KHOA CNTT & TRUYỀN THÔNG
BM KHOA HỌC MÁY TÍNH

ÔN TẬP KIẾN THỨC TOÁN

Nguyên lý máy học

Giáo viên giảng dạy:
TS. TRẦN NGUYỄN MINH THƯ
tnmthu@cit.ctu.edu.vn

1

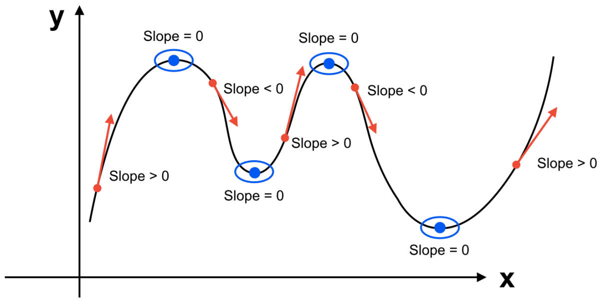
1

NỘI DUNG

- Đạo hàm
- Đạo hàm riêng
- Phương pháp giảm gradient (Gradient descend)

2

Đạo hàm



Đạo hàm - derivative /dɪˈrɪvətɪv/

Đạo hàm của hàm số f(x) khảo sát sự biến thiên của hàm số f(x) theo biến x

Đạo hàm là công cụ giúp ta tìm được xem tại 1 điểm, đồ thị đang đi lên (Tăng) hay đi xuống (giảm) hay đạt cực đại, cực tiểu (không tăng- không giảm).

Đạo hàm

Đạo hàm - derivative /dɪˈrɪvətɪv/

Đạo hàm là công cụ giúp ta tìm được xem tại 1 điểm, đồ thị đang đi lên hay đi xuống

Bản chất của đạo hàm f'(x) là tốc độ gia tăng của hàm f(x) theo sự tăng dần của biến số x ở ngay gần sát tại điểm x đang xét :

- Nếu giá trị của đạo hàm tại x là >0, chứng tỏ rằng tại đó hàm số tăng theo x
- Nếu giá trị của đạo hàm tại x là <0, chứng tỏ rằng tại đó hàm số giảm theo x
- Nếu giá trị của đạo hàm tại x là =0, chứng tỏ rằng tại đó hàm số không tăng cũng không giảm

Khi đạo hàm bằng 0 thì hàm số tại đó không tăng cũng không giảm khi x tăng

Điểm có đạo hàm bằng 0 có thể là cực đại; cực tiểu hoặc điểm uốn.

Cần phải biết dấu của đạo hàm cấp hai mới xác định được đó là CĐ, CT hay điểm uốn.

Đạo hàm

Đạo hàm của f(x) với x là biến số	Đạo hàm của f(u) với u là một hàm số
$(kx)' = k$	$(k \times u)' = k \times u'$
$(x^n)' = nx^{n-1}$	$(u^n)' = n \times u^{n-1} \times (u)'$
$\left(\frac{1}{x}\right)' = -\frac{1}{x^2}$	$\left(\frac{1}{u}\right)' = -\frac{(u)'}{u^2}$
$(\sqrt{x})' = \frac{1}{2\sqrt{x}}$	$(\sqrt{u})' = -\frac{(u)'}{2\sqrt{u}}$
$(\sin x)' = \cos x$	$(\sin u)' = \cos u \times (u)'$
$(\cos x)' = -\sin x$	$(\cos u)' = -\sin u \times (u)'$
$(\tan x)' = 1 + \tan^2 x = \frac{1}{\cos^2 x}$	$(\tan u)' = (1 + \tan^2 u) \times (u)' = \frac{(u)'}{\cos^2 u}$
$(\cot x)' = -(1 + \cot^2 x) = -\frac{1}{\sin^2 x}$	$(\cot u)' = -(1 + \cot^2 u) \times (u)' = -\frac{(u)'}{\sin^2 x}$
$(e^x)' = e^x$	$(e^u)' = e^u \times (u)'$
$(a^x)' = a^x \times \ln a$	$(a^u)' = a^u \times \ln a \times (u)'$
$(\ln x)' = (\ln\|x\|)' = \frac{1}{x}$	$(\ln u)' = (\ln\|u\|)' = \frac{(u)'}{u}$
$(\log_a x)' = (\log_a \|x\|)' = \frac{1}{x \times \ln a}$	$(\log_a u)' = (\log_a \|u\|)' = \frac{(u)}{u \times \ln a}$

5

5

Đạo hàm riêng

Đạo hàm riêng của một hàm số **đa biến** là đạo hàm theo một biến, các biến khác được xem như là hằng số.

Ký hiệu của đạo hàm riêng là

$$\frac{\partial f}{\partial x} \qquad \frac{d}{dx}$$

Ví dụ công thức sau: đạo hàm riêng của hàm số **J(θ)** theo **θ**$_j$

$$\frac{\partial}{\partial \theta_j} J(\theta)$$

https://vi.wikipedia.org/wiki/%C4%90%E1%BA%A1o_h%C3%A0m_ri%C3%AAng

6

6

Đạo hàm riêng: **một số công thức cơ bản**

$$\frac{d}{dx}(\alpha u) = \alpha \frac{du}{dx}$$

$$\frac{d}{dx}\sum u = \sum \frac{du}{dx}$$

$$\frac{d}{dx}u^n = nu^{n-1}\frac{du}{dx}$$

$$\frac{d}{dx}f(g(x)) = f'(g(x))g'(x)$$

7

7

Đạo hàm & Gradient

Gradient descend

"Derivative"- đạo hàm là khái niệm dành cho hàm số với một biến

Gradient /ˈgreɪdɪənt/ là khái niệm tương tự như derivative, nhưng dành cho hàm nhiều biến (multi-variable)

Đạo hàm giúp ta biết hướng xuống / lên dốc trong hàm 1 biến tại một điểm thì gradient giúp ta biết điều này trong hàm nhiều biến (2 trở lên).

Gradient descend là một thuật toán giúp tìm điểm cực tiểu cục bộ (local minimum) của hàm số bằng cách đi dần dần về phía giảm của gradient (gradient descend).

8

8

Giảm Gradient

Gradient descent (giảm gradient) là phương pháp dùng để tìm cực tiểu cục bộ của một hàm. Ý tưởng chính của phương pháp này là từ vị trí hiện tại, thực hiện một bước đi nhỏ, ngược lại với hướng của vector gradient tại vị trí đó.

Cách đơn giản nhất để hiểu gradient descent là từ vị trí hiện tại, ta đi theo chiều giảm của đạo hàm bậc nhất cho đến khi không thể giảm được nữa. Khi đó ta đã ở một điểm tối ưu cục bộ. Công thức cập nhật cho gradient descent là:

$$\theta_{k+1} = \theta_k - \alpha_k \frac{\partial}{\partial \theta} C(\theta_k)$$

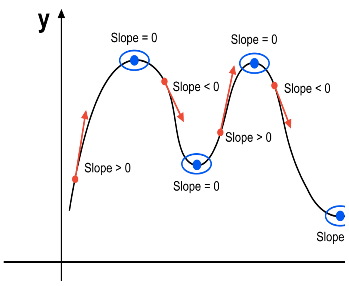
trong đó **α** gọi là learning rate, có thể được cố định hoặc thay đổi thích nghi trong suốt quá trình huấn luyện.

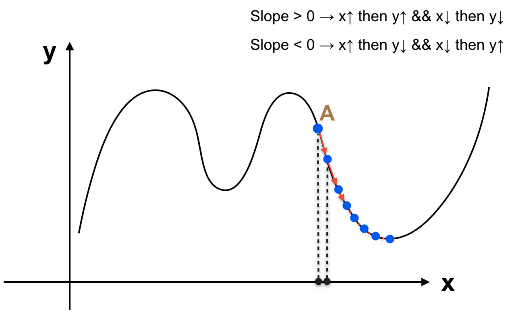
9

9

Giảm Gradient

Khi đó J(θ) giảm xuống (vì slope tại A 0 thì khi tăng θ thì J(θ) sẽ tăng, lúc này muốn giảm J(θ) ta phải giảm θ. Vậy với giá trị θ hiện tại, chúng ta đã biết tăng hay giảm θ để giá trị của J(θ) giảm xuống. Thử tưởng tượng rằng chúng ta làm việc này một cơ số lần, khi đó J(θ) sẽ giảm dần xuống giá trị cực tiểu gần nhất (cực tiểu địa phương – Local minimum).





10

10

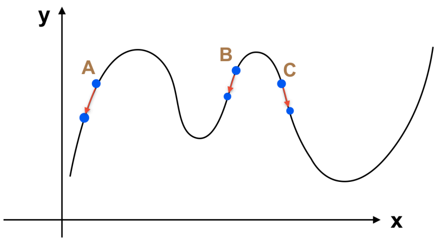
Giảm Gradient

Giải thuật giảm Gradient được mô tả như sau:

+ Khởi tạo θ bất kì (thường tất cả tham số θ được gán = 0).
+ Liên tiếp thay đổi **θ := θ – α * J'(θ)** đồng thời (tất cả các tham số θ phải được thay đổi cùng 1 lúc)
+ Dừng lại khi J(θ) có thay đổi không đáng kể.

Trong đó α được gọi là Learning rate (dùng để điều chỉnh tốc độ học của máy).

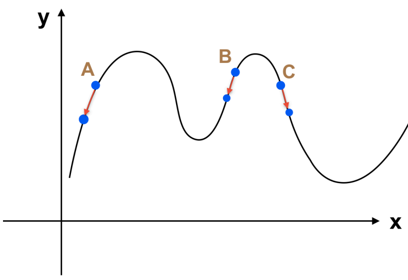
Gradient descent (giảm gradient) là phương pháp dùng để tìm cực tiểu cục bộ của một hàm. Ý tưởng chính của phương pháp này là từ vị trí hiện tại, thực hiện một bước đi nhỏ, ngược lại với hướng của vector gradient tại vị trí đó.

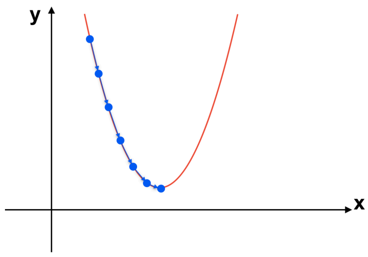


11

Giảm Gradient

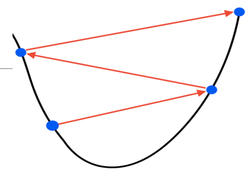
Khi ta đang ở điểm B và điểm C, áp dụng giải thuật trên sẽ đưa ta đến 2 điểm cực tiểu khác nhau. Rõ ràng tại C, ta sẽ giảm được J(θ) nhiều hơn, chính vì thế để đảm bảo J(θ) đạt được giá trị nhỏ nhất, ta phải cố gắng xây dựng J(θ) có dạng tương tự như parabolla. Khi đó J(θ) sẽ chỉ có một giá trị nhỏ nhất, và giải thuật chắc chắn đưa J(θ) xuống đáy của đồ thị.





12

Giảm Gradient – tốc độ học



Tốc độ học - Learning rate α.

Nếu như ta loại bỏ tham số α thì chẳng qua là để α = 1

giải thuật trở thành $\theta := \theta - J'(\theta)$

Việc xác định được các tham số θ nhanh hay chậm phụ thuộc vào tham số này, tuy nhiên nếu như lựa chọn không phù hợp có thể gây ra hệ luỵ nhất định. Cụ thể như sau:

- α quá nhỏ: α nhỏ sẽ khiến cho các bước đi xuống dốc của J(θ) cũng nhỏ => việc học mất nhiều thời gian hơn
- α quá lớn: α lớn sẽ khiến cho các bước đi trở nên quá lớn, có thể khiến J(θ) lớn hơn giá trị hiện tại.

Phân tích hồi quy

Phân tích hồi quy là nghiên cứu sự phụ thuộc của 1 biến (biến phụ thuộc) vào 1 hay nhiều biến khác (biến độc lập), nhằm mục đích ước lượng (hay dự đoán) giá trị trung bình của biến phụ thuộc trên cơ sở các giá trị biết trước của các biến độc lập.

Living area (feet^2)	Price (1000\$s)
2104	400
1600	330
2400	369

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Biến độc lập?

Biến phụ thuộc?

Square meters	Bedrooms	Floors	Age of building (years)	Price in 1000€
x1	*x2*	*x3*	*x4*	*y*
200	5	1	45	460
131	3	2	40	232
142	3	2	30	315

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$$