Cálculo

Conceptualización: Derivadas - Optimización

Sergio A. Cantillo
sacantillo@uao.edu.co

Facultad de Ingeniería y Ciencias Básicas
Universidad Autónoma de Occidente

Agenda

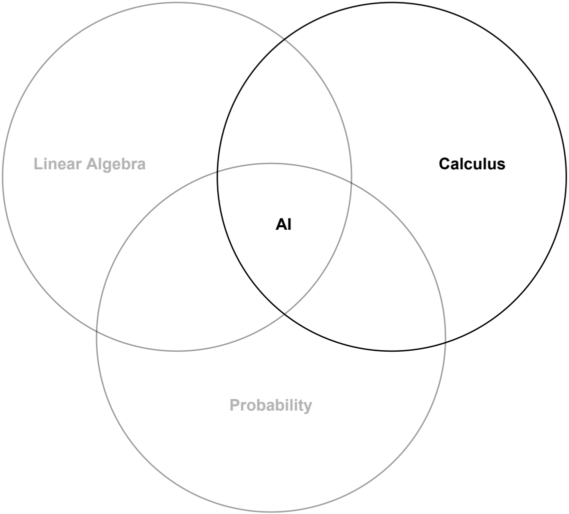
Agenda

Agenda

Agenda

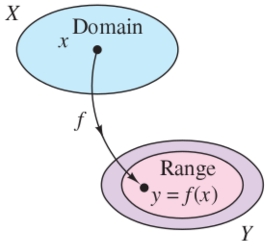
1. Las Matematicas en la IA
2. Cálculo
   - Funciones
   - Derivacion
   - Funciones de Varias Variables
   - Derivadas Parciales
3. Conceptos de Optimización
   - Optimización
   - Gradiente y Matriz Hessiana
   - Algoritmos de Optimización
     - Algoritmos de Gradiente
4. Referencias

Las Matemáticas en la IA



https://towardsdatascience.com/mathematics-for-ai-all-the-essential-math-topics-you-need-ed1d9c910baf

Agenda



Función Real

Una función de X a Y es una relación de correspondencia que asigna a cada número $x \in \mathbb{R}^D$ en X exactamente **un número** $y \in \mathbb{R}$ en Y.

- **Dominio:** conjunto de todos los números reales para los que esta definida la función.
- **Imagen/Codominio:** se denota mediante $f(x)$ y se le llama el valor de f **en** x.
- **Rango:** subconjunto de Y formado por todas las imágenes de los números de X.

Ejemplo

$$x^2 + 2y = 1 \quad \longrightarrow \quad y = \frac{1}{2}(1 - x^2) \quad \longrightarrow \quad f(x) = \frac{1}{2}(1 - x^2)$$

Funciones de una Variable

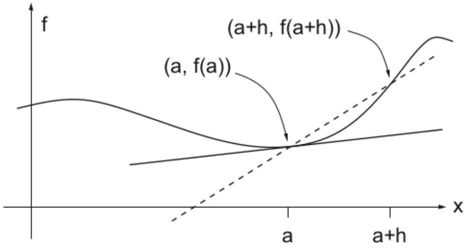
Ejemplo

Dadas las funciones:

1. $f(x) = \sqrt{x-1}$
2. $f(x) = \dfrac{1}{x^2-4} \quad ; \qquad 4 \leq x \leq 5$
3. $f(x) = \begin{cases} 1-x, & \text{if } x < 1 \\ \sqrt{x-1}, & \text{if } x \geq 1 \end{cases}$

a. Calcular el domino y el rango.
b. Realizar la gráfica en Python usando la librería Matplotlib.

Agenda



la pendiente de la recta tangente a la función f en el punto $(a, f(a))$ indica la pendiente de la grafica de f en $x = a$.

Ejemplo

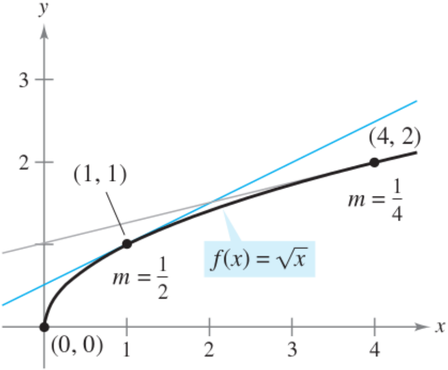
Graficar en Python las funciones $f(x) = 2x^3 - 4x^2 + 3x + 5$; $y = x - 5$; $y = 2x - 5$ y $y = 3x - 5$; analizar el punto $(0, -5)$

Derivada de una Función

Definición

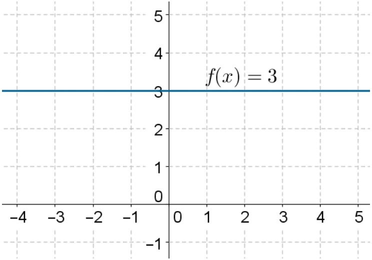
La derivada de una función proporciona **la pendiente de la recta tangente** a la grafica de f en el punto $(x, f(x))$, siempre y cuando exista.

Calcular la pendiente de una funcion en un punto:



$$f(x) = \sqrt{x} = x^{\frac{1}{2}}$$

$$\boxed{m = f'(x) = \frac{1}{2\sqrt{x}}}$$



Definición

La derivada de cualquier valor constante siempre será igual a **cero**.

$$f(x) = c$$
$$f'(x) = 0$$

Ejemplos

Calcular las derivadas:

1. $s(r) = -5$
2. $y(t) = k\pi$, donde k es una constante

Definición

Para cualquier entero n al que este elevado una variable a derivar, esta bajará a multiplicar la expresión y su exponente se reduce en una unidad.

$$f(x) = x^n$$
$$f'(x) = nx^{n-1}$$

Ejemplo

Calcular las derivadas y graficar en Python:

1. $y = x$
2. $y = x^2$
3. $y = \dfrac{1}{x^3}$

Definición

Si $f(x)$ y $g(x)$ son diferenciables en x, entonces su producto es diferenciable en x.

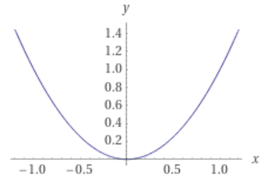
Sea $h(x) = f(x) \cdot g(x)$:

$$h'(x) = f(x) \cdot g'(x) + f'(x) \cdot g(x)$$

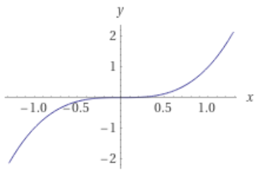
Ejemplos

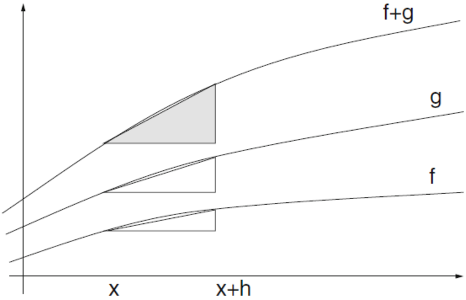
Calcular la derivada y graficar en Python:

$f(x) = x^2$



$f(x) = x^3$





Definición

Si f y g son diferenciables en x, y c es cualquier constante, entonces $f+g$, $f-g$ y cf son diferenciables en x.

$$(f+g)'(x) = f'(x) + g'(x)$$
$$(f-g)'(x) = f'(x) - g'(x)$$
$$(cf)'(x) = cf'(x)$$

Ejemplo

Calcular la derivada y graficar en Python.

$$p(x) = 2x^4 + 7x^3 + 6x^2 + 2x + 15$$

Definición

Si f es diferenciable en x y $f(x) \neq 0$, entonces $\frac{1}{f}$ es diferenciable en x

$$\left(\frac{1}{f}\right)'(x) = -\frac{f'(x)}{(f(x))^2}$$

Ejemplo

Calcular la derivada de y graficar en Python de la función:

$$f(x) = \frac{1}{x^2+1}$$

Definición

Si f y g son diferenciables en x y $g(x) \neq 0$, entonces su cociente es diferenciable en x,

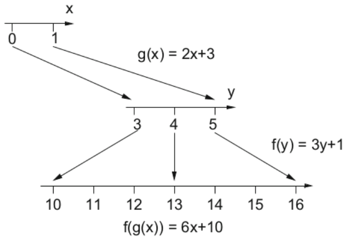
$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2}$$

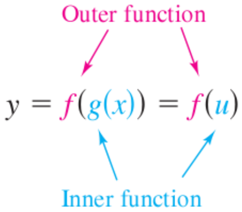
Ejemplo

Sea $f(x) = x$ y $g(x) = x^2 + 1$, calcular la derivada y realizar la gráfica en Python.

Funcion Compuesta

Sean f y g dos funciones. La función dada por $(f \circ g)(x) = f(g(x))$ se llama función **compuesta**.





Si f es diferenciable en y y $y = g(x)$ es diferenciable en x, entonces

$$(f \circ g)'(x) = f'(g(x))g'(x)$$

Regla de la Cadena

Ejemplo

Calcular la derivada de:

$$y = \left(x^2 - x + 5\right)^4$$

$$y = f(g(x))$$
$$u = g(x)$$
$$y = f(u)$$

Solución:

$$\frac{dy}{dx} = 4(x^2 - x + 5)^3(2x - 1)$$

$\frac{dy}{du}$; derivada de y con respecto a u
$\frac{du}{dx}$; derivada de u con respecto a x

Derivadas de Orden Superior

Definición

Sea f una función diferenciable, entonces f' es la **primera derivada** de f. Si $f'(x)$ tambien es derivable, entonces $f''(x)$ se le denomina **segunda derivada** de la función primitiva f. De esta manera es posible obtener derivadas de mayor orden.

$$f'(x) = y' = \frac{dy}{dx} = \frac{d}{dx}f(x)$$

Ejemplo

Calcular las derivadas de orden 1,2,3,4 y 5 para la función $f(x) = x^4$

~~Solución:~~

$$f'(x) = 4x^3$$
$$f''(x) = 12x^2$$
$$f'''(x) = 24x$$

$$f^{(4)}(x) = 24$$
$$f^{(5)}(x) = 0$$

Agenda

Funciones de Varias Variables

Definición

Una función es una relación entre dos conjuntos donde a cada elemento del primer conjunto le corresponde **un solo elemento** del segundo conjunto. En una función de varias variables, la variable dependiente estará regida por **más de una variable independiente**.

$$z = f(x, y)$$

- **Dominio:** (x, y)
- **Rango:** $f(x, y)$

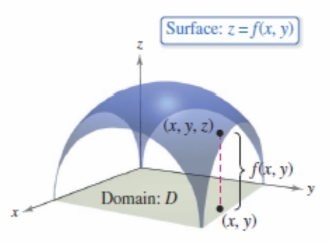


Figura: [1]

Funciones de Varias Variables

- **Función de dos variables**:

$$f(x, y) = x - y$$

- **Función de tres variables**:

$$h(x, y, z) = x + y + xz$$

Ejemplos

Graficar en Python

1. $z = x^2 + y^2$
2. $z = x + y$
3. $z = x^2 + y$
4. $z = \sqrt{x^2 + y^2}$
5. $z = \sqrt{1 - x^2 + y^2}$

Agenda

Derivadas Parciales

Definición

Las derivadas parciales de una función de dos o mas variables indica como el *cambio en una de las variables independientes* afecta a la función. Si $z = f(x, y)$, las derivadas parciales f_x y f_y se denotan por:

$$\frac{\partial}{\partial x} f(x, y) = f_x(x, y) = z_x = \frac{\partial z}{\partial x} \text{ y } \frac{\partial}{\partial y} f(x, y) = f_y(x, y) = z_y = \frac{\partial z}{\partial y}$$

Ejemplo

Hallar las derivadas parciales f_x y f_y de las funciones:

1. $f(x, y) = x^2y^2 + xy + y$
2. $f(x, y) = 3x - x^2y^2 + 2x^3y$

Derivadas Parciales

Ejemplo

Hallar las pendientes de la función de dos variables en las direcciones de x y de y en el punto $(\frac{1}{2}, 1, 2)$.

$$f(x,y) = -\frac{x^2}{2} - (y-2)^2 + \frac{25}{8}$$

Graficar con matplotlib esta función y las rectas tangentes en el mismo punto.

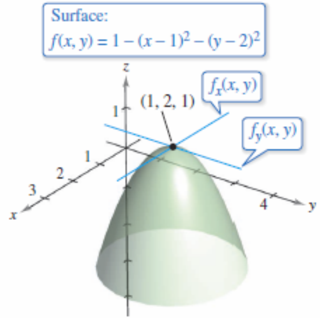
Derivadas Parciales

Ejemplo

Hallar las pendientes de la función de dos variables en las direcciones de x y de y en el punto $(1, 2, 1)$.

$$f(x, y) = 1 - (x - 1)^2 - (y - 2)^2$$

Graficar con matplotlib esta función y las rectas tangentes en el mismo punto.



De igual forma que con las funciones de una variable, es posible hallar las segundas, terceras, etc, derivadas parciales.

Orden de las Derivadas Parciales de una función de dos variables. Sea $z = f(x, y)$

1. Segunda derivada con respecto a x:

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} = f_{xx}$$

2. Segunda derivada con respecto a y:

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2} = f_{yy}$$

3. Derivar primero con respecto a x y luego con respecto a y:

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x} = f_{xy}$$

4. Derivar primero con respecto a y y luego con respecto a x:

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x \partial y} = f_{yx}$$

Ejemplo

Calcular las derivadas parciales de segundo orden de: $f(x, y) = 3xy^2 - 2y + 5x^2y^2$

Derivadas parciales de primer orden:

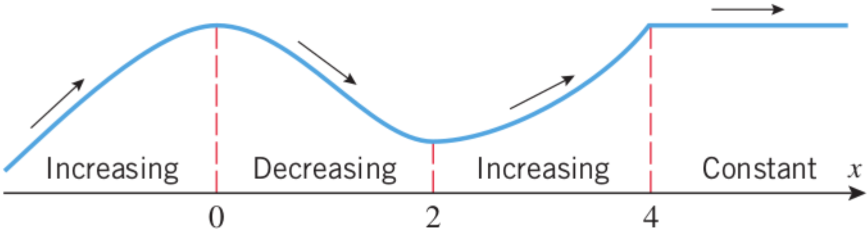
$$f_x(x, y) = 3y^2 + 10xy^2 \qquad f_y(x, y) = 6xy - 2 + 10x^2y$$

Derivadas parciales de segundo orden:

$$f_{xx}(x, y) = 10y^2 \qquad f_{yy}(x, y) = 6x + 10x^2$$

$$f_{xy}(x, y) = 6y + 20xy \qquad f_{yx}(x, y) = 6y + 20xy$$

Funciones Crecientes y Decrecientes

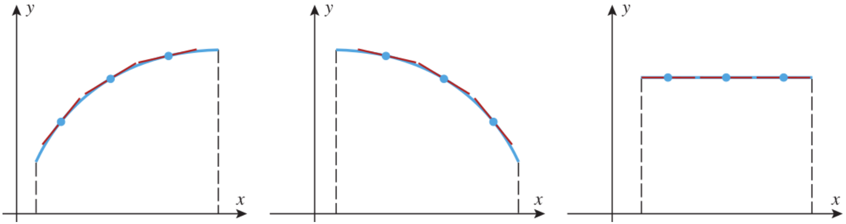


Definición

Sea f una funcion definida en un intervalo, y x_1 y x_2 puntos en el intervalo.

- f es **creciente** en el intervalo si $f(x_1) < f(x_2)$ siempre que $x_1 < x_2$.
- f es **decreciente** en el intervalo si $f(x_1) > f(x_2)$ siempre que $x_1 < x_2$.
- f es **constante** en el intervalo si $f(x_1) = f(x_2)$ para todo punto x_1 and x_2.

Funciones Crecientes y Decrecientes



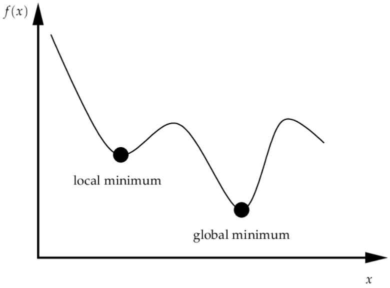
Sea f una función que es **continua** en un intervalo cerrado $[a, b]$ y **diferenciable** en el intervalo abierto (a, b).

- Si $f'(x) > 0$ para todo valor de x en (a, b), entonces f es creciente en $[a, b]$
- Si $f'(x) < 0$ para todo valor de x en (a, b), entonces f es decreciente en $[a, b]$
- Si $f'(x) = 0$ para todo valor de x en (a, b), entonces f es constante en $[a, b]$

Agenda

Optimización

- La solución de problemas de **optimización** es una de las principales aplicaciones de la diferenciación (derivadas).
- Un problema de optimización hace referencia a hallar el máximo o **mínimo** valor de una función.



- **Mínimo global:** $x^* \in S$, Si $f(x^*) \leq f(x)$ para todo $x \in S$
- **Mínimo local:** $x^* \in S$, Si $f(x^*) \leq f(x)$ for all x dentro de una distancia infinitesimalmente pequeña ϵ de x^*. Es decir, existe $\epsilon > 0$ tal que para todo x satisface $|x - x^*| < \epsilon$, $f(x^*) \leq f(x)$

La teoría de **optimización** provee las herramientas para encontrar un "buen"valor de x que corresponde a un *buen* valor de $f(x)$

Asumiendo que la primera y segunda derivada de $f(x)$ existen, las **condiciónes necesaria** para que x^* sea un **mínimo/máximo local** de la función $f(x)$ en un intervalo (a, b) son:

Mínimo Local

1. $\left.\frac{df}{dx}\right|_{x=x^*} = 0$
2. $\left.\frac{d^2f}{dx^2}\right|_{x=x^*} \geq 0$

Máximo Local

1. $\left.\frac{df}{dx}\right|_{x=x^*} = 0$
2. $\left.\frac{d^2f}{dx^2}\right|_{x=x^*} \leq 0$

Importante

Estas condiciones son necesarias, pero no suficientes. Que se cumplan no garantiza que x^* sea un mínimo o máximo local.

Sea x^* un punto en el cual la primera derivada es igual a cero y el orden de la primera derivada superior distinta de cero es n.

1. Si n es impar, entonces x^* es un **punto de inflexión**.
2. Si n es par, entonces x^* es un optimo local. Ademas:
   1. Si el valor de la derivada de $f''(x)$ en x^* es positiva, entonces el punto x^* es un **mínimo local**.
   2. Si el valor de la derivada de $f''(x)$ en x^* es negativa, entonces el punto x^* es un **máximo local**.

Ejemplo

Calcular los extremos relativos de la función $f(x) = -3x^5 + 5x^3$. Graficar la función en Python e identificar estos puntos.

Ejemplo

Calcular el máximo valor de la función $f(x) = -x^3 + 3x^2 + 9x + 10$ en el intervalo $-2 \leq x \leq 4$. Graficar Python e identificar este punto.

Solución:

$$\frac{df}{dx} = -3x^2 + 6x + 9$$

$$-3x^2 + 6x + 9 = 0$$

$$x = 3 \text{ y } x = -1$$

Evaluando $f(x)$ en $x = 3$ el resultado es 37, Entonces $x = 3$ corresponde al máximo de la función

Agenda

Gradiente

Definición

Dada una función $f(\cdot)$, su gradiente representa el vector cuya dirección y la magnitud del cambio en el valor de la función son **máximos**. Matemáticamente, corresponden a sus **primeras derivadas parciales** tal que:

$$\nabla(f(x)) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix}$$

donde x es un vector n-dimensional.

Esto implica que usar gradientes es una forma de **realizar optimización a través de derivadas** con funciones de 2 o mas variables.

Gradiente

Ejemplo

Dada la función $f(x) = 2x_1^4 + 4x_2^3 - 3x_1x_2^2$, Calcular el gradiente.

Solución:

$$\nabla(f(x)) = \begin{bmatrix} 8x_1^3 - 3x_2^2 \\ \\ 12x_2^2 - 6x_1x_2 \end{bmatrix}$$

Teniendo en cuenta en este ejemplo que $f(x)$ presenta dos variables (x_1 y x_2), el vector fila resultante será de **2x1**.

Matriz Hessiana

Definición

La Hessiana de una función $f(x)$ es la *matriz simétrica* que proporciona información sobre la **curvatura de la función**, conformada por las **segundas derivadas parciales**, la cual esta dada por:

$$\nabla^2(f(x)) = H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Brinda informacion de la convexidad de la función a minimizar.

Matriz Hessiana

Ejemplo

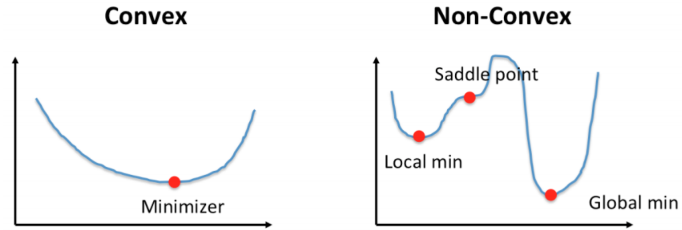
Dada la función $f(x) = 2x_1^4 + 4x_2^3 - 3x_1x_2^2$ calcular la Hessiana

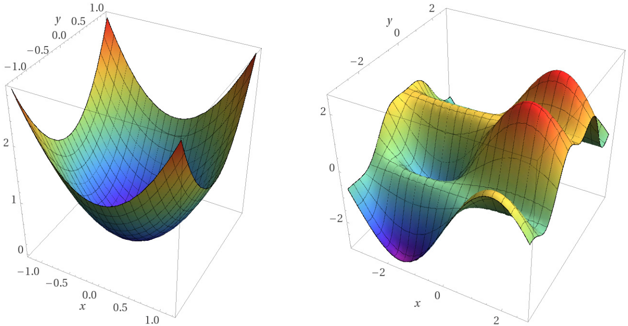
Solución:

$$\nabla^2(f(x)) = H(f) = \begin{bmatrix} 24x_1^2 & -6x_2 \\ \\ -6x_2 & 24x_2 - 6x_1 \end{bmatrix}$$

Teniendo en cuenta en este ejemplo que $f(x)$ también presenta dos variables (x_1 y x_2), la matriz resultante será de **2x2**.

Funciones Convexas y No Convexas



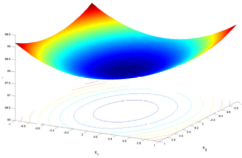


Matriz Hessiana

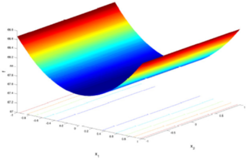
Convexidad (1/2)

De acuerdo al comportamiento de la hessiana de una función, se puede obtener:

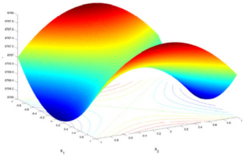
- Definida positiva: Todos los menores principales son positivos (mínimo local).
- Semi-definida positiva: Menores principales no negativos. (posible mínimo local).
- Definida negativa: Todos los menores principales son negativos (máximo local).
- Semi-definida negativa: Menores principales alternan entre cero o negativos. (posible máximo local)
- Indefinida: Menores principales varían en signo; (punto de silla)



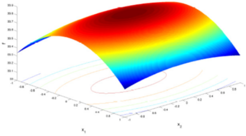
(a) Positive definite



(b) Positive semi-definite



(c) Indefinite



(d) Negative definite

Matriz Hessiana

Convexidad (2/2)

Definida Positiva

$$H = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

- $\det(H_1) = 2 > 0$
- $\det(H_2) = 3 > 0$
- $\det(H_3) = 2 \cdot 3 \cdot 4 = 24 > 0$

Semidefinida Positiva

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

- $\det(H_1) = 1 > 0$
- $\det(H_2) = 1 \times 1 = 1 > 0$
- $\det(H_3) = 0 \geq 0$

Definida Negativa

$$H = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -4 \end{pmatrix}$$

- $\det(H_1) = -2 < 0$
- $\det(H_2) = (-2) \cdot (-3) = 6 > 0$
- $\det(H_3) = (-2) \cdot (-3) \cdot (-4) = -24 < 0$

Semidefinida Negativa

$$H = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

- $\det(H_1) = -1 < 0$
- $\det(H_2) = (-1) \times (-1) = 1 > 0$
- $\det(H_3) = 0 \leq 0$

Indefinida

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Menores Principales:

- $\det(H_1) = 1 > 0$
- $\det(H_2) = 1 \times (-1) = -1 < 0$
- $\det(H_3) = 1 \times (-1) \times 1 = -1 < 0$

Agenda

Algoritmos de Optimización

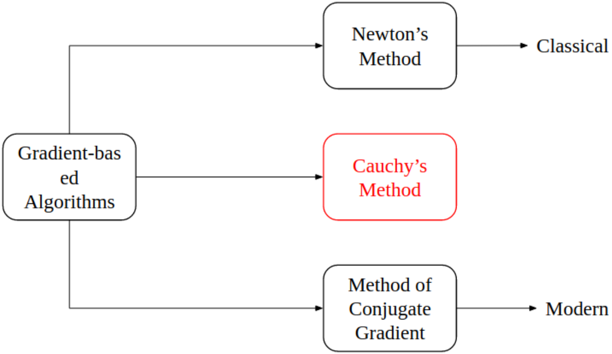
Definición

Métodos y técnicas utilizados para encontrar los *valores de las variables que maximizan o minimizan* una función objetivo. Estos algoritmos se pueden clasificar de muchas maneras, dependiendo del enfoque o las caracteristicas que estamos tratando de comparar[2]. Por ejemplo, podemos clasificarlos de la siguiente manera:

- Métodos basados en gradiente
- Métodos libres de gradiente
- Algoritmos metaheurísticos

Algoritmos de Gradiente

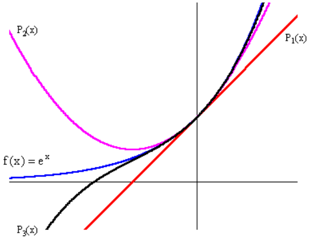
Algoritmos que utilizan la información de la **derivada** para su desarrollo. Calculan las derivadas de la función objetivo, ya sea por aproximación (diferencias finitas) o por procesos de diferenciación automática.



Series de Taylor

$$f(x) = f(a) + \underbrace{\frac{\dot{f}(a)}{1!}(x-a)}_{(1)} + \underbrace{\frac{\ddot{f}(a)}{2!}(x-a)^2}_{(2)} + \frac{\dddot{f}(a)}{3!}(x-a)^3 + \cdots$$

(1) Gradiente: Permite conocer la dirección de búsqueda de los parámetros de la función.
(2) Hessiana: Permite determinar la concavidad.



$$x(k+1) = x(k) + \alpha(k)s(x(k))$$

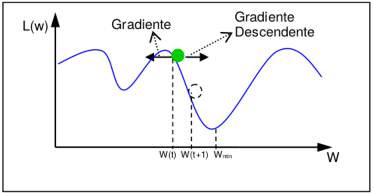
donde:

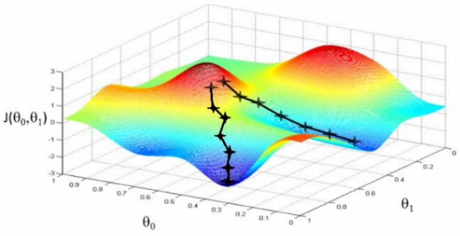
- $x(k) =$ estimación actual de x^*, la solución.
- $\alpha(k) =$ parámetro de longitud de paso
- $s(x(k)) =$ dirección de búsqueda en el espacio N de las variables de diseño x_i

La forma en que $s(x)$ y α se determinan en cada iteración, define un método particular.

Algoritmos de Gradiente

Método de Cauchy



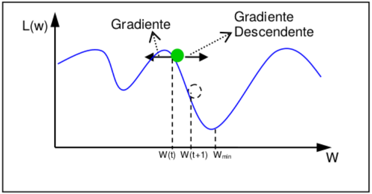


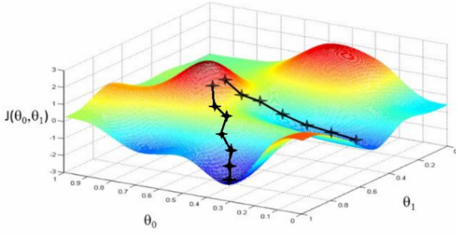
$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^T \Delta x + \cdots$$

$$x(k+1) = x(k) + \alpha(k) s(x(k)) \longrightarrow s(\bar{x}) = -\nabla f(\bar{x})^T$$

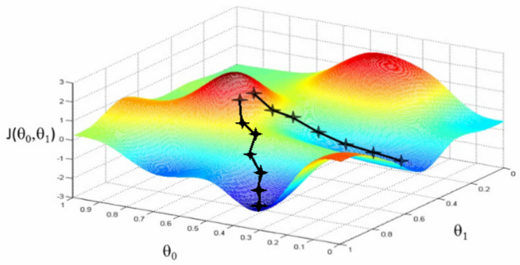
Algoritmos de Gradiente

Método de Cauchy





$$x(k+1) = x(k) - \alpha \nabla f(x(k)) \longrightarrow \textbf{Simple Gradient Method}$$



Método de Gradiente Simple

$$x(k+1) = x(k) - \alpha \nabla f(x(k))$$

- La elección de α debe ser adecuada.
- lentitud inherente cerca del mínimo debido a que las correcciones desaparecen cuando ∇f llega a cero.

Metodo de Cauchy

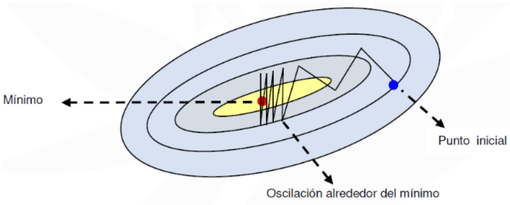
$$x(k+1) = x(k) - \alpha(k) \nabla f(x(k))$$

$\alpha(k)$ se determina de manera que $f(x(k+1))$ es un mínimo a lo largo de $\nabla f(x(k))$ utilizando un **método de búsqueda de una sola variable** apropiado.

Como el valor de α afecta al método?

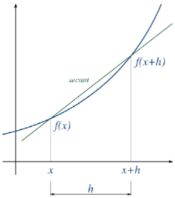
$$X(t+1) = X(t) + \alpha\left(-\frac{df(X)}{dX}\right)$$

$$\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \\ \cdots \\ x_N(t+1) \end{bmatrix} = \begin{bmatrix} x_1(t) + \alpha\left(-\frac{\partial f(X)}{\partial x_1}\right) \\ x_2(t) + \alpha\left(-\frac{\partial f(X)}{\partial x_2}\right) \\ \cdots \\ x_N(t) + \alpha\left(-\frac{\partial f(X)}{\partial x_N}\right) \end{bmatrix}$$



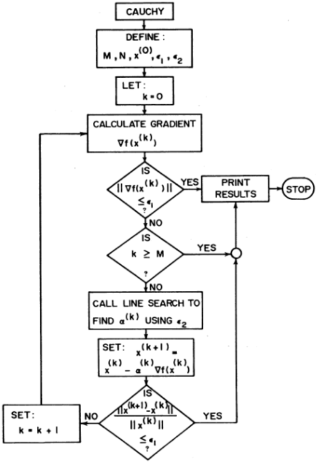
Gradiente Estimado

$$\underbrace{X(t+1) = X(t) + \alpha\left(-\frac{df(X)}{dX}\right)}_{\text{Analytic Gradient}} \longrightarrow \underbrace{\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} = \begin{bmatrix} x_1(t) + \alpha\left(-\dfrac{f(x_1+h, x_2) - f(x_1, x_2)}{h}\right) \\ x_2(t) + \alpha\left(-\dfrac{f(x_1, x_2+h) - f(x_1, x_2)}{h}\right) \end{bmatrix}}_{\text{Estimated gradient}}$$



https://en.wikipedia.org/wiki/Numerical_differentiation

Diagrama de Flujo



Conclusiones

- Este método es más confiable que el método de gradiente simple, sin embargo, la velocidad de convergencia será extremadamente lenta para la mayoría de los problemas prácticos.
- Este método normalmente produce buenos resultados desde puntos alejados del mínimo. Esto debido a:

$$f(x(k+1) \leq f(x(k))$$

- El criterio de parada se da cuando el gradiente se vuelve cero o se cumplen el numero de iteraciones.
- La convergencia del método depende de la condición inicial y del valor de α.

Ejemplo

Minimizar la función:

$$f(x) = x_1^2 - 24x_1 + x_2^2 - 10x_2$$

Sea $x^0 = (8, 7)$ un punto de partida cualquiera.

Solución:

Pasos 1 y 2: Calcular el gradiente (vector de ascenso/descenso) y evaluarlo en el punto de partida

$$f_{x_1} = \frac{\partial f}{\partial x_1} = 2x_1 - 24 \qquad\qquad f_{x_2} = \frac{\partial f}{\partial x_2} = 2x_2 - 10$$

$$\nabla f(x) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \\ \dfrac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 - 24 \\ \\ 2x_2 - 10 \end{bmatrix}$$

$$-\nabla f(x^0) = S^0 = -\begin{bmatrix} 2(8) - 24 \\ \\ 2(7) - 10 \end{bmatrix} = -\begin{bmatrix} -8 \\ \\ 4 \end{bmatrix} \text{ (estamos minimizando)}$$

Solución:

Paso 3: Plantear la función de alfa - g(α)

$$g(\alpha) = f(x^0 + \alpha S^0)$$

$$g(\alpha) = f((8,7) + \alpha(8,-4)) = f((8+8\alpha, 7-4\alpha))$$

$$g(\alpha) = (8+8\alpha)^2 - 24(8+8\alpha) + (7-4\alpha)^2 - 10(7-4\alpha)$$

$$g(\alpha) = (64 + 128\alpha + 64\alpha^2) - (192 + 192\alpha) + (49 - 56\alpha + 16\alpha) - (70 - 40\alpha)$$

$$g(\alpha) = 80\alpha^2 - 80\alpha - 149$$

Solución:

Paso 4: Encontrar el valor de alfa (α)

$$\frac{\partial g}{\partial \alpha} = 160\alpha - 80$$

$$\frac{\partial g}{\partial \alpha} = 0 \rightarrow 160\alpha - 80 = 0$$

$$\alpha = \frac{80}{160} = \frac{1}{2} \geq 0 \rightarrow \text{Valor minimo de } g(x)$$

Solución:

Paso 5: Actualizar el valor de los puntos iniciales ($x^0 \to x^1$)

$$x(k+1) = x^1 = x^0 + \alpha S^0 \longrightarrow x^1 = (8,7) + \frac{1}{2}(8,-4)$$

$$x^1 = (8+4, 7-2) = (12,5)$$

Repetimos los pasos usando estos nuevos puntos hasta que el valor de S^n sea cero (esto indica la aparición de un maximo/minimo segun sea el caso, donde n será el numero de iteraciones donde esto ocurra).

Algoritmos de Gradiente

Consideraciones sobre el metodo de Cauchy

- Se aplica en funciones de optimización sin restricciones
- Presenta problemas en situaciones de alta dimensionalidad (muchas características)
- Uso bajo un contexto muy reducido (funciones cuadráticas), por lo que es poco versátil)
- Es sensible a los valores de iniciación (baja velocidad de convergencia).
- Orientado a investigación académica.
- Alternativa para I.A: Metodos basados en gradiente descendente.

Referencias

Ron Larson, Robert P Hostetler, and Bruce H Edwards.
Calculus.
Houghton Mifflin Boston, 2006.

Gintaras V Reklaitis, A Ravindran, and Kenneth M Ragsdell.
Engineering optimization: Methods and applications.
Wiley New York, 1983.

https://medium.com/ai-society/hello-gradient-descent-ef74434bdfa5