Cálculo, optimización y cálculo multivariable

Universidad Internacional de Valencia

Máster Universitario en Inteligencia Artificial

02MIAR | Matemáticas:

Matemáticas para la Inteligencia Artificial

Profesor:

David Zorío Ventura

universidadviu.com

De: Planeta Formación y Universidades

Funciones monótonas

Definición

Sea $f : A \to \mathbb{R}$*. Diremos que* f *es (estrictamente) creciente en* $B \subseteq A$ *si* $\forall a, b \in B$, $a < b$, $f(a) \leq f(b)$ *(resp.,* $f(a) < f(b)$*).*

Definición

Sea $f : A \to \mathbb{R}$*. Diremos que* f *es (estrictamente) decreciente en* $B \subseteq A$ *si* $\forall a, b \in B$, $a < b$, $f(a) \geq f(b)$ *(resp.,* $f(a) > f(b)$*).*

Teorema

Sea $f : A \to \mathbb{R}$ *y* $B \subseteq A$ *intervalo abierto, con* f *derivable en* B*. Entonces:*

- *Si* $f'(x) \geq 0$ *(resp.,* $f'(x) > 0$*)* $\forall x \in B$*, entonces* f *es creciente (resp., estrictamente creciente) en* B*.*
- *Si* $f'(x) \leq 0$ *(resp.,* $f'(x) < 0$*)* $\forall x \in B$*, entonces* f *es decreciente (resp., estrictamente decreciente) en* B*.*

Definición

Sea $f : A \to \mathbb{R}$ *y* $B \subseteq A$. *Diremos que un punto* $c \in B$ *es un* **máximo absoluto de** f **en** B *si* $f(c) \geq f(x)$ $\forall x \in B$.

Definición

Sea $f : A \to \mathbb{R}$ *y* $B \subseteq A$. *Diremos que un punto* $c \in B$ *es un* **mínimo absoluto de** f **en** B *si* $f(c) \leq f(x)$ $\forall x \in B$.

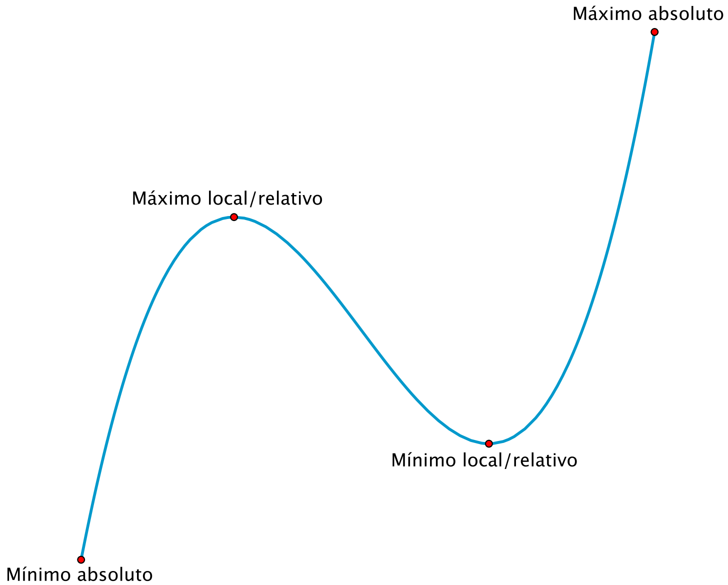
Definición

Sea $f : A \to \mathbb{R}$. *Diremos que un punto* $c \in A$ *es un* **máximo relativo o local de** f *si* $\exists \rho > 0$ *tal que* $f(c) \geq f(x)$ $\forall x \in (c - \rho, c + \rho)$.

Definición

Sea $f : A \to \mathbb{R}$. *Diremos que un punto* $c \in A$ *es un* **mínimo relativo o local de** f *si* $\exists \rho > 0$ *tal que* $f(c) \leq f(x)$ $\forall x \in (c - \rho, c + \rho)$.

Extremos


Máximo absoluto
Máximo local/relativo
Mínimo local/relativo
Mínimo absoluto

Puntos críticos

Definición

Sean $f : A \to \mathbb{R}$, $c \in A$. c *es un* **punto crítico** *de* f *si* $\exists f'(c) = 0$.

Ejemplo

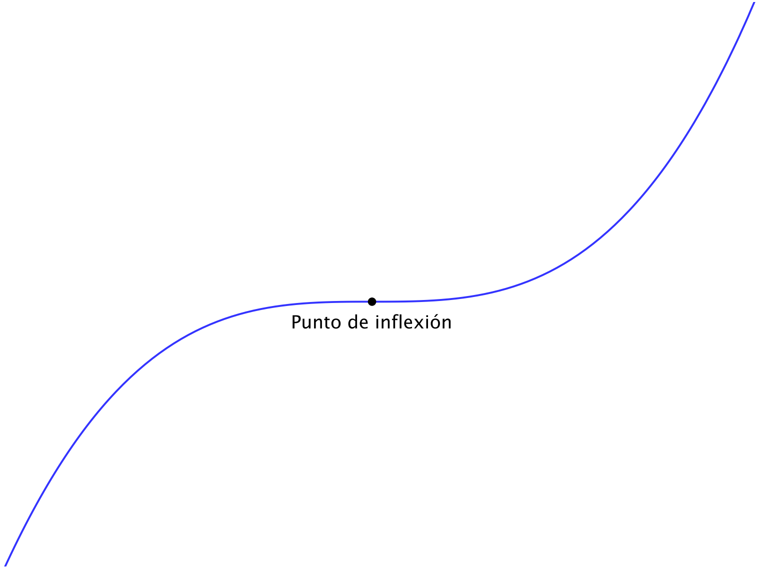
Sea $f(x) = 2x^3 - 3x^2 \to f'(x) = 6x^2 - 6x$. *Imponiendo* $f'(x) = 0$, *se obtiene* $x = 0$ *y* $x = 1$, *por lo que* f *tiene dos puntos críticos:* 0 *y* 1.

Teorema

Sea c *un punto crítico de* f *y* $n \geq 2$ *el primer natural tal que* $f^{(n)}(c) \neq 0$. *Entonces:*

- *Si* n *es par, se tienen dos casos:*
  - *Si* $f^{(n)}(c) < 0$, *entonces* c *es un máximo relativo de* f.
  - *Si* $f^{(n)}(c) > 0$, *entonces* c *es un mínimo relativo de* f.
- *Si* n *es impar, entonces* c *es un* **punto de inflexión** *de* f.

Puntos críticos



Puntos críticos

Ejemplos

1. *Vimos anteriormente que los puntos críticos de* $f(x) = 2x^3 - 3x^2$ *son* $x = 0$ *y* $x = 1$. *Tenemos* $f'(x) = 6x^2 - 6x$ *y* $f''(x) = 12x - 6$.
   - *Por una parte, para* $x = 0$ *se tiene* $f''(0) = -6 < 0$, *por lo que* 0 *es un máximo relativo de* f.
   - *Por otra parte, para* $x = 1$, *tenemos* $f''(1) = 6 > 0$, *por lo que* 1 *es un mínimo relativo de* f.
2. *Sea* $f(x) = x^3$. *En este caso,* $f'(x) = 3x^2$, *por lo que su único punto crítico es* $x = 0$. *Si calculamos* f'', *obtenemos* $f''(x) = 6x$, *con lo cual* $f''(0) = 0$. *Por otra parte,* $f'''(x) = 6$, *luego* $f'''(0) = 6$. *Por tanto,* 0 *es un punto de inflexión y no es extremo local, puesto que* $f(x) < f(0)$ *para* $x < 0$ *y* $f(x) > f(0)$ *para* $x > 0$.

Ejemplo

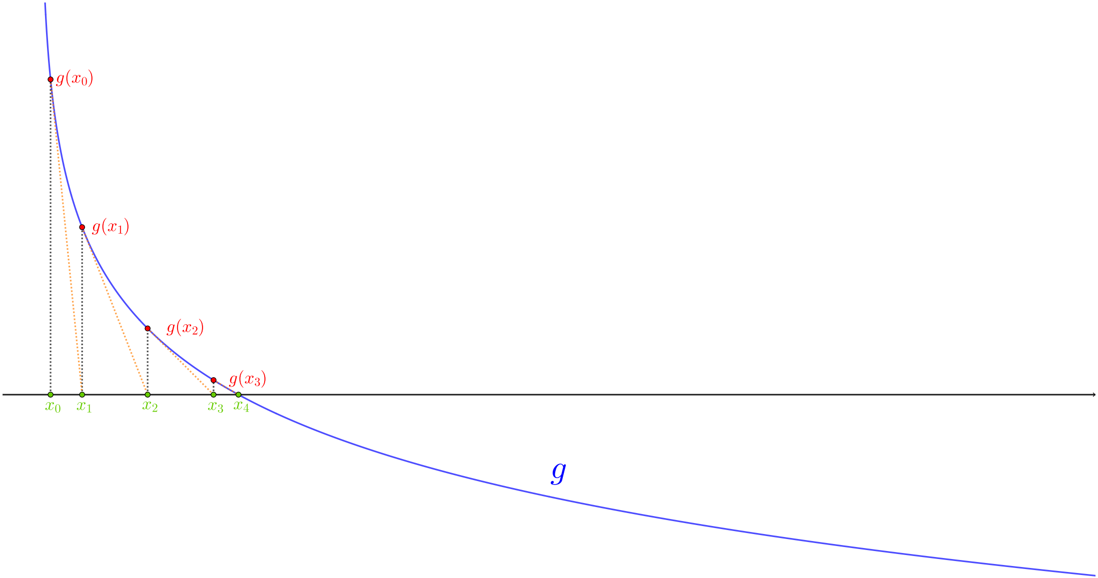
Un comerciante vende un determinado producto por un valor de 50 *euros, con una media de* 200 *clientes mensuales. Por otra parte, sabe que por cada euro que aumente el precio, perderá un promedio de* 2 *clientes mensuales. ¿Cuánto tiene que aumentar el precio para aspirar a obtener el mayor beneficio posible?*

- *Si* x *es el aumento de precio, la función de ganancia mensual es* $f(x) = (50 + x)(200 - 2x) = -2x^2 + 100x + 10000$.
- *Analizamos* f' *con el fin de buscar puntos críticos:* $f'(x) = -4x + 100$, *luego el único punto crítico correspondiente a la identidad* $f'(x) = 0$ *es* $x = 25$.
- *Calculamos* $f''(x) = -4$ *y obtenemos* $f''(25) = -4 < 0$, *luego hay un máximo relativo en* $x = 25$.
- *Por tanto, deberá aumentar el precio* 25 *euros.*

Ejercicio

Un algoritmo satisface la siguiente propiedad. Si éste se ejecuta en paralelo a través de dos procesadores, cada procesador pierde un 0.8 % *de eficiencia con respecto a su funcionamiento original sin ir en paralelo. Si es ejecutado en paralelo a través de tres procesadores, cada procesador pierde un* 1.6 % *de eficiencia con respecto al funcionamiento original. Si hay cuatro, un* 2.4 % *con respecto al funcionamiento original, y así sucesivamente (tras cada procesador añadido en paralelo, cada uno de ellos pierde un* 0.8 % *de eficiencia con respecto a su funcionamiento original sin ir en paralelo con otros). ¿Cuál es el número óptimo de procesadores en paralelo para ejecutar el algoritmo lo más rápidamente posible?*

Método de Newton



Descenso de gradiente

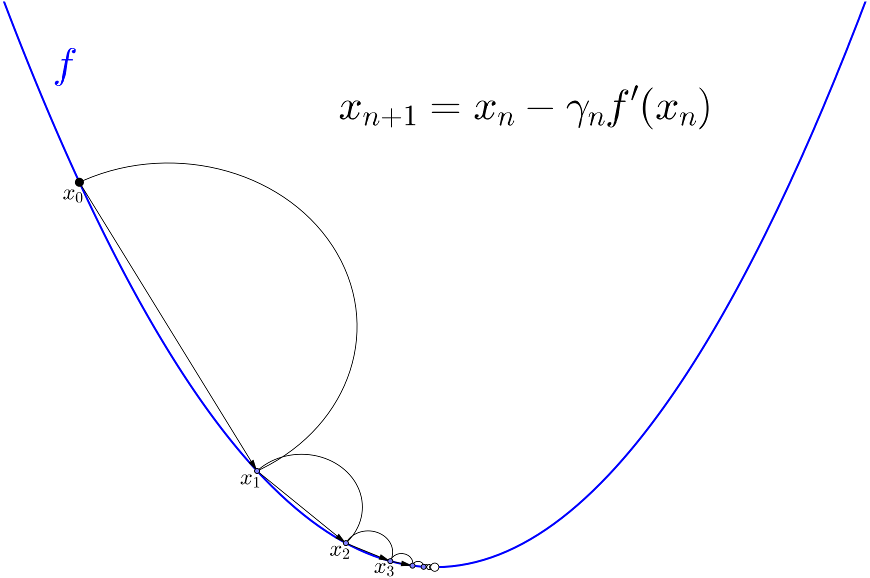
- No todas las funciones que se desean optimizar tienen una expresión sencilla, con una derivada que se pueda conocer globalmente, obteniéndose directamente los puntos críticos.
- **Ejemplo:** funciones de coste en el entrenamiento de una red neuronal.
- Esto precisa de métodos que permitan aproximarse a mínimos locales vía un conjunto discreto de evaluaciones.
- **Descenso de gradiente** (caso de una variable):
  - Dada $f : A \to \mathbb{R}$, elegimos un valor inicial $x_0 \in A$.
  - Para $n \geq 0$, obtenemos x_{n+1} como sigue:

$$x_{n+1} = x_n - \gamma_n f'(x_n),$$

donde $\gamma_n > 0$ recibe el nombre de **ratio de aprendizaje**.

Descenso de gradiente

$$x_{n+1} = x_n - \gamma_n f'(x_n)$$



Posibles elecciones del ratio de aprendizaje:

- $\gamma_n = \gamma > 0$, valor constante, lo suficientemente pequeño como para garantizar la convergencia.
- **Método de Newton modificado** sobre f':
  - Aproximación local de f' vía recta tangente:
  $$f'(x) \approx f'(x_n) + (x - x_n)f''(x_n).$$
  - Aproximación de un x tal que $f'(x) = 0$:
  $$f'(x_n) + (x - x_n)f''(x_n) = 0 \leftrightarrow x = x_n - \frac{f'(x_n)}{f''(x_n)}.$$
  - Construcción del método iterativo:
  $$x_{n+1} = x_n - \frac{f'(x_n)}{|f''(x_n)|} = x_n - \frac{1}{|f''(x_n)|}f'(x_n) = x_n - \gamma_n f'(x_n),$$
  donde $\gamma_n = \dfrac{1}{|f''(x_n)|}$.

Descenso de gradiente

- **Método de la secante modificado** sobre f' (evita tener que trabajar sobre f''):
  - Aproximación de $f''(x_n)$ en términos de f':

$$f''(x_n) \approx \frac{f'(x_n) - f'(x_{n-1})}{x_n - x_{n-1}}.$$

  - Redefinición del método:

$$x_{n+1} = x_n - \gamma_n f'(x_n), \quad n \geq 1,$$

  donde $\gamma_0 = h$, con $h > 0$ pequeño, y

$$\gamma_n = \frac{1}{\left|\frac{f'(x_n)-f'(x_{n-1})}{x_n-x_{n-1}}\right|} = \frac{|x_n - x_{n-1}|}{|f'(x_n) - f'(x_{n-1})|}, \quad n \geq 1.$$

- En resumen, posibles elecciones de γ_n (entre muchas otras):

$$\gamma_n = h, \quad \gamma_n = \frac{1}{|f''(x_n)|}, \quad \gamma_n = \begin{cases} h & n = 0 \\ \dfrac{|x_n - x_{n-1}|}{|f'(x_n) - f'(x_{n-1})|} & n > 0 \end{cases}$$

Funciones de varias variables

Universidad Internacional de Valencia

Definición

Una función real de **varias variables** *es una función de la forma* $f : A \to \mathbb{R}$, *con* $A \subseteq \mathbb{R}^n$, $n > 1$.

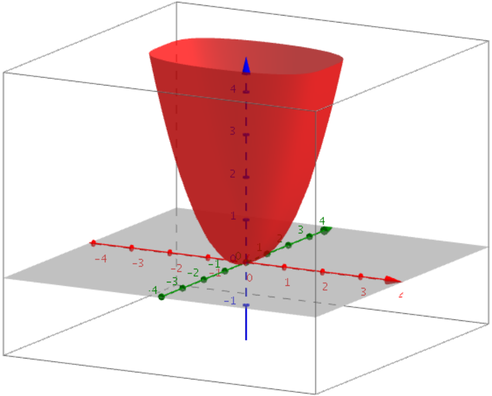
Ejemplos

1. $f : \mathbb{R}^2 \to \mathbb{R}$ *dada por* $f(x, y) = \sqrt{x^2 + y^2}$ *es una función de dos variables (que en este caso representa el módulo de un vector con coordenadas* (x, y)*).*
2. *Sea* $A = \{(x, y) \in \mathbb{R}^2 \mid x > y\} \subseteq \mathbb{R}^2$. *Entonces* $f : A \to \mathbb{R}$ *dada por* $f(x, y) = \ln(x - y)$ *es una función de dos variables.*
3. $f : \mathbb{R}^3 \to \mathbb{R}$ *dada por* $f(x, y, z) = x^2 + yz - \sin(\pi(x + z))$ *es un ejemplo de una función de tres variables.*

Funciones de varias variables

Al igual que en el caso de las funciones reales de variable real, las funciones de varias variables de la forma $f : A \to \mathbb{R}$, con $A \subseteq \mathbb{R}^n$ pueden representarse gráficamente, en forma del siguiente conjunto:

$$\{(x; f(x)) \mid x \in A\} \subseteq \mathbb{R}^{n+1}.$$



Derivadas parciales

Definición

Sea $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ y $(x_1, \dots, x_n) \in A$. Diremos que f admite derivada parcial en la dirección del eje i si

$$\exists \lim_{h \to 0} \frac{f(x_1, \dots, x_i + h, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{h}.$$

En tal caso, llamamos **derivada parcial** *i-ésima de f en $(x_1, \dots, x_n)$ al valor anterior, y la denotamos por*

$$\frac{\partial f}{\partial x_i}(x_1, \dots, x_n) = \lim_{h \to 0} \frac{f(x_1, \dots, x_i + h, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{h}.$$

Además, si todas las derivadas parciales existen, llamamos **gradiente** *de f en $(x_1, \dots, x_n)$ a la expresión*

$$\nabla f(x_1, \dots, x_n) = \left(\frac{\partial f}{\partial x_1}(x_1, \dots, x_n), \dots, \frac{\partial f}{\partial x_n}(x_1, \dots, x_n) \right).$$

Derivadas parciales

Ejemplos

1. *Sea* $f(x,y) = 2x + 3y$. *Entonces* $\frac{\partial f}{\partial x}(x,y) = 2$ *y* $\frac{\partial f}{\partial y}(x,y) = 3$, *luego* $\nabla f(x,y) = (2,3)$.
2. *Consideremos* $f(x,y) = x^2 + y^2$. *Entonces* $\frac{\partial f}{\partial x}(x,y) = 2x$ *y* $\frac{\partial f}{\partial y}(x,y) = 2y$, *luego* $\nabla f(x,y) = (2x,2y)$.
3. *Tomemos* $f(x,y,z) = xy + z$. *Entonces* $\frac{\partial f}{\partial x}(x,y,z) = y$, $\frac{\partial f}{\partial y}(x,y,z) = x$ *y* $\frac{\partial f}{\partial z}(x,y,z) = 1$, *luego* $\nabla f(x,y,z) = (y,x,1)$.
4. *Sea* $f(x,y,z) = x\ln(x^2+y^2+z^2) + \sin(\pi yz)$. *Entonces* $\frac{\partial f}{\partial x}(x,y,z) = \ln(x^2+y^2+z^2) + \frac{2x^2}{x^2+y^2+z^2}$, $\frac{\partial f}{\partial y}(x,y,z) = \frac{2xy}{x^2+y^2+z^2} + \pi z\cos(\pi yz)$, $\frac{\partial f}{\partial z}(x,y,z) = \frac{2xz}{x^2+y^2+z^2} + \pi y\cos(\pi yz)$.

Derivadas direccionales

Universidad Internacional de Valencia

Definición

Sea $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$. Diremos que f admite derivada en la dirección de $v \in \mathbb{R}^n$ en $x \in \mathbb{R}^n$ si

$$\exists \lim_{h \to 0} \frac{f(x+hv) - f(x)}{h}.$$

En tal caso, denotaremos por $D_v f(x)$ al límite anterior, que llamaremos **derivada direccional** *de f en x en la dirección de v.*

Obsérvese que en el caso particular de los vectores de la base canónica de $\mathbb{R}^n$, $e_i = (0, \dots, 0, \underbrace{1}_{i}, 0, \dots, 0)$, se cumple $D_{e_i} f(x) = \frac{\partial f}{\partial x_i}(x)$.

Teorema

En las condiciones anteriores, se cumple $D_v f(x) = \nabla f(x) \cdot v$.

Derivadas direccionales

- Combinando lo anterior con propiedades del producto escalar:

$$D_v f(x) = \nabla f(x) \cdot v = \|\nabla f(x)\|_2 \|v\|_2 \cos(\theta),$$

donde $\theta \in [0, 2\pi[$ es el ángulo que forman $\nabla f(x)$ y v.

- En particular:

$$\underbrace{-\|\nabla f(x)\|_2 \|v\|_2}_{\theta=\pi} \leq D_v f(x) \leq \underbrace{\|\nabla f(x)\|_2 \|v\|_2}_{\theta=0}$$

- Si $\theta = 0$, entonces $v \propto \nabla f(x)$ y

$$D_v f(x) = \|\nabla f(x)\|_2 \|v\|_2 \cos(0) = \|\nabla f(x)\|_2 \|v\|_2.$$

$\nabla f(x)$ **es la dirección de mayor crecimiento de** f **en** x.

- Si $\theta = \pi$, entonces $v \propto -\nabla f(x)$ y

$$D_v f(x) = \|\nabla f(x)\|_2 \|v\|_2 \cos(\pi) = -\|\nabla f(x)\|_2 \|v\|_2.$$

$-\nabla f(x)$ **es la dirección de mayor decrecimiento de** f **en** x.

Regla de la cadena

Teorema (Regla de la cadena)

Sean $g : \mathbb{R}^n \to \mathbb{R}^m$, $f : \mathbb{R}^m \to \mathbb{R}$ *y* $a \in \mathbb{R}^n$ *tales que* $\exists \nabla g_i(a)\ \forall i \in \{1, \ldots, m\} \wedge \exists \nabla f(g(a))$*. Entonces la composición* $f \circ g : \mathbb{R}^n \to \mathbb{R}$*, dada por* $(f \circ g)(a) = f(g(a)) = f(g_1(a), \ldots, g_m(a))$*, cumple*

$$\exists \frac{\partial (f \circ g)}{\partial x_j}(a) = \nabla f(g(a)) \cdot \frac{\partial g}{x_j}(a), \text{ donde } \frac{\partial g}{x_j} = \left(\frac{\partial g_1}{x_j}, \ldots, \frac{\partial g_m}{x_j} \right).$$

Notación sin ambigüedades (referenciando sólo direcciones):

$$D_{e_j}(f \circ g)(a) = \nabla f(g(a)) \cdot D_{e_j} g(a) = \sum_{i=1}^{m} D_{e_i} f(g(a)) D_{e_j} g_i(a).$$

Forma más informal e imprecisa (referenciando variables):

$$\frac{\partial (f \circ g)}{\partial x_j} = \sum_{i=1}^{m} \frac{\partial f}{\partial g_i} \frac{\partial g_i}{\partial x_j}.$$

Regla de la cadena

Ejemplos

1. *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ dada por*
$$f(s,t) = s^2 - t^2 + 1.$$
Se pide calcular
$$\frac{\partial}{\partial x}[f(x+2y, 3x-y)].$$
   - *Método 1. Proceder directamente:*
$$f(x+2y, 3x-y) = (x+2y)^2 - (3x-y)^2 + 1.$$
   Entonces
$$\frac{\partial}{\partial x}[f(x+2y, 3x-y)] = 2(x+2y) - 2(3x-y)\cdot 3 = 10y - 16x.$$

Regla de la cadena

Ejemplos

1. *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ dada por*

$$f(s,t) = s^2 - t^2 + 1. \to \frac{\partial f}{\partial s}(s,t) = 2s, \; \frac{\partial f}{\partial t}(s,t) = -2t$$

- *Método 2. Aplicar la regla de la cadena:*

$$f(x+2y, 3x-y) = f(s(x,y), t(x,y)) = (f \circ g)(x,y),$$

donde $g : \mathbb{R}^2 \to \mathbb{R}^2$ viene dada por

$$g(x,y) = (s(x,y), t(x,y)) = (x+2y, 3x-y) \to \begin{cases} \frac{\partial s}{\partial x}(x,y) = 1 \\ \frac{\partial t}{\partial x}(x,y) = 3 \end{cases}$$

Entonces

$$\frac{\partial (f \circ g)}{\partial x}(x,y) = \frac{\partial f}{\partial s}(g(x,y))\frac{\partial s}{\partial x}(x,y) + \frac{\partial f}{\partial t}(g(x,y))\frac{\partial t}{\partial x}(x,y)$$
$$= 2(x+2y) \cdot 1 - 2(3x-y) \cdot 3 = 10y - 16x.$$

Ejemplos

2. *Sea* $f : \mathbb{R}^2 \to \mathbb{R}$, *verificando*

$$\exists D_{e_1} f = D_{e_2} f.$$

Demuéstrese que la función $g : \mathbb{R} \to \mathbb{R}$ *dada por* $g(x) = f(x, -x)$ *es constante.* *Podemos ver* $g = f \circ w$, *donde* $w : \mathbb{R} \to \mathbb{R}^2$ *dada por* $w(x) = (x, -x)$. *Por tanto:*

$$\begin{aligned} g'(x) &= \frac{\partial (f \circ w)}{\partial x}(x) = D_{e_1} f(x, -x) \frac{\partial w_1}{\partial x}(x) + D_{e_2} f(x, -x) \frac{\partial w_2}{\partial x}(x) \\ &= D_{e_1} f(x, -x) \cdot 1 + D_{e_2} f(x, -x) \cdot (-1) \\ &= D_{e_1} f(x, -x) - D_{e_2} f(x, -x) = 0. \end{aligned}$$

Como $f'(x) = 0 \ \forall x \in \mathbb{R}$, *entonces* f *es constante.*

Derivadas de segundo orden

Al igual que en el caso de funciones de variable real, podemos considerar las derivadas de orden superior. Las derivadas de orden 2 son de gran interés en términos de optimización, que son de la forma

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial \frac{\partial f}{\partial x_i}}{\partial x_j}, \quad 1 \leq i, j \leq n.$$

Ejemplos

1. *Sea* $f(x, y) = xy + \sin(x)$, *luego* $\frac{\partial f}{\partial x}(x, y) = y + \cos(x)$, $\frac{\partial f}{\partial y}(x, y) = x$. *Por tanto:*
   - $\frac{\partial^2 f}{\partial x^2}(x, y) = -\sin(x)$.
   - $\frac{\partial^2 f}{\partial x \partial y}(x, y) = 1$.
   - $\frac{\partial^2 f}{\partial y \partial x}(x, y) = 1$.
   - $\frac{\partial^2 f}{\partial y^2}(x, y) = 0$.

Ejemplos

2. *Sea* $f(x,y) = e^{x^2-y^2} + 2x^5y^3$
   - $\frac{\partial f}{\partial x}(x,y) = 2xe^{x^2-y^2} + 10x^4y^3.$
     - $\frac{\partial^2 f}{\partial x^2}(x,y) = 2e^{x^2-y^2} + 4x^2e^{x^2-y^2} + 40x^3y^3.$
     - $\frac{\partial^2 f}{\partial x \partial y}(x,y) = -4xye^{x^2-y^2} + 30x^4y^2.$
   - $\frac{\partial f}{\partial y}(x,y) = -2ye^{x^2-y^2} + 6x^5y^2.$
     - $\frac{\partial f}{\partial y \partial x}(x,y) = -4xye^{x^2-y^2} + 30x^4y^2.$
     - $\frac{\partial^2 f}{\partial y^2}(x,y) = -2e^{x^2-y^2} + 4y^2e^{x^2-y^2} + 12x^5y.$

Teorema (Teorema de Schwarz)

Si existen las segundas derivadas de f y son continuas, entonces $\forall i,j \in \{1,2,\ldots,n\}$ se cumple $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$

Puntos críticos

Definición

Sea $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$*. Diremos que* $x \in A$ *es un* **punto crítico** *de* f *si* $\nabla f(x) = (0, \ldots, 0)$.

Ejemplo

Sea $f(x,y) = x^2 + y^2 - xy - 3x$*. Calculemos sus puntos críticos. Para ello, lo primero que debemos calcular es su gradiente:*

$$\nabla f(x,y) = (2x - y - 3, 2y - x).$$

Imponiendo la condición $\nabla f(x,y) = (0,0)$*, se obtiene el siguiente sistema de ecuaciones:*

$$\left.\begin{array}{rl} 2x - y &= 3 \\ -x + 2y &= 0 \end{array}\right\}$$

de donde se obtiene como solución $(x,y) = (2,1)$.

Hessiano

Definición

Sea $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ dos veces derivable en $x \in A$. Definimos la **matriz hessiana** *de f en el punto $x = (x_1, \ldots, x_n) \in A$ como la siguiente matriz $n \times n$:*

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

Teorema

Sea $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ *y* $a \in A$ *cumpliendo que* $\exists \nabla f(a) = 0$ *(es decir,* a *es un punto crítico de* f*) y* $\exists H_f(a)$. *Entonces:*

- *Si* $H_f(a)$ *es una matriz definida positiva, entonces* a *es un mínimo local de* f.
- *Si* $H_f(a)$ *es una matriz definida negativa, entonces* a *es un máximo local de* f.
- *Si los autovalores de* $H_f(a)$ *tienen signos diferentes y ninguno es nulo, entonces* a *es un* **punto de silla** *de* f.
- *Si hay algún valor propio nulo, entonces el criterio no es concluyente.*

Hessiano

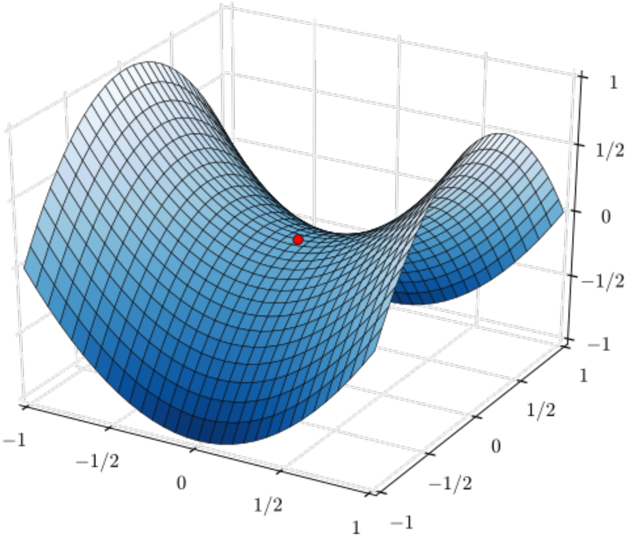
Caso particular $n = 2$:

Teorema

Sea $f : A \subseteq \mathbb{R}^2 \to \mathbb{R}$ *y* $(x_0, y_0) \in A$ *cumpliendo que* $\exists \nabla f(x_0, y_0) = 0$ *(es decir,* (x_0, y_0) *es un punto crítico de* f*) y* $\exists H_f(x_0, y_0)$*. Entonces:*

- *Si* $\det\left(H_f(x_0, y_0)\right) > 0$*, tenemos las siguientes posibilidades:*
  - *Si* $\frac{\partial^2 f}{\partial x^2}(x_0, y_0) > 0$*, entonces* (x_0, y_0) *es un mínimo local de* f*.*
  - *Si* $\frac{\partial^2 f}{\partial x^2}(x_0, y_0) < 0$*, entonces* (x_0, y_0) *es un máximo local de* f*.*
- *Si* $\det\left(H_f(x_0, y_0)\right) < 0$*, entonces* (x_0, y_0) *es un punto de silla de* f*.*
- *Si* $\det\left(H_f(x_0, y_0)\right) = 0$*, entonces el criterio no es concluyente.*

Hessiano



Punto de silla.

Hessiano

Ejemplo

La matriz hessiana asociada a la función del ejemplo anterior, $f(x,y) = x^2 + y^2 - xy - 3x$, *es*

$$H_f(x,y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(x,y) & \frac{\partial^2 f}{\partial x \partial y}(x,y) \\ \frac{\partial^2 f}{\partial y \partial x}(x,y) & \frac{\partial^2 f}{\partial y^2}(x,y) \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

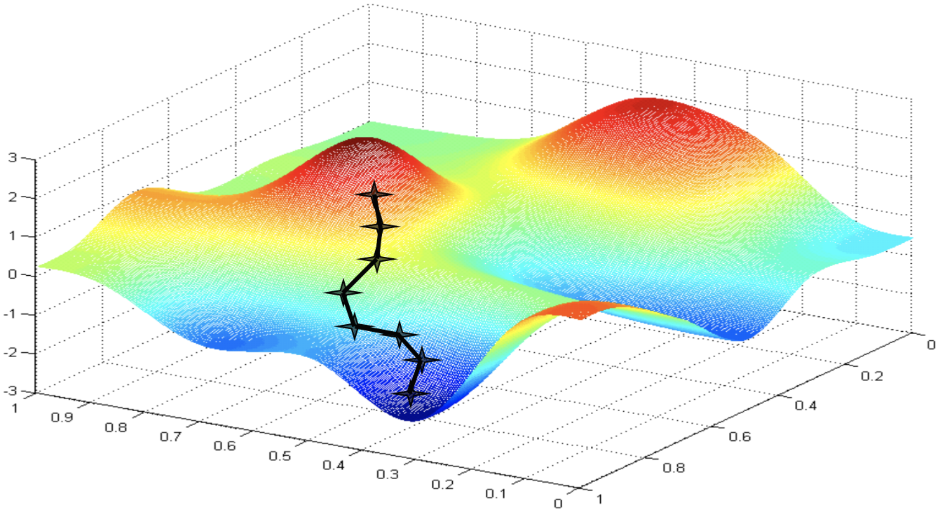
En particular, como $H_f(x,y)$ *es constante, si evaluamos en el punto crítico,* $(2,1)$, *obtenemos*

$$H_f(2,1) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

Como $\det\left(H_f(2,1)\right) = 3 > 0$ *y* $\frac{\partial^2 f}{\partial x^2}(2,1) = 2 > 0$, *podemos concluir que* $(2,1)$ *es un mínimo local de* f.

Descenso de gradiente

Descenso de gradiente: $x_{n+1} = x_n - \gamma_n \nabla f(x_n)$



Descenso de gradiente

- Sea $f : A \subseteq \mathbb{R}^k \to \mathbb{R}$ tal que $\exists \nabla f$ y $x_0 \in A$ un punto de partida. Entonces el método de descenso de gradiente se puede expresar como sigue:

$$x_{n+1} = x_n - \gamma_n \nabla f(x_n),$$

  donde $\gamma_n > 0$ es el ratio de aprendizaje.
- Posibles elecciones del ratio de aprendizaje (entre muchas otras):
  - Constante: $\gamma_n = h$, para $h > 0$ suficientemente pequeño.
  - Variable. Dado $h > 0$ suficientemente pequeño:

$$\gamma_n = \begin{cases} h & n = 0 \\ \dfrac{|(x_n - x_{n-1}) \cdot [\nabla f(x_n) - \nabla f(x_{n-1})]|}{\|\nabla f(x_n) - \nabla f(x_{n-1})\|_2^2} & n \geq 1 \end{cases}$$

  - Obsérvese que el caso variable se reduce al método de la secante modificado visto para $k = 1$ (ejercicio).

¡Muchas gracias!

Universidad Internacional de Valencia

Contacto:

david.zorio@campusviu.es