

5: Optimización y Método del Gradiente Descendiente

Vamos a describir el método del gradiente descendente, que sirve para encontrar un punto donde la función $f: \mathbb{R}^n \rightarrow \mathbb{R}$ alcanza un valor mínimo.

El proceso se puede resumir en:

- Comenzamos eligiendo un valor arbitrario inicial de \mathbf{x} .
 - Buscamos la dirección \mathbf{v} en la que f decrece más rápidamente alrededor de \mathbf{x} . A \mathbf{v} le llamamos *dirección de máximo descenso*.
 - Variamos “un poco” los valores iniciales en la dirección \mathbf{v} :
- $$\mathbf{x} \rightarrow \mathbf{x} + \varepsilon \mathbf{v} \quad (5.1)$$
- Repetimos.

Nos detenemos ahora en el segundo paso de este proceso: cómo determinar la dirección de descenso más rápido. Como se ha visto en la sección anterior:

$$f(\mathbf{x} + \mathbf{v}) \approx f(\mathbf{x}) + \mathbf{v} \nabla f(\mathbf{x}). \quad (5.2)$$

Por la definición de producto escalar, suponiendo que \mathbf{v} es un vector de módulo 1, la expresión anterior se escribe

$$f(\mathbf{x} + \mathbf{v}) \approx f(\mathbf{x}) + \mathbf{v} \nabla f(\mathbf{x}) = f(\mathbf{x}) + \|\nabla f(\mathbf{x})\| \cos(\theta), \quad (5.3)$$

donde θ es el ángulo que forman los vectores \mathbf{v} y $\nabla f(\mathbf{x})$.

Con vistas a que f decrezca lo más rápido posible, debemos elegir θ de forma que $\cos(\theta)$ sea lo menor posible, esto es, $\cos(\theta) = -1$.

Esto ocurre cuando θ vale π (180°) o, lo que es lo mismo, si \mathbf{v} tiene el sentido opuesto a $\nabla f(\mathbf{x})$.

El proceso queda reescrito como:

- Comenzamos eligiendo un valor arbitrario inicial de \mathbf{x} .
 - La dirección \mathbf{v} que hace descender más rápidamente a f alrededor de \mathbf{x} es $-\nabla f(\mathbf{x})$. Calculamos $\nabla f(\mathbf{x})$.
 - Variamos un poco los valores iniciales en la dirección \mathbf{v} :
- $$\mathbf{x} \rightarrow \mathbf{x} - \varepsilon \nabla f(\mathbf{x}) \quad (5.4)$$
- Repetimos.

Para aplicar este algoritmo iterativo es necesario fijar dos parámetros: el número de iteraciones y el valor de ε (que se conoce como *tasa de aprendizaje*).

- Se dirá que el algoritmo converge si la sucesión de soluciones en el proceso iterativo se acerca a la solución óptima (esto es, el mínimo global). La cantidad de iteraciones que se necesita para que el método del gradiente descendente converja puede variar mucho (dependiendo del problema, desde unas pocas iteraciones hasta miles de ellas). En cualquier caso, la tasa de aprendizaje condicionará el número mínimo de iteraciones necesario para la convergencia.
- La tasa de aprendizaje ε es una constante, con $0 < \varepsilon < 1$, que afectará de forma significativa la efectividad del algoritmo: un valor demasiado pequeño puede llevar a que el algoritmo precise un número elevado de iteraciones antes de converger; mientras que si el valor asignado es demasiado grande, puede saltarse el mínimo global (pasando a tomar valores mayores) y no proporcionar una buena solución.

This page titled 5: Optimización y Método del Gradiente Descendiente is shared under a [not declared](#) license and was authored, remixed, and/or curated by Joaquín López Herraiz.