

## Journal Name

Crossmark

TAREA 3

RECEIVED  
dd Month yyyy

## Descenso de Gradiente

REVISED  
dd Month yyyyDaniel González<sup>1</sup> <sup>1</sup>Facultad de Ingeniería, Universidad Autónoma de Querétaro, Querétaro, México  
21/08/2025**E-mail:** dgonzalez16@alumnos.uaq.mx**Keywords:** descenso gradiente, plano, mínimo, derivada, epocas**Abstract**

El descenso de gradiente consiste en tomar el principio de la gradiente la cual nos indica en tomar la ruta del punto más alto de una función, sin embargo en este caso al tomar la dirección contraria encontraríamos el punto más bajo. Imaginemos que estamos en una montaña, y quisiéramos encontrar el punto más bajo con esto podríamos avanzar paso a paso en la dirección de mayor pendiente descendente hasta llegar al mínimo.

**1 Introducción**

El descenso de gradiente es un método iterativo para encontrar el mínimo de una función. Dada una función

$$f(\theta),$$

donde  $\theta$  representa los parámetros, el objetivo es encontrar el valor de  $\theta$  que minimiza  $f$ .

- El gradiente es el vector de derivadas parciales:

$$\nabla f(\theta) = \left( \frac{\partial f}{\partial \theta_1}, \frac{\partial f}{\partial \theta_2}, \dots, \frac{\partial f}{\partial \theta_n} \right). \quad (1)$$

- La actualización de parámetros se realiza con la regla:

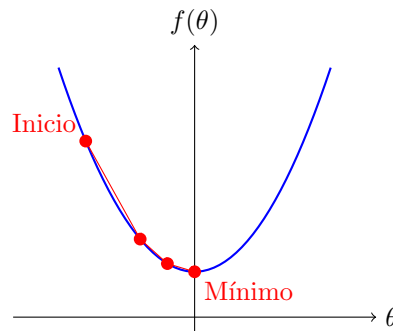
$$\theta \leftarrow \theta - \eta \nabla f(\theta),$$

donde  $\eta > 0$  es la tasa de aprendizaje.

Al repetir este proceso varias veces, los parámetros  $\theta$  se acercan a un mínimo de la función.

En machine learning, la función  $f(\theta)$  corresponde a la función de pérdida o error, que mide qué tan bien predice el modelo. El descenso de gradiente ajusta los parámetros (por ejemplo, los *pesos* de una red neuronal) para minimizar dicho error y así mejorar el desempeño del modelo.

*Ejemplo gráfico*



## 2 Desarrollo

Se plantea la función inicial:

$$\begin{aligned} z &= x^3 + x^2 + y^3 + y^2 \\ z &= \theta_0^3 + \theta_0^2 + \theta_1^3 + \theta_1^2 \end{aligned} \quad (2)$$

Partiendo de la ecuación 1 calculamos la derivada parcial de esta para obtener:

$$\frac{\partial}{\partial x} = 3x^2 + 2x \quad (3)$$

$$\frac{\partial}{\partial y} = 3y^2 + 2y \quad (4)$$

Una vez que tenemos esto, ya podemos obtener nuestra función para calcular el descenso de gradiente. El **gradiente** de  $z$  es:

$$\nabla z(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial z}{\partial \theta_0} \\ \frac{\partial z}{\partial \theta_1} \end{bmatrix} = \begin{bmatrix} 3\theta_0^2 + 2\theta_0 \\ 3\theta_1^2 + 2\theta_1 \end{bmatrix}. \quad (5)$$

Con un *learning rate*  $\eta > 0$ , la regla de actualización iterativa (descenso de gradiente) es:

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - \eta \nabla z(\boldsymbol{\theta}^{(k)}), \quad (6)$$

es decir, componente a componente:

$$\theta_0^{(k+1)} = \theta_0^{(k)} - \eta (3(\theta_0^{(k)})^2 + 2\theta_0^{(k)}), \quad (7)$$

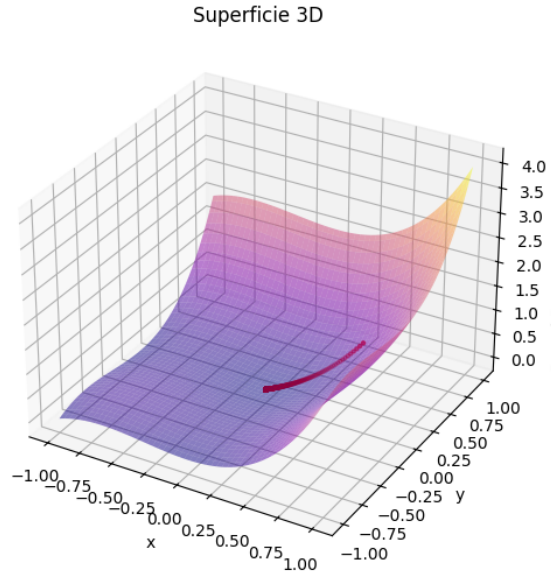
$$\theta_1^{(k+1)} = \theta_1^{(k)} - \eta (3(\theta_1^{(k)})^2 + 2\theta_1^{(k)}). \quad (8)$$

Con esto en mente pasamos al código para poder evaluar el plano de nuestra función, con nuestras  $\theta_0$  y  $\theta_1$  iniciales así como un learning rate que sera  $\alpha$  y un numero de épocas.

Nuestros valores serán:

$$\begin{aligned} \theta_0 &= 0.5 \\ \theta_1 &= 0.5 \\ \alpha &= 0.01 \\ epoch &= 5000 \end{aligned} \quad (9)$$

Con estos valores vamos a mostrar las iteraciones en nuestra gráfica y obtenemos lo siguiente:



**Figure 1.** Grafica del plano de la ecuacion 2 con gradiente descendente apartir de las condiciones 9

Como se puede ver nuestro gradiente tiende a cero.

### 3 Conclusiones

El descenso de gradiente confirma encontrar mínimos en funciones de varias variables. A partir de la función propuesta y del cálculo de sus derivadas parciales, se pudo construir la regla de actualización que, con ayuda de un ciclo de iteraciones, nos guía hacia valores donde el gradiente tiende a cero como se ve en la figura 1.

Durante el desarrollo podemos ver cómo la elección de una tasa de aprendizaje adecuada y un número suficiente de épocas permiten visualizar la convergencia de los parámetros al mínimo local de la función. También se destacó la relevancia de este procedimiento en machine learning, donde el descenso de gradiente se utiliza para ajustar los parámetros de un modelo y minimizar la función de pérdida, mejorando así su capacidad predictiva.

Con esto es posible demostramos la base matemática así como la aplicación práctica del descenso de gradiente, la cual es una herramienta fundamental en optimización y aprendizaje automático.