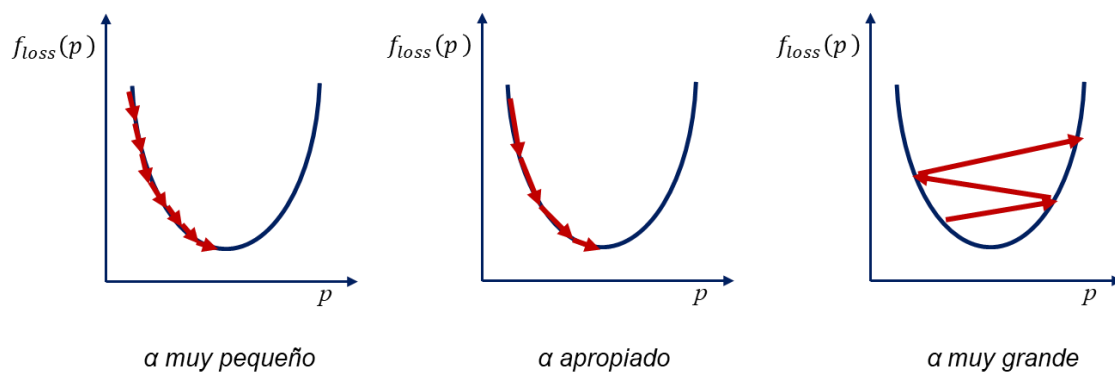


Figura 2. Avance del aprendizaje para distintas tasas de aprendizaje.