Steepest Descent in elliptical error surface

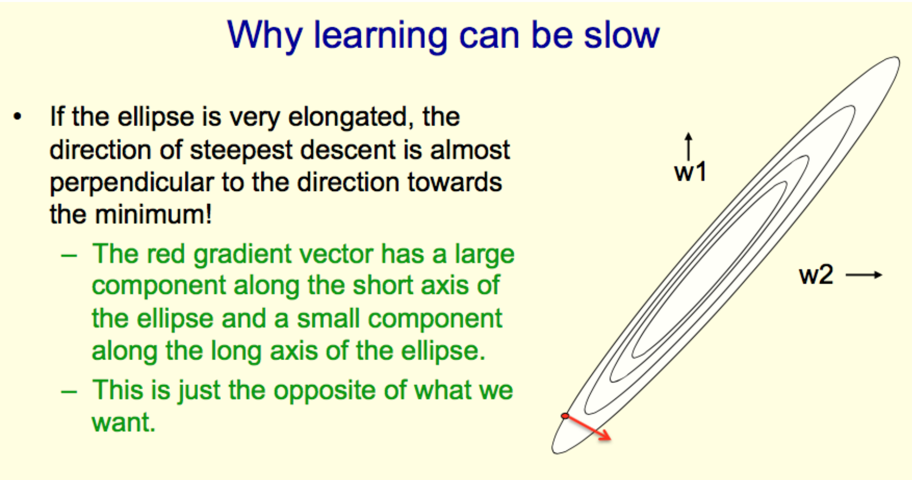
1

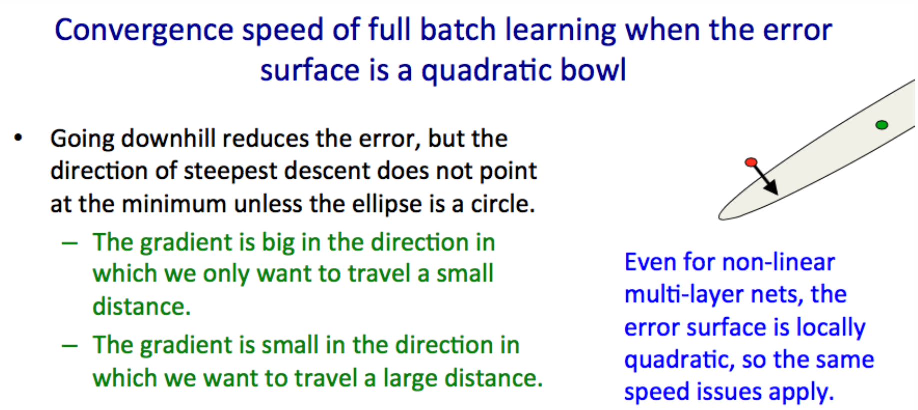
I am watching the Neural Network videos by Prof. Geoff Hinton. In there he talks about the problem with elliptical error surface.

In particular, he says, if the error surface is very elliptical, the direction of steepest descent is ***perpendicular to the direction towards the minimum***.

My understanding of steepest descent was that it is ***perpendicular to the error surface*** and it points towards the direction of minimum.

Slides:





Link to timestamped videos:
https://youtu.be/tIovUOirJkE?list=PLoRl3Ht4JOcdU872GhiYWf6jwrk_SNhz9&t=224
https://youtu.be/4BZBog1Zx6c?list=PLoRl3Ht4JOcdU872GhiYWf6jwrk_SNhz9&t=66

Questions:
1) By definition, shouldn't steepest descent point in the direction of minimum?
2) Can you help me understand why is steepest descent perpendicular to the direction of minimum in case of elliptical error surfaces? But points towards minimum in case of circular error surface.

PS: responding to Amakelov's answer (modifying question, since I can't upload images in comment):

asked 6 months ago
viewed 159 times
active 5 months ago

Linked

12 Gradient is NOT the direction that points to the minimum or maximum

Related

1 Trying to understand Lagrange multipliers

0 Convergence of steepest gradient descent

0 Steepest descent for a function defined over a specific domain

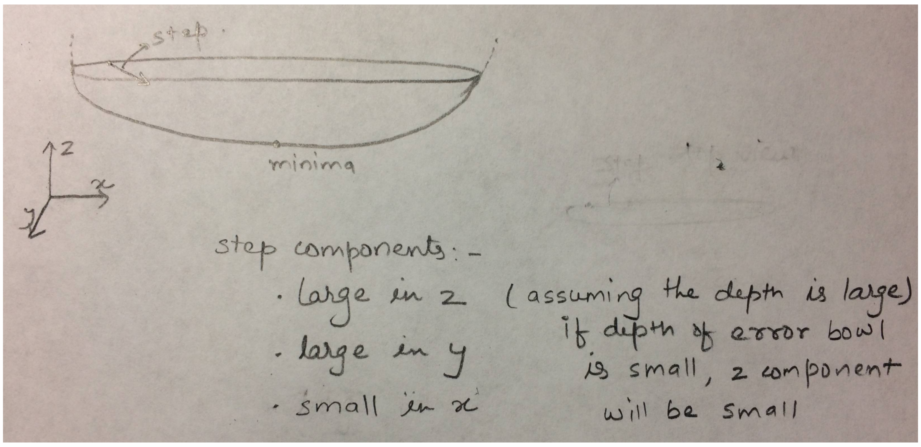
2 Difference between Gradient Descent method and Steepest Descent

1 Prove the steepest descent algorithm for solving $Ax = b$

0 Why is gradient describe the steepest ascent direction and not the steepest descent

2 Gradient Descent: transforming error surface from elliptical to circular

2



Let's assume this is my elongated elliptical error surface bowl. In this for steepest descent, I will have a large `y` component, small `x` component and variable `z` component (depending how low in the error bowl I am).

So, if I look at the top down view, I will see a big movement in the `y` axis and a small movement in the `x` axis but the movement will still be in the direction of minima (and not perpendicular to it). What am I missing here?

(optimization) (convex-optimization) (neural-networks) (gradient-descent)

share cite improve this question

edited Apr 29 at 0:03

asked Apr 28 at 23:23

The Wanderer
204 1 13

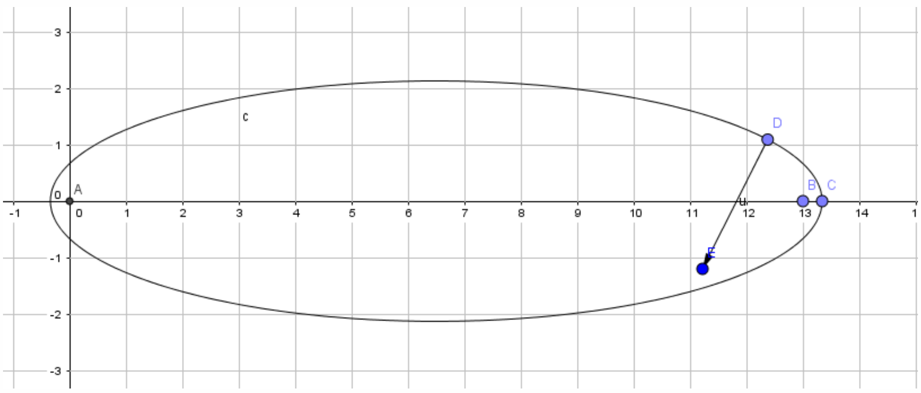
add a comment

2 Answers

active oldest **votes**

2

Here is a figure that may help. I draw a long ellipse, which is intended to be one contour line, and a gradient vector that I drew by eye. The gradient vector is always perpendicular to the contour lines. It should be clear that the vector does not point toward the center of the ellipse. It is also not really close to perpendicular to the direction toward the center of the ellipse. I think the slides you cite exaggerate the perpendicularity. It does get worse if the ellipse is even longer than I have drawn.



Added: To my eye, the gradient is closer to 45° from the direction to the minimum than perpendicular. I think this is correct. If you minimize along the direction of the gradient you will cross the axis of the ellipse and stop at a point where your direction of travel is along the contour line. You stop and turn a right angle as I pointed out in my answer to your other question. Now you are again following the gradient and again about 45° from the direction to the minimum. You will zigzag across the axis. The good news is your total travel is only $\sqrt{2}$ times as far as the direct route. The bad news is that each turn requires evaluating the gradient, which can be expensive. The point I picked is about the worst case. If you start closer to the end of the ellipse the gradient will be closer to the direction to the center. If you start farther from the end, the first step will be farther from the direction to the center but the next will be better.

share cite improve this answer

edited May 6 at 19:24

answered May 6 at 0:25

Ross Millikan
243k ● 20 ■ 162 ▲ 302

Thank you Ross. This is what I was struggling to understand, ie. why is the learning slow for elliptical curves, which led to me asking all those other questions. – The Wanderer May 6 at 0:51

add a comment

2

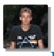
1) No, and you can easily convince yourself by a mental experiment. Imagine you're literally standing on this elliptical ravine at the point on the figure. Remember that the ellipses denote locations with the same altitude. Then, when you're near the narrow end of the ravine (**but not exactly at the narrow end!**), the fastest way to descrease your altitude is not to walk towards the minimum (which would involve a lot of sideways motion that doesn't contribute to decreasing your altitude) but to walk directly down towards the long axis of the ellipse.

Formally, what we mean by 'steepest descent' in math is that, locally, the direction opposite to the gradient $\nabla f(x)$ provides the most decrease in altitude for a given horizontal distance δ you are willing to travel away from x, *in the limit* $\delta \to 0$.

2) The steepest descent direction is not literally perpendicular to the direction of the minimum. Rather, in this extreme case it is *almost* perpendicular, and the reason is as above. Another way to look at it is to observe that the gradient is orthogonal to the level sets of your function, and since the level sets of the elliptical ravine near its narrow ends (**but not exactly at the narrow ends!**) are closely aligned with the direction toward the minimum, this leads to your gradient being almost orthogonal to the direction to the minimum.

share cite improve this answer

answered Apr 28 at 23:40



amakelov
2,337 ■ 3 ▲ 15

Hey Amakelov, thank you for your detailed response. But it was still not clear to me. So, I drew a small example and updated my question. Can you please help me understanding based on what I drew? I would greatly appreciate – The Wanderer Apr 29 at 0:05

add a comment

Your Answer

Sign up or log in

Sign up using Google

Sign up using Facebook

Sign up using Email and Password

Post as a guest

Name

Email

required, but never shown

Post Your Answer

By posting your answer, you agree to the privacy policy and terms of service.

Not the answer you're looking for? Browse other questions tagged (optimization) (convex-optimization) (neural-networks) (gradient-descent) or ask your own question.

question feed