OBS Business School

RED DE EDUCACIÓN SUPERIOR
Planeta Formación y Universidades

Moisés Cantón Jara

Fundamentos de Matemáticas

Cálculo y Optimización en IA

Partner académico | unie Universidad

Cálculo y Optimización en IA

1. Introducción

El cálculo, como uno de los pilares fundamentales de las matemáticas, juega un papel esencial en el mundo del ML y la Inteligencia Artificial. A través de este tema, exploraremos las herramientas y conceptos fundamentales del cálculo que le permitirán abordar problemas de optimización con confianza y precisión.

2. Fundamentos de Cálculo[i]

¿Qué es una variable? ¿y un algoritmo?

Un concepto fundamental en matemáticas y programación es la variable, que actúa como un símbolo o espacio para almacenar valores en distintos contextos:

- En el ámbito de las matemáticas, las variables son utilizadas para representar cantidades que pueden tomar diversos valores. Por ejemplo, en un estudio de alturas de personas, una variable podría representar la altura de un individuo en particular.
- En programación, las variables son nombres que se asocian con valores almacenados en la memoria de una computadora. En un programa de calculadora, por ejemplo, la variable "x" podría representar el número ingresado como entrada.

Paralelamente, los algoritmos son elementos esenciales en la ciencia de la computación, ya que son conjuntos de instrucciones precisas y ordenadas que se emplean para resolver problemas o realizar tareas específicas. Estos algoritmos desempeñan un papel crucial en diversas aplicaciones, desde el procesamiento de datos hasta la toma de decisiones y la automatización de tareas.

Algunas de las características clave de los algoritmos son:

- Precisión: Los algoritmos deben definirse de manera precisa y detallada, con cada paso claramente especificado para evitar ambigüedades en su ejecución.
- Orden: Los pasos de un algoritmo deben seguir un orden específico, estableciendo qué acción se realiza primero, cuál continúa después y así sucesivamente.
- Entrada y Salida: Los algoritmos operan a partir de entradas (datos o información) y producen salidas (resultados) en función de esas entradas. Los pasos del algoritmo describen cómo se transforman las entradas en salidas.
- Finitud: Los algoritmos deben concluir en un número finito de pasos; no pueden quedar atrapados en bucles infinitos ni ejecutarse indefinidamente.
- Efectividad: Los algoritmos deben ser efectivos, es decir, capaces de resolver el problema para el cual fueron diseñados en un tiempo razonable.

Ecuaciones y Funciones

Una ecuación se define como una igualdad matemática que contiene una o más incógnitas, representadas como variables, y expresa la relación en la que dos expresiones o valores son iguales. Las ecuaciones tienen un propósito fundamental en la resolución de problemas matemáticos, ya que permiten determinar el valor desconocido de las incógnitas que satisface la igualdad. Por ejemplo, consideremos la ecuación elemental $2x = 8$, donde x es la incógnita; su solución se obtiene como $x = \frac{8}{2} = 4$, lo que significa que 2 multiplicado por 4 es igual a 8.

Por otro lado, una función se define como una relación matemática que asigna de manera unívoca cada elemento de un conjunto de entrada, denominado dominio, a un elemento correspondiente en otro conjunto. En términos más simples, una función toma uno o varios valores de entrada y genera un valor de salida específico.

Las funciones suelen representarse mediante notación como f(x) o y = f(x), donde x representa la variable de entrada y f(x) denota el valor de salida correspondiente. Por ejemplo, en la función $f(x) = 2x$, se toma un valor de entrada, x, y se produce un valor de salida igual al doble del valor de x.

Simplificando, sus principales diferentes son:

- Las ecuaciones no necesariamente tienen una variable de entrada específica y pueden contener múltiples incógnitas
- Las funciones se utilizan para representar relaciones y modelar situaciones en matemáticas y ciencia

En resumen, una ecuación es una igualdad matemática con incógnitas que se resuelve para encontrar valores específicos, mientras que una función describe una relación matemática entre entradas y salidas, y se utiliza para mapear valores de entrada a valores de salida en una forma organizada y consistente. Las funciones pueden definirse mediante ecuaciones, pero no todas las ecuaciones representan funciones.

Significado y Cálculo de la pendiente en funciones

La pendiente de una función es una medida que describe la inclinación o tasa de cambio de una función en un punto específico o entre dos puntos de su gráfico y representa cuanto cambia la variable de salida en relación con un cambio en la variable de entrada.

$$p = \frac{\Delta y}{\Delta \mathrm{x}}$$

Simplificando, la pendiente mide cuando asciende o desciende la función en relación con un cambio unitario en la variable de entrada. Algunos puntos clave sobre ella:

- Dirección de la Pendiente: La pendiente puede ser positiva, negativa o cero. Una pendiente positiva indica que la función aumenta a medida que la variable de entrada aumenta, mientras que una pendiente negativa indica que la función disminuye a medida que la variable de entrada aumenta. Una pendiente de cero indica que la función es constante en ese intervalo.
- Magnitud de la Pendiente: Cuanto mayor sea la magnitud de la pendiente (ya sea positiva o negativa), más empinada será la función en ese punto o intervalo. Una pendiente más grande significa que la función cambia más rápidamente en respuesta a cambios en la variable de entrada.
- Interpretación: La pendiente tiene significados específicos en diferentes contextos. Por ejemplo, en un gráfico de distancia contra tiempo, la pendiente representa la velocidad. En un gráfico de cantidad de ventas contra tiempo, la pendiente representa la tasa de cambio de ventas por unidad de tiempo.
- Derivada: En cálculo, la derivada de una función en un punto dado proporciona la pendiente de la función en ese punto. Esto es fundamental para analizar funciones y calcular tasas de cambio instantáneas.

Fundamentos y Calculo de Derivadas[ii]

Anteriormente mencionada, la derivada de una función f(x) en un punto x_0 representa la tasa de cambio instantánea de la función en ese punto. Matemáticamente, se denota como $f'(x_0)$ o $\frac{df}{dx}$.

Normalmente, a la hora de calcularlas, se utiliza la regla de potencias como punto de partida. Para una función $f(x)$, la derivada de x_n es $nx^{(n-1)}$, por ejemplo, para derivar la función $f(x) = 3x^2 - 2x + 1$, aplicamos la regla de potencias para cada término:

$$\frac{d}{dx}(3x^2) - \frac{d}{dx}(2x) + \frac{d}{dx}(1) = 6x - 2$$

A partir de la regla anterior, se pueden derivar funciones más complejas utilizando reglas adicionales, como la regla de la suma, la regla del producto y la regla de la cadena.

python:
import sympy as sp x = sp.symbols('x') f = sp.sin(x) + x**2 f prime = sp.diff(f, x) print(f_prime)

Teoría de Integrales y Tipos (Definidas e Indefinidas)

Una integral indefinida, también conocida como antiderivada, es la operación inversa a la derivación y se denota como: $\int f(x)dx$.

El resultado de una integral indefinida es una familia de funciones, ya que se agrega una constante arbitraria de integración, denotada comúnmente como C, debido a que una derivada única no permite determinar completamente la función original.

$$\int \mathrm{f(x)\ dx} = F(x) + C$$

donde F(x) representa la función antiderivada de f(x) y C es la constante de integración.

Un ejemplo del cálculo de una antiderivada mediante la regla de las potencias inversa sería:

$$\int (2x+3)dx = \int 2xdx + \int 3dx = x^2 + 3x + C$$

donde C es una constante de integración.

Si concretamos las integrales indefinidas a un intervalo concreto tendremos un nuevo concepto, las integrales definidas que representan el área bajo la curva entre dos puntos específicos del eje x, y se denota como sigue $\int_a^b \mathrm{f(x)\ dx}$, donde a y b son los límites de integración.
Geométricamente, la integral definida se interpreta como el área entre la curva de la función y el eje x en el intervalo [a,b]. Si la función está por encima del eje x en ese intervalo, el valor será positivo; si está por debajo, será negativo

python:

f_integral = sp.integrate(f, x)
print(f_integral) # Indefinida
f_integral_def = sp.integrate(f, (x, 0, 1)) # Entre 0 y 1

```
print(f_integral_def)
```

Teorema Fundamental del Calculo

El Teorema Fundamental del cálculo establece es uno de los resultados más fundamentales en Cálculo Integral. Desarrollo por Newton y Leibniz en el siglo XVII establece una conexión profunda entre la derivación e integración de funciones.

a) Derivación de una Integral Definida

Su primera parte establece que si tienes una función continua f(x) definida en un intervalo cerrado [a,b], y si F(x) es cualquier función primitiva (antiderivada) de f(x) en ese intervalo, entonces la integral definida de f(x) en ese intervalo se puede calcular como la diferencia entre los valores de F(x) en los extremos del intervalo:

$$\int_a^b \mathrm{f(x)\,dx} = F(b) - F(a)$$

lo que significa que puedes encontrar el valor de una integral definida simplemente evaluando la antiderivada en los límites de integración y tomando la diferencia.

b) Cálculo de Derivadas de Integrales Definidas

Su segunda parte indica que que si tienes una función f(x) continua en un intervalo que incluye a y una función F(x) es una antiderivada de f(x) en ese intervalo, entonces la derivada de la integral definida de f(x) con respecto a un límite (por ejemplo, b) es igual a la función f(x) evaluada en ese límite:

$$\frac{d}{dx}\left(\int_a^x \mathrm{f(t)\,dt}\right) = f(x)$$

Esto significa que puedes calcular la derivada de una integral definida simplemente evaluando la función f(x) en el límite superior de integración.

Explorando la integración numérica

A mayores de lo que comentado anteriormente, si no es posible calcular analíticamente una integral, existen varios métodos de integración numérica que nos permiten obtener aproximaciones numéricas a la integral definida de una función.

Algunos de estos métodos son:

- **Sumas de Riemann:**
- **Concepto**: Las Sumas de Riemann dividen un intervalo en pequeños subintervalos y aproximan el área bajo la curva de la función utilizando valores específicos en cada subintervalo.
- **Fórmula**: Para una partición $a = x_0 < x_1 < \cdots < x_n = b$ del intervalo [a,b], la suma de Riemann se define como:

$$\int_a^b \mathrm{f(x)\,dx} \approx \sum_{i=1}^{n} f(x_i^*) * \Delta x_i$$

Donde x_i^* es un punto en el i-ésimo subintervalo y Δx_i es la longitud de ese subintervalo.

- **Ejemplo**: Aproximemos la integral definida de $f(x) = x^2$ en el intervalo [0,1] utilizando la suma de Riemann con cuatro subintervalos.

$$\int_0^1 \mathrm{x}^2\,\mathrm{dx} = \left(\frac{1}{16} * 0^2\right) + \left(\frac{1}{16} * \left(\frac{1}{4}\right)^2\right) + \left(\frac{1}{16} * \left(\frac{1}{2}\right)^2\right) + \left(\frac{1}{16} * 1^2\right) = \frac{1}{4}$$

- **Regla del Trapecio:**
- **Concepto**: La Regla del Trapecio aproxima el área bajo la curva de la función dividiendo el intervalo en trapecios y calculando el área de cada trapecio.
- **Fórmula**: Para una partición $a = x_0 < x_1 < \cdots < x_n = b$ del intervalo [a,b], la aproximación de la integral definida utilizando la Regla del Trapecio es:

$$\int_a^b \mathrm{f(x)}\,\mathrm{dx} \approx \frac{h}{2}\sum_{i=1}^{n}[f(x_{i-1}) + f(x_i)]$$

Donde h es la longitud de cada subintervalo $h = \frac{b-a}{n}$.

- **Ejemplo**: Aproximemos la integral definida de $f(x) = x^2$ en el intervalo [0,1] utilizando la Regla del Trapecio con cuatro subintervalos.

$$\int_0^1 \mathrm{x}^2\,\mathrm{dx} \approx \frac{1}{8}\left(0 + 2\left(\frac{1}{4}\right) + 2\left(\frac{1}{2}\right) + 1\right) = \frac{5}{8}$$

- **Regla de Simpson:**
- **Concepto**: La Regla de Simpson utiliza segmentos de parábola para aproximar el área bajo la curva de la función.
- **Fórmula**: Para una partición $a = x_0 < x_1 < \cdots < x_n = b$ (con un numero par de subintervalos) del intervalo [a,b], la aproximación de la integral definida con la Regla de Simpson es:

$$\int_a^b \mathrm{f(x)}\,\mathrm{dx} \approx \frac{h}{3}[f(x_0) + 4f(x_1) + 2(x_2) + 4f(x_3) + \cdots + 2f(x_{n-2}) + 4f(x_n)]$$

Donde h es la longitud de cada subintervalo $h = \frac{b-a}{n}$.

- **Ejemplo**: Aproximemos la integral definida de $f(x) = x^2$ en el intervalo [0,1] utilizando la Regla de Simpson con cuatro subintervalos.

$$\int_0^1 \mathrm{x}^2\,\mathrm{dx} \approx \frac{1}{3}\left(0 + 4\left(\frac{1}{4}\right) + 2\left(\frac{1}{2}\right) + 4\left(\frac{3}{4}\right) + 1\right) = \frac{1}{3}$$

Al ser aproximaciones, es útil introducir el concepto de error aproximación que indica la diferencia entre el valor real de la integral y la aproximación numérica obtenida por el método. A la hora de evaluar la precisión numérica es habitual calcular el error absoluto o el error relativo.

- **Error Absoluto** (E_a) que se define como la diferencia entre el valor real de la integral (I) y la aproximación numérica (In):

$$E_a = |I - I_n|$$

- **Error Relativo** (E_r) que es el resultado de dividir el error absoluto dividido por el valor real de la integral, generalmente multiplicado por 100 para expresarlo como un porcentaje:

$$E_r = \left|\frac{I - I_n}{I}\right| * 100\%$$

Sí suponemos que conocemos el valor real de la integral definida de $f(x) = x^2$ en el intervalo [0,1], que es $\frac{1}{3}$, y hemos calculado una aproximación numérica de $\frac{5}{8}$ utilizando la Regla del Trapecio, el error absoluto y el error relativo serán como sigue:

$$E_a = |I - I_n| = \left|\frac{1}{3} - \frac{5}{8}\right| = \left|\frac{8}{24} - \frac{15}{24}\right| = \frac{7}{24} \approx 0.2917$$

$$E_r = \left|\frac{I - I_n}{I}\right| * 100\% = \left|\frac{\frac{1}{3} - \frac{5}{8}}{\frac{1}{3}}\right| * 100\% = \left|\frac{\frac{7}{24}}{\frac{1}{3}}\right| * 100\% \approx 87.5\%$$

3. Fundamentos de Optimización

Aspectos Claves de la Optimización

A la hora de trabajar con modelos de Machine Learning, nos veremos obligados en diversos casos a encontrar, por ejemplo, los valores que maximicen ciertos parámetros del modelo o que minimicen el error. Al acto de encontrar esos valores máximos o mínimos de una función, se le conoce como optimizar una función.

A estos puntos se los conoce como puntos críticos o extremos y se clasifican según su localización en locales y globales.

- Locales. Un máximo local es el valor más alto que puede alcanzar en un intervalo pequeño alrededor de un punto en particular, mientras que un mínimo es lo opuesto.
  - Estos extremos se pueden encontrar mediante el cálculo de derivadas buscando los puntos donde esta sea 0, aunque esto no es condición suficiente para su existencia, por lo que se refieren pruebas adicionales como la de la segunda derivada.
- Globales. Un máximo global es el valor más alto que alcanza una función en todo su dominio, siendo un mínimo global lo contrario.
  - A diferencia de para máximos y mínimos locales, encontrar máximos y mínimos globales es más complejo y no se garantiza que haya una solución única.
  - A la hora de calcular estos máximos y mínimos globales, se suelen emplear técnicas como el descenso del gradiente que mencionaremos más adelante en el módulo.

Términos y Requisitos para la Identificación de Puntos Críticos en Funciones

A continuación, se enumeran algunas de las condiciones fundamentales para la existencia de estos puntos:

- Continuidad. La función debe ser continua en el intervalo donde se están buscando los puntos críticos, es decir, no debe haber discontinuidades, ni saltos abruptos en la función.
- Diferenciabilidad de la función. La función debe tener derivadas en el punto que estamos analizando.

- Derivada igual a 0. Un punto crítico se encuentra donde la derivada de la función es igual a cero (f′(x)=0).
- Prueba de la segunda derivada. Si la segunda derivada en el punto crítico es positiva, el punto es un mínimo local; si es negativa, es un máximo local; y si es igual a cero, la prueba no concluye y se necesita un análisis adicional.

Optimización Multivariada: Derivadas Parciales y Gradientes

La optimización multivariada consiste en la búsqueda de máximos y mínimos de una función de múltiples variables, representada como $f(x_1, x_2, \ldots, x_n)$.

Aquí están los conceptos clave en relación con la optimización univariada:

- Máximos y Mínimos Locales. Al igual que para Optimización univariada, los máximos y mínimos locales representan valores donde la función es mayor o menor que en los puntos cercanos. Sin embargo, aquí hay que considerar múltiples direcciones.
- El gradiente de una función es vector que apunta en la dirección de la máxima tasa de cambio.
- Los puntos críticos en la optimización multivariada son aquellos en los que el gradiete de la función es igual a cero ($\nabla f(x) = 0$) o donde no existe el gradiente.
- El hessiano es una matriz que describe cómo cambian las derivadas parciales del gradiente.

En este contexto, se definen

- derivada parcial al cálculo que nos indica cómo cambia una función de varias variables respecto a una de ellas manteniendo el resto constante y se denotan como $\frac{\partial f}{\partial x}$ (derivada parcial con respecto a x) y $\frac{\partial f}{\partial y}$ (derivada parcial con respecto a y)
- gradiente como un vector que contienen todas las derivadas parciales de una función de varias variables y para función f(x,y), el gradiente se denota como $\nabla f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$

Tomando, la función $f(x, y) = x^2 + 2xy + y^2$, como ejemplo, tendríamos:

1. $\frac{\partial f}{\partial x}$:

Tratamos x como la variable de interés y y como una constante.

$$\frac{\partial f}{\partial x} = 2x + 2y$$

2. $\frac{\partial f}{\partial y}$

$$\frac{\partial f}{\partial y} = 2x + 2y$$

3. ∇f:

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (2x + 2y, 2x + 2y)$$

4. Descenso del Gradiente

Función Objetivo y Restricciones

A la hora de afrontar problemas de cálculo y optimización, conviene destacar los conceptos de funciones objetivo y las restricciones:

- Una función objetivo es una función matemática que se emplea para expresa el objetivo principal de un problema de optimización. Este objetivo puede consistir en maximizar o minimizar esta función en función de las variables del problema.
    - Algunos ejemplos de funciones objetivo pueden ser: la maximización de las ganancias, la minimización de los costos, la maximización de la eficiencia, la minimización de la distancia de viaje, … y su elección es fundamental.
- Las restricciones son condiciones adicionales que limitan las soluciones viables del problema de optimización. Estas restricciones pueden ser ecuaciones o desigualdades que las variables deben cumplir.
    - Algunos ejemplos de restricciones incluyen restricciones de recursos, limitaciones de capacidad, restricciones de tiempo, restricciones de calidad, …

Método de Descenso del Gradiente[iii]

El Método de Descenso del Gradiente es una técnica fundamental en el campo de la optimización numérica y desempeña un papel crucial en una amplia variedad de aplicaciones, incluyendo el aprendizaje automático y la inteligencia artificial. En su esencia, este método se utiliza para encontrar el mínimo (o máximo) de una función, lo que lo convierte en una herramienta esencial en la resolución de problemas de optimización.

La idea detrás del Descenso del Gradiente es relativamente simple pero poderosa. En lugar de buscar directamente el mínimo de una función en todo su dominio, el algoritmo se inicia desde un punto inicial y utiliza información local, en forma del vector gradiente, para tomar pasos sucesivos hacia el mínimo. El gradiente en un punto dado señala la dirección en la que la función crece más rápidamente, por lo que avanzar en la dirección opuesta del gradiente nos acerca al mínimo deseado.

A lo largo de una serie de iteraciones, el algoritmo ajusta gradualmente la posición en la que se encuentra, lo que lo hace converger hacia un mínimo local o global de la función objetivo. Este proceso iterativo continúa hasta que se cumple un criterio de parada predefinido, como un número máximo de iteraciones o una pequeña diferencia entre valores sucesivos de la función.

A continuación, se describen a alto nivel los pasos que este método nos requiere seguir antes de continuar con un par de ejemplos.

1) Inicialización:
    a. Se comienza seleccionando un punto de inicio que denominaremos x_0 en el espacio de búsqueda. Este punto puede elegirse aleatoriamente o en base a algún conocimiento sobre el problema que nos indique que el mínimo se encuentra en su entorno.
    b. Se define una tasa de aprendizaje que representará un hiperpárametro que controla el tamaño de los pasos que dará el algoritmo en cada iteración.
    c. Se especifica un criterio de parada, como un número máximo de iteraciones o una pequeña diferencia entre los valores de la función objetivo en iteraciones sucesivas.
2) Cálculo del Gradiente.
    a. Se calcula el gradiente de la función objetivo en el punto actual x_k, siendo este un vector que indica la dirección y la magnitud del cambio más pronunciado de la función en ese punto que se denomina como $\nabla f(x_k)$
    b. El gradiente se calcula tomando las derivadas parciales de la función objetivo con respecto a cada una de las variables si la función es multivariable.
3) Actualización

a. Utiliza el gradiente calculado para ajustar el valor de x_k en la dirección opuesta al gradiente. Esto se hace para moverse hacia un mínimo local o global.
b. La actualización se realiza mediante la siguiente formula:

$$x_{k+1} = x_k + \alpha * \nabla f(x_k)$$

4) Convergencia.
    a. Verifica si se cumple algún criterio de parada. Los criterios de parada típicos pueden incluir:
        i. Un número máximo de iteraciones alcanzado.
        ii. Una pequeña diferencia entre los valores de la función objetivo en dos iteraciones sucesivas (indicando convergencia).
    b. Si se cumple el criterio de parada, el algoritmo se detiene. De lo contrario, regresa al Paso 2 para calcular el gradiente en el nuevo punto x_{k+1} y repetir el proceso.
5) Resultado
    a. El algoritmo se termina cuando se alcanza un mínimo local o global (o cuando se cumple el criterio de parada).
    b. El valor de x en el final es una aproximación al mínimo de la función objetivo.

Si tomamos como ejemplo una función objetivo univariable como $f(x) = 2x^2 - 4x + 5$. El objetivo será encontrar el mínimo de la función objetivo.

1) Inicializacion
    a. Empezamos con un valor inicial para nuestra variable, denotado como x_0. Supongamos que inicializamos con $x_0 = 3$.
    b. También definimos la tasa de aprendizaje como $\alpha = 0.1$
    c. Y especificamos un criterio de parada, como el número máximo de iteraciones o una pequeña diferencia entre valores sucesivos.
2) Cálculo del gradiente
    a. Calculamos la derivada en el punto actual x_0 para obtener el gradiente $f'(x_0)$. Para nuestra función, $f'(x) = 4x - 4$
3) Actualización
    a. Utilizamos la formula del descenso de gradiente para actualizar x, donde sustituyendo tenemos:

$$x_1 = x_0 - \alpha * f'(x_0) = 3 - 0.1(4 * 3 - 4) = 2.2$$

4) Convergencia
    a. Comprobamos si hemos alcanzado el criterio de parada. Si no, repetimos los pasos 2 y 3.
    b. Supongamos que hemos decidido detenernos después de 10 iteraciones. Continuamos iterando hasta que alcancemos ese límite o hasta que se cumpla nuestro criterio de parada.
5) Iteraciones posteriores
    a. Iteramos el proceso utilizando x_1 como punto de partida. Calculamos $f'(x_1)$, actualizamos x a x_2, y así sucesivamente.
    b. Después de 10 iteraciones, supongamos que hemos obtenido $x_{10} \sim 0.9999$ como nuestro valor final.

Y si, en cambio, tomamos, como ejemplo, la siguiente función objetivo multivariable $f(x, y) = x^2 + y^2$, y queremos encontrar el mínimo de esta función utilizando el método del descenso del gradiente tendremos:

1) Inicializacion

   a. Empezamos con un valor inicial para nuestra variable, denotado como x_0. Supongamos que inicializamos con $(x_0, y_0) = (3,4)$.
   b. También definimos la tasa de aprendizaje como $\alpha = 0.1$
   c. Y especificamos un criterio de parada, como el número máximo de iteraciones o una pequeña diferencia entre valores sucesivos.

2) Cálculo del gradiente
   a. Calculamos la derivada en el punto actual x_0 para obtener el gradiente $\nabla f(x, y)$. Para nuestra función, $\nabla f(x, y) = (2x, 2y)$.
3) Actualización
   a. Utilizamos la formula del descenso de gradiente para actualizar x, donde sustituyendo tenemos:

$$x_1 = x_0 - \alpha \left(\frac{\partial f}{\partial x}\right)_{x_0} = x_0 - 2\alpha x_0 = 3 - 2(0.01)(3) = 2.94$$
$$y_1 = y_0 - \alpha \left(\frac{\partial f}{\partial y}\right)_{y_0} = y_0 - 2\alpha y_0 = 4 - 2(0.01)(4) = 3.92$$

4) Convergencia
   a. Comprobamos si hemos alcanzado el criterio de parada. Si no, repetimos los pasos 2 y 3.
   b. Supongamos que hemos decidido detenernos después de 10 iteraciones. Continuamos iterando hasta que alcancemos ese límite o hasta que se cumpla nuestro criterio de parada.
5) Iteraciones posteriores
   a. Iteramos el proceso utilizando (x_1, y_1) como punto de partida. Calculamos $f'(x_1, y_2)$, actualizamos (x, y) a (x_2, y_2), y así sucesivamente.
   b. Después de 10 iteraciones, supongamos que hemos obtenido $(x_{10}, y_{10}) = (0.362388, 0.429497)$ como nuestro valor final, donde comprobamos como los valores de x, y se han ido acercando al mínimo global (0,0).

Aplicación: Descenso del Gradiente en Python

A modo de consolidación de nuestra comprensión del descenso del gradiente, consideremos la función $f(x) = x^2$, por simplicidad, ya que su mínimo global es en 0.

Implementaremos el descenso del gradiente para esta función y veamos cómo converge al mínimo.

python:

```
import numpy as np

def gradient_of_f(x):
    return 2*x

alpha = 0.1  # Tasa de aprendizaje
x = 5  # Punto inicial
num_iterations = 50

for i in range(num_iterations):
    x = x - alpha * gradient_of_f(x)
```

```
print(x)  # Debería estar cerca de 0, que es el mínimo de f(x) = x^2
```

[i] Ron Larson, Robert P. Hostetler, Bruce H. Edwards. (2006). Cálculo, 8a ed. Mc. Graw-Hill
[ii] G. Bradley y K.J. Smith. (1998). Cálculo de una variable, Vol. 1, Ed. Prentice Hall.
[iii] Boyd, S., & Vandenberghe, L. (2004). Convex Optimization. Cambridge University Press.