

Práctica 1: Aprendizaje Automático

1. Ejercicios sobre la búsqueda iterativa de óptimos

Eloy Bedia García

10 de Marzo de 2018

EJERCICIO 1

1. Implementar el algoritmo de gradiente descendente.

$var \leftarrow Punto_{inicio}$

$f : \mathbb{R}^n \rightarrow \mathbb{R}$

$fderivadas \leftarrow (\frac{\partial f}{\partial x_0}, \dots, \frac{\partial f}{\partial x_i}, \dots, \frac{\partial f}{\partial x_n})$

$\eta \leftarrow Tasa_{aprendizaje}$

Sea $a \in \mathbb{R}^n$ el anterior punto evaluado segun BGD, $\epsilon \in \mathbb{R} \mid |f(x) - f(a)| \geq \epsilon$

$limit \leftarrow Max_{iteraciones}$

```
BatchGradientDescent = function(var, f, fderivadas, mu, epsilon = -Inf, limit = Inf) {  
  #el vector 'y' sera utilizado para almacenar las nuevas componentes  
  #Si las guardaramos directamente en 'var' alterariamos el punto a evaluar en ese momento.  
  y <- var  
  dim(y) <- dim(var)  
  iteraciones <- 0  
  
  eval_anterior <- Inf  
  while(iteraciones < limit && abs(eval_anterior - eval(f)) >= epsilon) {  
  
    #Recorremos cada una de las componentes independientes de la funcion  
    #Para cada componente de la funcion hay una derivada parcial  
    for(i in 1:dim(var)) {  
      #Evaluamos la derivada en el punto 'var' y seguimos el sentido negativo de la misma  
      #Si eval(fderivada[i]) < 0 la variable 'var[i]' aumentara siguiendo el sentido negativo  
      #Si eval(fderivada[i]) > 0 la variable 'var[i]' disminuira siguiendo el sentido negativo  
      #Si eval(fderivada[i]) = 0 la variable 'var[i]' habra llegado a su valor optimo, por tanto no se  
      y[i] <- var[i] - mu * eval(fderivadas[i])  
    }  
  
    eval_anterior <- eval(f)  
    #Actualizamos el punto que evaluaremos  
    var <- y  
    iteraciones <- iteraciones + 1  
  
  }  
  
  list(y, iteraciones)  
}
```

EJERCICIO 2

2. Considerar la función $f(u, v) = (u^3 e^{v-2} - 4v^3 e^{-u})^2$. Usar gradiente descendente para encontrar un mínimo de esta función, comenzando desde el punto $(u, v) = (1, 1)$ y usando una tasa de aprendizaje $\eta = 0,1$.

a) Calcular analíticamente y mostrar la expresión del gradiente de la función $f(u, v)$

$$\nabla f = \left(\frac{\partial (u^3 e^{v-2} - 4v^3 e^{-u})^2}{\partial u}, \frac{\partial (u^3 e^{v-2} - 4v^3 e^{-u})^2}{\partial v} \right)$$

$$\nabla f = (2(u^3 e^{v-2} - 4v^3 e^{-u})(3u^2 e^{v-2} + 4v^3 e^{-u}), 2(u^3 e^{v-2} - 4v^3 e^{-u})(u^3 e^{v-2} - 12v^2 e^{-u}))$$

b) Cuántas iteraciones tarda el algoritmo en obtener por primera vez un valor de $f(u, v)$ inferior a 10^{-14} .

Para al vanzar el minimo de la función f, el algoritmo utiliado invierte 620198

c) ¿En qué coordenadas (u, v) se alcanzó por primera vez un valor igual o menor a 10^{-14} en el apartado anterior.

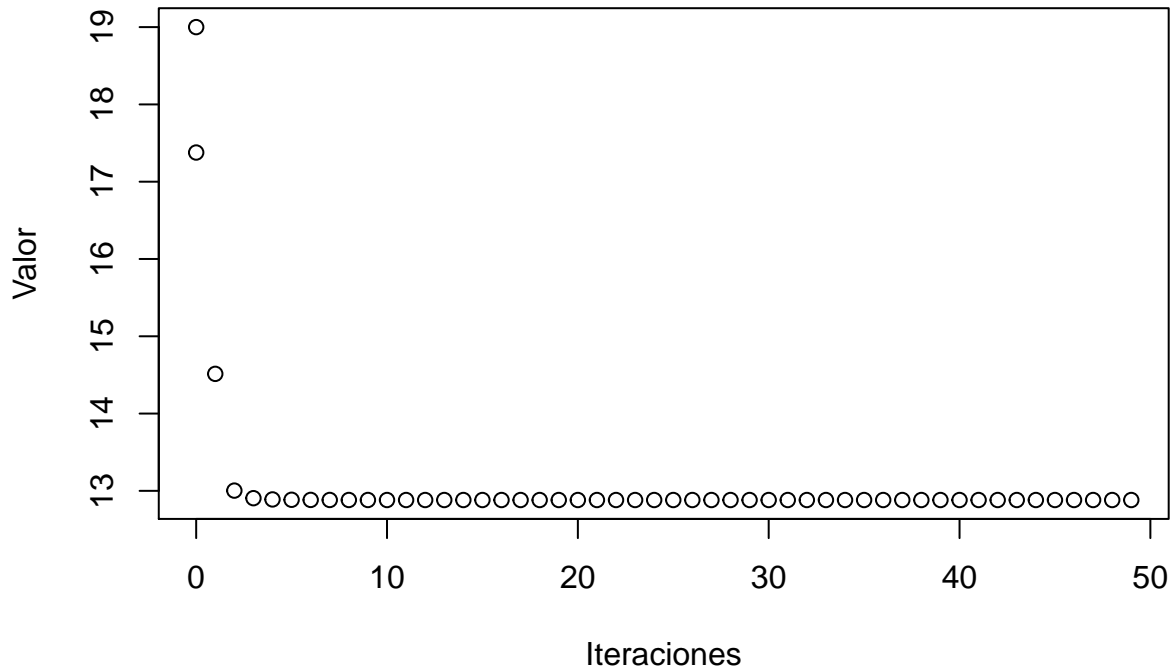
En el punto (0.07794586 , -0.004884877). El valor f(u,v) en ese punto es 4.122791e-09

EJERCICIO 3

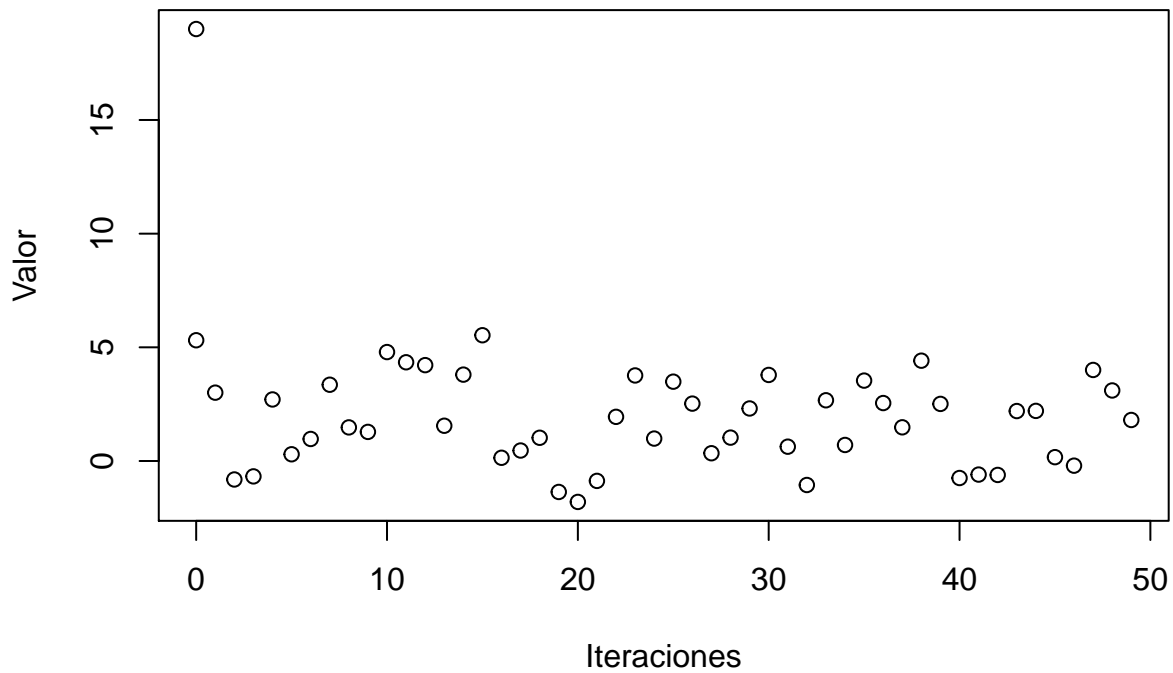
3. Considerar ahora la función $f(x, y) = (x - 2)^2 + 2(y + 2)^2 + 2\sin(2\pi x)\sin(2\pi y)$

a) Usar gradiente descendente para minimizar esta función. Usar como punto inicial $(x_0 = 1, y_0 = 1)$, tasa de aprendizaje $\eta = 0.01$ y un máximo de 50 iteraciones. Generar un gráfico de cómo desciende el valor de la función con las iteraciones. Repetir el experimento pero usando $\eta = 0, 1$, comentar las diferencias y su dependencia de η .

1º Experimento: Aplicamos el gradiente descendente con $\eta = 0.01$



2º Experimento: Aplicamos el gradiente descendente con $\eta = 0.1$



La única diferencia entre ambos experimentos ha sido el valor η , sin embargo, con esta alteración podemos observar que el primero a conseguido alcanzar el valor mínimo posible, por el contrario, el segundo ha estado lejos de conseguirlo.

El valor $f(u, v)$ en el segundo experimento podemos ver que ha estado aproximadamente en el intervalo $(-1, 7)$.

Recordemos que η es un factor multiplicativo que condiciona el incremento (o decremento) de la variable independiente.

El segundo experimento tenía una tasa de aprendizaje (η) con valor 0,1. Con lo dicho anteriormente, podemos deducir que el valor $f(u, v)$ ha estado oscilando entorno al mínimo debido a su alta tasa de aprendizaje. Es decir, cuando estaba cerca del mínimo (tanto en sentido positivo como negativo) debido a su tasa de aprendizaje incrementabamos (o decrementabamos) demasiado el valor de la variable independiente, por lo que en vez de llegar al valor del mínimo, pasabamos al otro lado del mismo. Este proceso se ha repetido numerosas veces y por eso su grafica muestra tantos picos.

En el caso del 1º experimento, la tasa de aprendizaje era adecuada y por ello en ningún momento hemos oscilado entorno al mínimo.

b) Obtener el valor mínimo y los valores de las variables (x, y) en donde se alcanzan cuando el punto de inicio se fija: (2,1, -2,1), (3, -3), (1,5, 1,5), (1, -1). Generar una tabla con los valores obtenidos

##		X	Y	Z
##	P(2.1, -2.1)	2.242323	-2.236772	-1.8199354
##	P(3, -3)	2.732558	-2.714534	-0.3810873
##	P(1.5, 1.5)	1.753628	1.058348	18.0509626
##	P(1, -1)	1.267442	-1.285466	-0.3810873

EJERCICIO 4

4. ¿Cuál sería su conclusión sobre la verdadera dificultad de encontrar el mínimo global de una función arbitraria?
 - a) Encontrar un punto de inicio que te permita llegar al óptimo global sin quedar atascado en óptimos locales.
 - b) Encontrar una tasa de aprendizaje que sea lo suficientemente alta como para encontrar el óptimo rápidamente pero sin llegar a oscilar entorno a él.
 - c) Encontrar un criterio de parada fiable