Machine Learning

Rodrigo Barrera

Sea $S \subseteq R^n$, un conjunto convexo y no vacío, y sea $f : S \to R$, decimos que f es una función convexa en S, si y solo si:

$$\mathrm{f}\left[\lambda \mathrm{x}_1 + (1-\lambda)\mathrm{x}_2\right] \leq \lambda \mathrm{f}\left(\mathrm{x}_1\right) + (1-\lambda)\mathrm{f}\left(\mathrm{x}_2\right)$$

$$\forall \lambda \in [0,1] \quad \text{y} \quad \forall \mathrm{x}_1, \mathrm{x}_2 \in \mathrm{S}.$$

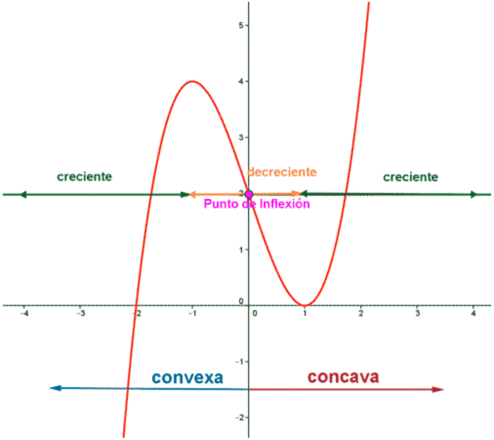
creciente
decreciente
creciente
Punto de Inflexión
convexa
concava

Figure:

Sea $S \subseteq R^n$, un conjunto convexo y no vacío, y sea $f : S \rightarrow R$, decimos que f es una función estrictamente convexa en S, si:

$$\mathrm{f}\left[\lambda \mathrm{x}_1 + (1-\lambda)\mathrm{x}_2\right] < \lambda \mathrm{f}\left(\mathrm{x}_1\right) + (1-\lambda)\mathrm{f}\left(\mathrm{x}_2\right)$$

$$\forall \lambda \in]0,1[\quad y \quad \forall \mathrm{x}_1, \mathrm{x}_2 \in \mathrm{S}.\,\text{con}\,\mathrm{x}_1 \neq \mathrm{x}_2$$

Sea $S \subseteq R^n$, un conjunto convexo y no vacío, y sea $f : S \to R$, decimos que f es una función cóncava en S, si y solo si:

$$\mathrm{f}\left[\lambda \mathrm{x}_1 + (1-\lambda)\mathrm{x}_2\right] \geq \lambda \mathrm{f}\left(\mathrm{x}_1\right) + (1-\lambda)\mathrm{f}\left(\mathrm{x}_2\right)$$

$$\forall \lambda \in [0,1] \quad y \quad \forall \mathrm{x}_1, \mathrm{x}_2 \in \mathrm{S}.$$

Una función es estrictamente cóncava si la desigualdad se verifica en sentido estricto, es decir:

$$f\left[\lambda x_1 + (1-\lambda)x_2\right] > \lambda f\left(x_1\right) + (1-\lambda) f\left(x_2\right)$$

$$\forall \lambda \in (0,1) \quad \text{y} \quad \forall x_1, x_2 \in S. \text{ con } x_1 \neq x_2$$

Para las funciones con derivadas continuas hasta el segundo orden, es posible emplear una caracterización fundamentada en el hessiano de la función. Consideremos una función $f : S \subset \mathbb{R}^n \to \mathbb{R}$, donde S es un conjunto convexo y no vacío, y supongamos que f es suficientemente suave, es decir, $f \in C^2(S)$, lo que indica que tiene derivadas segundas continuas en todo S. En este contexto, se pueden aplicar ciertos criterios basados en el hessiano de f para determinar propiedades de convexidad o concavidad de la función en el conjunto S.

a) f es convexa en S sii se cumple $Hf(x)$ es semidefinida positiva en S.

b) f es cóncava en S sii se cumple que $Hf(x)$ es semidefinida negativa en S.

c) f es estrictamente convexa solamente si $Hf(x)$ es definida positiva en S.

d) f es estrictamente cóncava solamente si $Hf(x)$ es definida negativa en S.

La matriz hessiana es una matriz cuadrada de segundas derivadas parciales de una función multivariable. Se utiliza para analizar la curvatura de dicha función, lo que permite estudiar extremos locales y puntos de silla. La matriz se define como:

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

donde f es la función y $x_1, \ldots, x_n$ son las variables. En el contexto de optimización, la matriz hessiana ayuda a determinar si un punto crítico es un mínimo, máximo o punto de silla.

Ejercicio:

Sea $f : \mathbb{R}^3 \to \mathbb{R}$ definida por $f(x, y, z) = -\left(x^2 + y^2 + z^2\right)$. Demuestra que f es cóncava en todo su dominio.

Para demostrar que la función $f(x, y, z) = -(x^2 + y^2 + z^2)$ es cóncava en todo su dominio $\mathbb{R}^3$, calculamos primero las segundas derivadas parciales de f:

$$\frac{\partial^2 f}{\partial x^2} = -2,$$
$$\frac{\partial^2 f}{\partial y^2} = -2,$$
$$\frac{\partial^2 f}{\partial z^2} = -2.$$

Las derivadas cruzadas son todas cero porque f no tiene términos mixtos:

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 0,$$
$$\frac{\partial^2 f}{\partial x \partial z} = \frac{\partial^2 f}{\partial z \partial x} = 0,$$
$$\frac{\partial^2 f}{\partial y \partial z} = \frac{\partial^2 f}{\partial z \partial y} = 0.$$

La matriz hessiana de f es entonces:

$$H_f = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

Una función es cóncava si su matriz hessiana es semidefinida negativa. En este caso, H_f es claramente semidefinida negativa, ya que todos sus autovalores son negativos (-2 para cada una de las direcciones x, y, z). Por lo tanto, podemos concluir que $f(x, y, z) = -(x^2 + y^2 + z^2)$ es cóncava en todo $\mathbb{R}^3$.

Una matriz definida positiva es una matriz cuadrada A tal que para cualquier vector no nulo x, la forma cuadrática $x^T Ax$ es siempre positiva. Esto implica que todos sus valores propios son positivos y la matriz es invertible. En términos de optimización, una matriz hessiana definida positiva en un punto crítico indica un mínimo local.

Una matriz semidefinida positiva también es una matriz cuadrada A pero permite que la forma cuadrática $x^T Ax$ sea cero o positiva para cualquier vector no nulo x.

Para demostrar que la matriz H_f dada es semidefinida negativa sin usar valores propios, debemos mostrar que para todo vector no nulo $x \in \mathbb{R}^3$, la forma cuadrática $x^T H_f x$ es menor o igual a cero. La matriz H_f es:

$$H_f = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

Tomamos un vector arbitrario $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$. Calculamos la forma cuadrática $x^T H_f x$:

$$x^T H_f x = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Multiplicamos la matriz por el vector x:

$$= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} -2x_1 \\ -2x_2 \\ -2x_3 \end{bmatrix}$$

Realizamos el producto escalar:

$$= -2x_1^2 - 2x_2^2 - 2x_3^2$$

Dado que x_1^2, x_2^2, y x_3^2 son siempre no negativos para cualquier $x_1, x_2, x_3 \in \mathbb{R}$, la expresión $-2x_1^2 - 2x_2^2 - 2x_3^2$ es siempre no positiva. Esto demuestra que para cualquier vector no nulo x, $x^T H_f x \leq 0$, lo que significa que H_f es semidefinida negativa.

Ejercicios

1. Considera la función $f : \mathbb{R}^3 \to \mathbb{R}$ dada por $f(x, y, z) = \ln(x + y + z)$, donde $x + y + z > 0$. Demuestra que f es cóncava en su dominio.
2. Sea $f : \mathbb{R}^3 \to \mathbb{R}$ definida como $f(x, y, z) = -(x^4 + y^4 + z^4)^{\frac{1}{4}}$. Verifica que f es cóncava en $\mathbb{R}^3_+$, el conjunto de todos los puntos en $\mathbb{R}^3$ con coordenadas no negativas.