Gradiente Descendiente

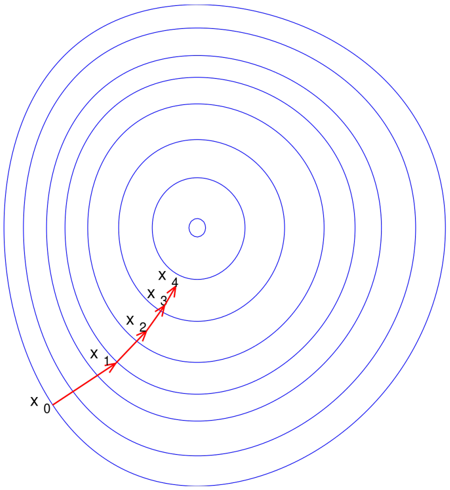
Luis Norberto Zúñiga Morales

21 de agosto de 2022

Gradiente Descendiente

- Gradiente Descendiente es un algoritmo de optimización iterativo de primer orden.
- Permite encontrar mínimos locales en una función diferenciable.
- La idea básica es dar pasos pequeños en dirección contraria al gradiente.

Gradiente Descendiente



Gradiente Descendiente

Requisitos:

- La función debe ser diferenciable.

Requisitos:

- La función debe ser diferenciable.
  - ¿Qué significa que una función sea *diferenciable*?

Gradiente Descendiente

Requisitos:

- La función debe ser diferenciable.
  - ¿Qué significa que una función sea *diferenciable*?
- La función debe ser convexa.

Gradiente Descendiente

Requisitos:

- La función debe ser diferenciable.
  - ¿Qué significa que una función sea *diferenciable*?
- La función debe ser convexa.
  - ¿Qué significa que una función sea *convexa*?

$$f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2)$$

Gradiente Descendiente

Ejercicio #1

Dibujen una función diferenciable.

Gradiente Descendiente

Ejercicio #1

Dibujen una función diferenciable.

Ejercicio #2

Dibujen un función convexa y una que no sea convexa.

Gradiente Descendiente

Definición: Gradiente

Para $f : \mathbb{R}^n \to \mathbb{R}$ su gradiente $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ en un punto $p = (x_1, \ldots, x_n)$ se define como:

$$\nabla f(p) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(p) \\ \vdots \\ \frac{\partial f}{\partial x_n}(p) \end{bmatrix}.$$

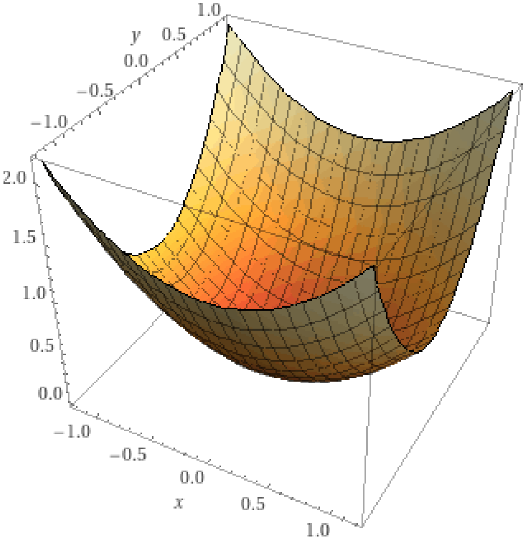
Ejercicio #3

Determinar al gradiente de la función

$$f(x, y) = 0{,}5x^2 + y^2$$

en el punto $p = (5, 5)$.

Gradiente Descendiente



- Gradiente Descendiente calcula iterativamente el siguiente punto usando el gradiente en el punto en turno.
- Lo escala (razón de aprendizaje).
- Resta este resultado a la posición actual.

- Gradiente Descendiente calcula iterativamente el siguiente punto usando el gradiente en el punto en turno.
- Lo escala (razón de aprendizaje).
- Resta este resultado a la posición actual.

Gradiente Descendiente

$$p_{m+1} = p_m - \eta \nabla f(p_m)$$

Gradiente Descendiente

$$p_{m+1} = p_m - \eta \nabla f(p_m)$$

- El parámetro η permite escalar el valor del gradiente, lo que hace cada paso más grande o más pequeño.

Gradiente Descendente

$$p_{m+1} = p_m - \eta \nabla f(p_m)$$

- El parámetro η permite escalar el valor del gradiente, lo que hace cada paso más grande o más pequeño.
- En Machine Learning, η es la razón de aprendizaje (*learning rate*).

Gradiente Descendiente

$$p_{m+1} = p_m - \eta \nabla f(p_m)$$

- El parámetro η permite escalar el valor del gradiente, lo que hace cada paso más grande o más pequeño.
- En Machine Learning, η es la razón de aprendizaje (*learning rate*).
  - Si es muy pequeño, tarda más en converger.

Gradiente Descendiente

$$p_{m+1} = p_m - \eta \nabla f(p_m)$$

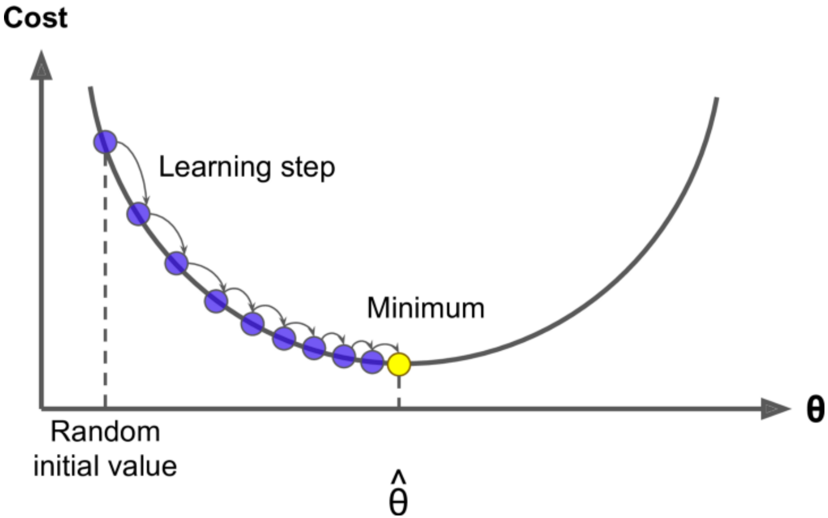
- El parámetro η permite escalar el valor del gradiente, lo que hace cada paso más grande o más pequeño.
- En Machine Learning, η es la razón de aprendizaje (*learning rate*).
    - Si es muy pequeño, tarda más en converger.
    - Si es muy grande, da saltos grandes, inclusive no llegando a converger.

1. Elegir un punto de salida (random).
2. Calcular el gradiente en ese punto.
3. Determinar el nuevo punto según

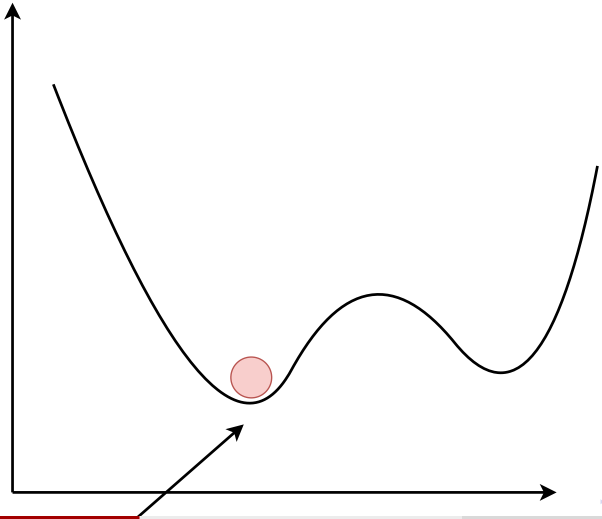
$$p_{m+1} = p_m - \eta \nabla f(p_m)$$

   .
4. Condición de paro:
   - Número máximo de iteraciones.
   - El tamaño del paso es menor que un valor de tolerancia.

Gradiente Descendiente



Ejercicio: Consideren el siguiente estado del gradiente descendiente. ¿Qué resultaría en la siguiente iteración del algoritmo?



Consideraciones Prácticas

Una de las preguntas que se pueden realizar es cómo elegir la razón de aprendizaje η ya que no es un parámetro obvio.

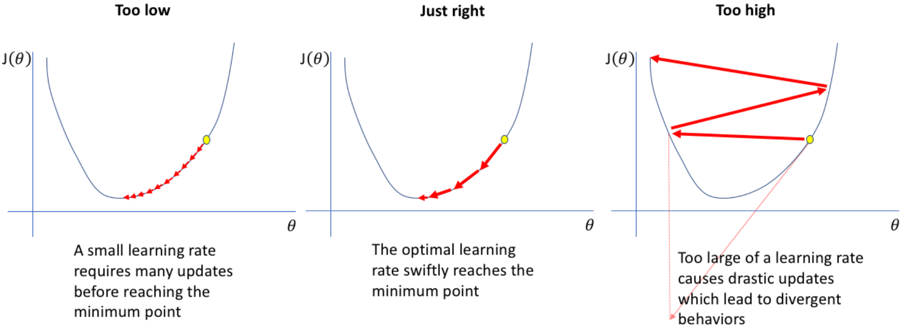
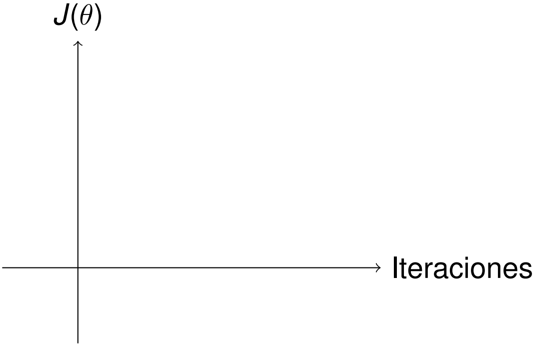


Figura: Efectos de la razón de aprendizaje.

Consideraciones Prácticas

La mejor forma es graficar diferentes valores de η vs el número de iteraciones.



Consideraciones Prácticas

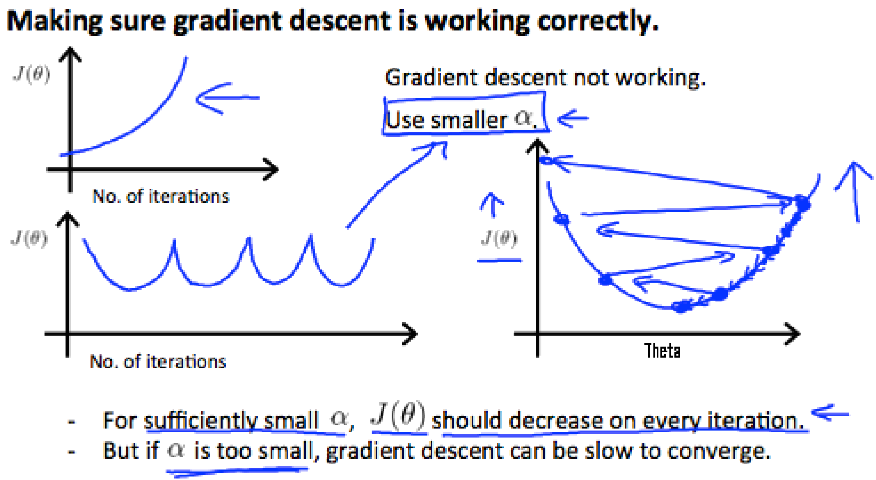


Figura: Probables problemas con el gradiente descendiente.

Consideraciones Prácticas

Razón de Aprendizaje

Lo mejor es elegir η pequeño. Si es lo suficientemente pequeño, gradiente descendiente converge. Si es muy pequeño, tarda en converger.

Lo mejor en práctica es utilizar valores iniciales de $\eta = 0{,}001, 0{,}01, 0{,}1$ y graficar su rendimiento. Aumentar gradualmente (doble o triple) el mejor valor para encontrar un buen parámetro.