Chain Rule

quantra

Table of Contents

This unit on "Derivatives and chain rule" is a prerequisite to the next unit on "Math behind the back propagation". To get a theoretical understanding of how backpropagation works in a neural network, you first need to understand the derivations and the chain rule.
To be able to understand this unit, you should know what a derivative is. Don't sweat, in case you don't know or don't remember the same, you can learn about it on the glossary section of Quantra website.

Chain rule

The chain rule is basically a formula for computing the derivative of the composition of two or more functions. Let us say that, if f and g are functions, then the chain rule expresses the derivative of their composition f ∘ g (the function which maps x to f(g(x))). The derivative of this composition is calculated as mentioned below.

$$(f \circ g)' = (f' \circ g) \cdot g'.$$

Another way of writing the above rule:
Let F = f ∘ g, or equivalently,
F(x) = f(g(x)) for all x. Then one can also write

$$F'(x) = f'(g(x))g'(x).$$

Yet, another way of writing the above rule!
The chain rule may be written in Leibniz's notation in the following way. If a variable z depends on the variable y, which itself depends on the variable x, so that y and z are therefore dependent variables, then z, via the intermediate variable of y, depends on x as well.

Simply put, the chain rule then states:

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}.$$

This, final formula, is the one that we will be using in backpropagation.

Let us say that z is function of y, z = f(y), and y is a function of x, y= g(x), then

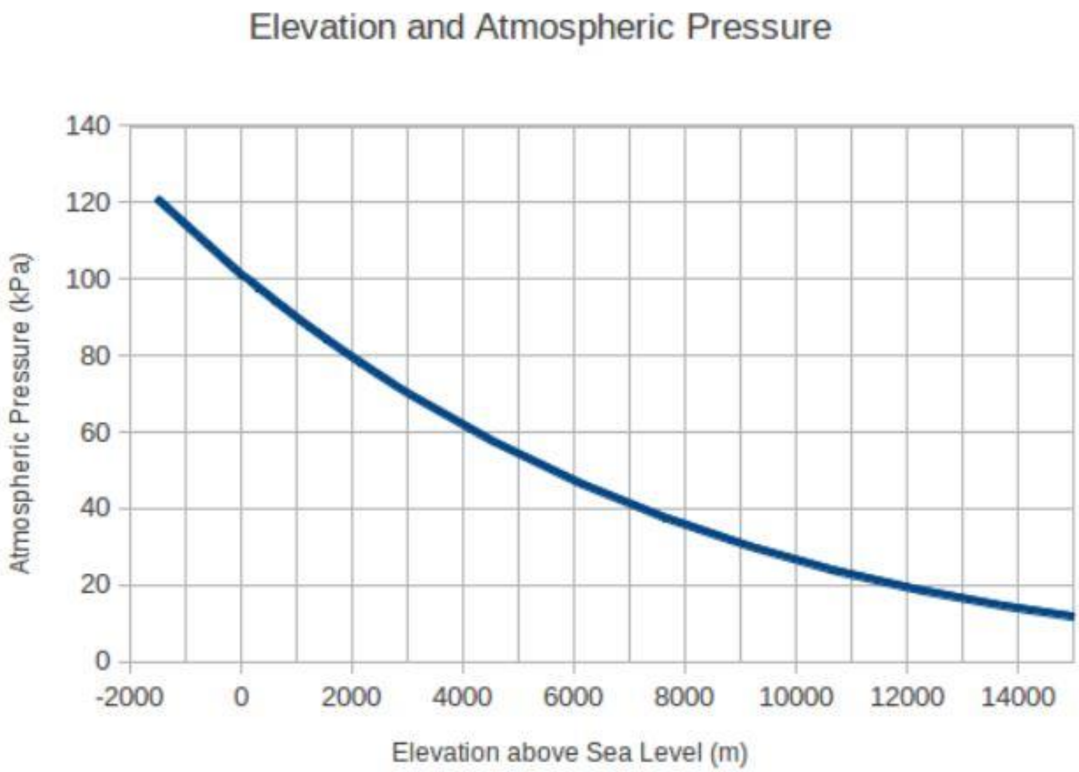
$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx} = f'(y)g'(x) = f'(g(x))g'(x).$$

Let us understand this better with the help of an example.

Example

When you fall from the sky, the atmospheric pressure keeps changing during the fall.

Checkout the graph below to understand this change.



Chain rule of derivatives is used to be able to understand the rate of change in atmospheric pressure. Let us assume that a physicist fell out of. At the time of his fall, from

4000 metres above sea level, his initial velocity was zero. As we know, the gravity is 9.8 meters per second-squared. Let us see how the chain rule helps us here.

In this example:

If we assume that the variable x is time 't', then the variable y or g(t), which is the distance travelled by the physicist since the beginning of his fall is given by $0.5*9.8t^2$, so his height from the mean sea level is given by the variable z or f(y), which is 4000 - g(t). Also, based on a model, the atmospheric pressure at a height h is given by $f(h) = 101325\ e^{-0.0001h}$.

These two equations can be differentiated and combined in various ways to produce the following data:

$g'(t) = -9.8t$ is the velocity of the skydiver at time t

$f'(h) = -10.1325^{e-0.0001h}$ is the rate of change in atmospheric pressure with respect to height h

Now let us understand what we can get by combining these two equations:

$(f \circ g)(t)$ is the atmospheric pressure the skydiver experiences t seconds after his jump. Similarly the differential of this combined function, $(f \circ g)'(t)$, is the rate of change in the atmospheric pressure with respect to time at t seconds after the skydiver's jump.

Using the chain rule we can compute the rate of change in the atmospheric pressure with respect to time at t in terms of f' and g'.

The chain rule states that,

$$(f \circ g)'(t) = f'(g(t)) \cdot g'(t).$$

In this example this will lead to:

$$(f \circ g)'(t) = \left(-10.1325e^{-0.0001(4000-4.9t^2)}\right) \cdot \left(-9.8t\right).$$

We will see an application of this chain rule in the back propagation of the Neural Networks in the next unit.