Thuật Toán Gradient Descent.

1. Cách tìm điểm cực trị, điểm yên ngựa của hàm f(θ) theo kiến thức đã học.

Cho f(x,y)

Cách làm:

* Đạo hàm riêng theo x, y: f’X , f’Y
* Giải hệ phương trình để tìm ra điểm M(xM, yM) :
  + f’x = 0
  + f’y = 0
* Đạo hàm riêng lần 2 theo x, y: f’’XX , f’’YY, f’’XY
* Tính D = f’’XX \* f’’YY – ( f’’XY) 2
* Tính
  + A = D(xM, yM)
  + Nếu A < 0 thì f đạt yên ngựa tại điểm M
  + Nếu A = 0 thì không xác định (Do chương trình học hiện tại, học sâu hơn sẽ rõ hơn)
  + Nếu A > 0 và f’’XX(xM, yM) > 0 thì tại M là cực tiểu của f
  + Nếu A > 0 và f’’XX(xM, yM) < 0 thì tại M là cực đại của f

2. Thuật Toán Gradient Descent

Giả sử ta cần tìm global minimum cho hàm f(θ) trong đó θ (theta) là một vector, thường được dùng để ký hiệu tập hợp các tham số của một mô hình cần tối ưu (trong Linear Regression thì các tham số chính là hệ số w). Đạo hàm của hàm số đó tại một điểm θ bất kỳ được ký hiệu là ∇θf(θ) (hình tam giác ngược đọc là nabla). Thuật toán GD cho hàm nhiều biến cũng bắt đầu bằng một điểm dự đoán θ0, sau đó, ở vòng lặp thứ t, quy tắc cập nhật là:

θt +1 = θt – η.∇θf(θt)

Hoặc viết dưới dạng đơn giản hơn: θ = θ – η.∇θf(θ).

3. Phần mềm sử dụng, hàm và thuật toán được sử dụng.

3.1. Phần mềm sử dụng

Phần mềm matlab (<https://www.mathworks.com/>)

3.2. Các lệnh trong mã đoạn.

Trong mã đoạn để viết thuật toán Gradient Descent, chủ yếu sử dụng các lệnh cơ bản:

|  |  |  |
| --- | --- | --- |
| STT | Tên lệnh | Công dụng |
| 1 | disp() | Xuất dữ liệu ra màn hình |
| 2 | diff() | Đạo hàm |
| 3 | subs() | Tính giá trị của hàm số tại điểm nào đó |
| 4 | while | Tạo một vòng lặp |
| 5 | If - else | Câu lệnh điều kiện |
| 6 | Abs() | Trị tuyệt đối |

3.3 Thuật toán sử dụng trong quá trình thực hiện:

Noisy Gradient Descent: Làm nhiễu giá trị của theta tiếp theo để tránh giá trị ở yên ngựa và tiếp tục đến điểm cực trị. Quy tắc là:

θt+1 = θt – η.∇θf(θt) + ᶓ

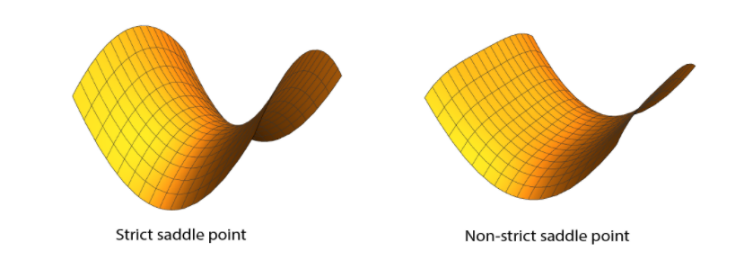
Lưu ý thuật toán Noisy Gradient Descent: thuật toán chỉ đúng khi có điểm ngặt. Nếu không có điểm ngặt mã đoạn sẽ tiếp tục chạy mãi có thể không dừng được.

Figure Điểm ngặt – Điểm không ngặt

4. Mã đoạn của thuật toán Gradient Descent. (Sử dụng Matlab)

4.1. Mã đoạn thuật toán Gradient Descent.

|  |  |
| --- | --- |
| 1 | function myGD(f,x1,x2) |
| 2 |  |
| 3 | warning off |
| 4 | format compact |
| 5 | format long g |
| 6 | syms x y |
| 7 | %% |
| 8 |  |
| 9 | while true |
| 10 | a = x1 - 0.1\*subs(subs(diff(f,x),x,x1),y,x2) ; |
| 11 | b = x2 - 0.1\*subs(subs(diff(f,y),x,x1),y,x2) ; |
| 12 | x1 = a; |
| 13 | x2 = b; |
| 14 |  |
| 15 |  |
| 16 | x1 = double(x1); |
| 17 | x2 = double(x2); |
| 18 | disp('x = '); |
| 19 | disp(x1); |
| 20 | disp('y = '); |
| 21 | disp(x2); |
| 22 | dhx=(subs(subs(diff(f,x),x,x1),y,x2)); |
| 23 | dhx = double(dhx); |
| 24 | disp(dhx); |
| 25 | dhy=(subs(subs(diff(f,y),x,x1),y,x2)); |
| 26 | dhy = double(dhy); |
| 27 | disp(dhy); |
| 28 | pause(0.5); |
| 29 |  |
| 30 |  |
| 31 | if (abs(subs(subs(diff(f,x),x,x1),y,x2)) < 10^-3) && (abs(subs(subs(diff(f,y),x,x1),y,x2)) < 10^-3) |
| 32 | A = diff(diff(f,x),x); |
| 33 | B = diff(diff(f,y),y); |
| 34 | C = diff(diff(f,x),y); |
| 35 | D = A\*B - C^2; |
| 36 | if subs(subs(D,x,x1),y,x2) < 0 |
| 37 | x1 = x1 + 10; |
| 38 | x2 = x2 + 10; |
| 39 | continue |
| 40 | else |
| 41 | break |
| 42 | end |
| 43 | end |
| 44 | end |
| 45 | %% |
| 46 | dhx=(subs(subs(diff(f,x),x,x1),y,x2)); |
| 47 | dhx = double(dhx); |
| 48 | disp(dhx); |
| 49 | dhy=(subs(subs(diff(f,y),x,x1),y,x2)); |
| 50 | dhy = double(dhy); |
| 51 | disp(dhy); |
| 52 | x1 = double(x1); |
| 53 | x2 = double(x2); |
| 54 | disp('x = '); |
| 55 | disp(x1); |
| 56 | disp('y = '); |
| 57 | disp(x2); |

4.2. Mô tả mã đoạn:

Dòng 10 đến dòng 13, sử dụng phương pháp đã đề cập ở mục 2 để tìm ra theta tiếp theo (với η = 0.1). η là learning rate.

Dòng 16 đến dòng 28, để in ra màn hình giá trị của theta mới – chính là 2 thành phần vector x và y, và đạo hàm f theo x và theo y mới.

Theo lý thuyết, cực trị của hàm số f(θ) được xác định ở những điểm làm đạo hàm của fx và fy = 0. Ở đây các điểm θ gần với giá trị θ\*(là cực trị chính xác của hàm số) nên ta xét các điều kiện của đạo hàm tiến gần đến 0, ở dòng 31, ta chọn điều kiện nhỏ hơn 10-3 là gần đến 0.

Sử dụng lí thuyết để kiểm tra điểm yên ngựa. Nếu rơi vào điểm yên ngựa, sử dụng thuật toán Noisy Gradient Descent để bỏ qua điểm yên ngựa.

4.3 Cách sử dụng mã đoạn.

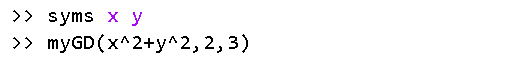
4.3.a. f(x,y) = x2+y2

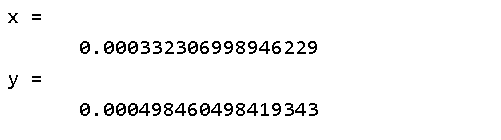
Trên Command Window, ta khai báo x và y thông qua lệnh:

|  |
| --- |
| syms x y |

Và gọi lại hàm myGD truyền vào các thông số cần thiết, đó là hàm số và giá trị theta bắt đầu để hàm số tiến về cực trị

|  |
| --- |
| myGD(x^2+y^2,2,3) |

Phần nhập:

Kết quả:

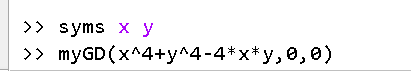
Chúng ta thấy x và y đang tiến dần về (0,0), là kết quả của thuật toán Gradient Descent với hàm f(x,y) = f(x,y) = x2+y2.

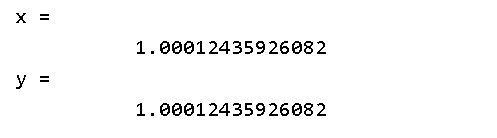
Ở bài đầu tiên chúng ta không có điểm yên ngựa nên việc xử lí của thuật toán rất đơn giản.

4.3.b. f(x,y) = x4+ y4-4xy

Bài tiếp theo này, chúng ta bị vướng phải một điểm yên ngựa, và bằng cách tính cơ bản theo lí thuyết đã học thì điểm yên ngựa là điểm (0,0). Vì thế ở bước nhập vào để chứng minh thuật toán Noisy Gradient Descent có thể bỏ qua được điểm yên ngựa, điểm bắt đầu sẽ là (0,0). Tuy nhiên để có độ chính xác hơn thì ta xét learning rate = 0.001.

Nhập



Xuất

Như thế thuật toán Noisy Gradient Descent giúp ta bỏ qua điểm yên ngựa để đến điểm cực trị. Tuy nhiên nó phụ thuộc vào lượng epsilon mà ta cộng vào và do độ lớn của learning rate.

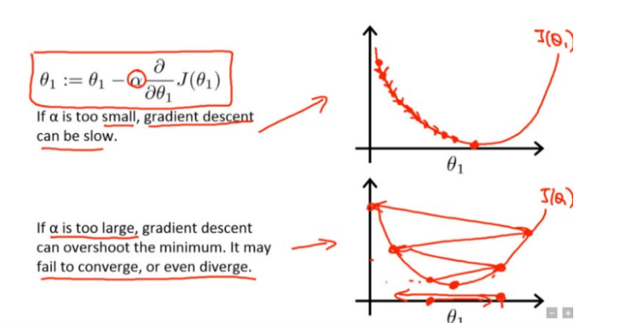
5. Nhược điểm của thuật toán Gradient Descent.

5.1 Learning Rate.

Như đã nói trên, Gradient Descent có nhược điểm đó là phụ thuộc vào điểm learning rate.

Nếu learning rate càng nhỏ, tuy mất rất nhiều thời gian, nhưng ta có thể dễ dàng tiếp cận đến điểm cực tiểu dễ dàng.

Nếu learning rate quá lớn, tuy là thời gian rút ngắn nhưng có thể ta sẽ bỏ qua điểm cực tiểu của hàm số.

 Vì thế không có giá trị nào là tối ưu cho learning rate để dễ dàng tiếp cận đến giá trị cực tiểu của hàm số.

5.2 Điểm yên ngựa.

Điểm yên ngựa là điểm giống cực trị, tại đó đạo hàm của hàm số bằng 0 nên nó thỏa mãn điều kiện của dòng lệnh if. Nếu không có giải pháp để bỏ qua điểm yên ngựa, hay nhận biết điểm yên ngựa, có thể lầm lẫn đó là cực trị của hàm số.