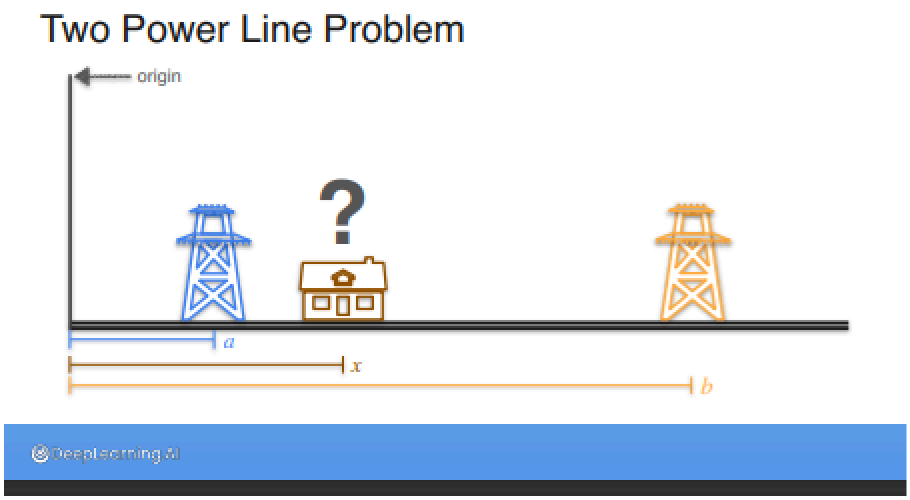
OPTIMIZATION

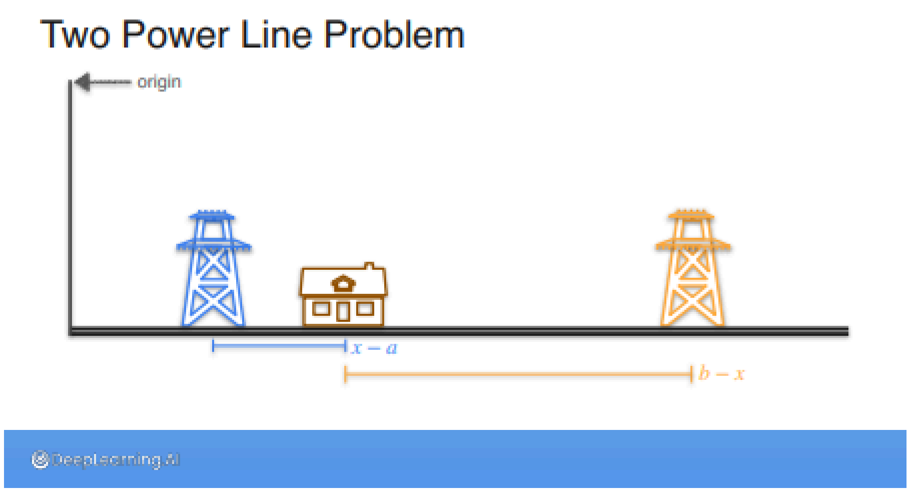
As I promised. Today we're going to talk about Calculus usage in optimization. But what do I mean by optimization? Imagine living in a strange city where you can drive your car strictly with required speed, but you're new to this city. How do you obey this rule? Well you can choose random speed and see the slope (speed and bill can be drawn as a graph). You will be changing your speed until you reach the *local minimum.* This local minimum is that speed needed. It's that gap in the graph. In other words: It's the point with 0 slope.

Powerline problem.
Idk if this problem is famous or not, but it represents optimization in the best way. I'll show only optimization for 2 powerlines, but DeepLearningAI calculus course offers you more examples at the end of week 1.

Imagine building a house. You must connect your house to 2 powerlines to provide electricity to your house. Company that offers you electricity have an interesting bill formula. Bill equals distance to your house^2

What is the best (cheapest) positioning for a house?





Well our total cost can be represented as:

$$(x - a)^2 + (b - x)^2$$

To find the best positioning x we need to find point where derivative = 0

Let's do it:

$$d/dx((x - a)^2 + (x - b)^2) = 0$$

Let's find derivatives.

We know that derivative of $f(x) = x^2$ is $f'(x) = 2x$

(or $f(x) = x^n$ is $f'(x) = nx^{n-1}$

So it's gonna be:

$$2(x - a) + 2(x - b) = 0$$

Solve this easy equation:

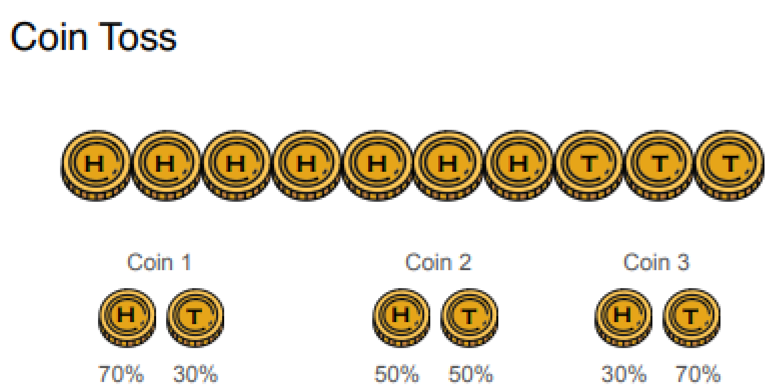
$$2x - 2a + 2x - 2b = 0$$
$$4x = 2b + 2a$$
$$2x = b + a$$
$$x = (b + a) / 2$$

It means that we have to place our house in the very middle (between those 2 powerlines)

Logarithms cost functions optimization.

Again. I'll give example from DeepLearningAI course (as imho there's no better explanation):

Imagine:
You are playing a game with coins. If your 10 coin flips will look like this:



You win.

And you are given 3 coins.
Which one will you use to maximize your winning chance?
Most likely you've chosen the 1st one (and it's right) but why?

We can represent our chances as:

Chances of winning: $p^7(1-p)^3 = g(p)$

Our goal is to maximize g(p)

There are 2 ways to do it:
1st way: hard (using derivative properties)

$$\frac{dg}{dp} = \frac{d}{dp}(p^7(1-p)^3) \overset{\text{Product rule}}{=} \frac{d(p^7)}{dp}(1-p)^3 + p^7\boxed{\frac{d((1-p)^3)}{dp}}$$

$$= 7p^6(1-p)^3 + p^7\boxed{3(1-p)^2(-1)} \quad \text{Chain rule}$$

$$= p^6(1-p)^2[7(1-p) - 3p]$$

$$= \boxed{p^6}\boxed{(1-p)^2}\boxed{(7-10p)} = 0$$

$\cancel{p = 0}$ $\quad p = 1 \quad p = 0.7$

As you can see we've set derivative to 0, but why didn't we take p=0 or p=1:
Because it doesn't satisfy our goal. p = 0 means we have no chances to get H, while p = 1 means we have 100% chance to get H, but as a result 0% chance to get T.

But as I mentioned it's pretty hard as it required us to make some advanced calculations.
So there's other way:
2nd way: easy (use logarithms)

$$\log(g(p)) = \log(p^7(1-p)^3) = \log\left(p^7\right) + \log\left((1-p)^3\right)$$
$$= 7\log(p) + 3\log(1-p) = G(p)$$
$$\frac{dG(p)}{dp} = \frac{d}{dp}(7\log(p) + 3\log(1-p)) = 7\frac{1}{p} + 3\frac{1}{1-p}(-1)$$
$$= \frac{7(1-p) - 3p}{p(1-p)} = 0$$
$$7(1-p) - 3p = 0 \quad p = 0.7$$

We have used logarithm properties, but we achieved our result!

So interesting point here is to keep this logarithm optimization in mind
As it can be used when we have some advanced calculations.

So in the end of the 1st week we've seen some practical applications of calculus in Machine Learning.

____

Uses materials by DeepLearningAI
DeepLearningAI course - link
Written by Venchislav for the GitHub community❤.
21.03.2024

____
GoodBye!