**Content**

[1. Gradient Descent 2](#_Toc16086984)

1. Termiology
   1. Gradient Descent

Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or approximate gradient) of the function at the current point. If, instead, one takes steps proportional to the positive of the gradient, one approaches a local maximum of that function; the procedure is then known as gradient ascent.

Chain rule – derivative of nested functions

Calculate the back propagation for a fully connected neural network.

* + 1. What is the Gradient?

Gradient shows us the maximum of the function.

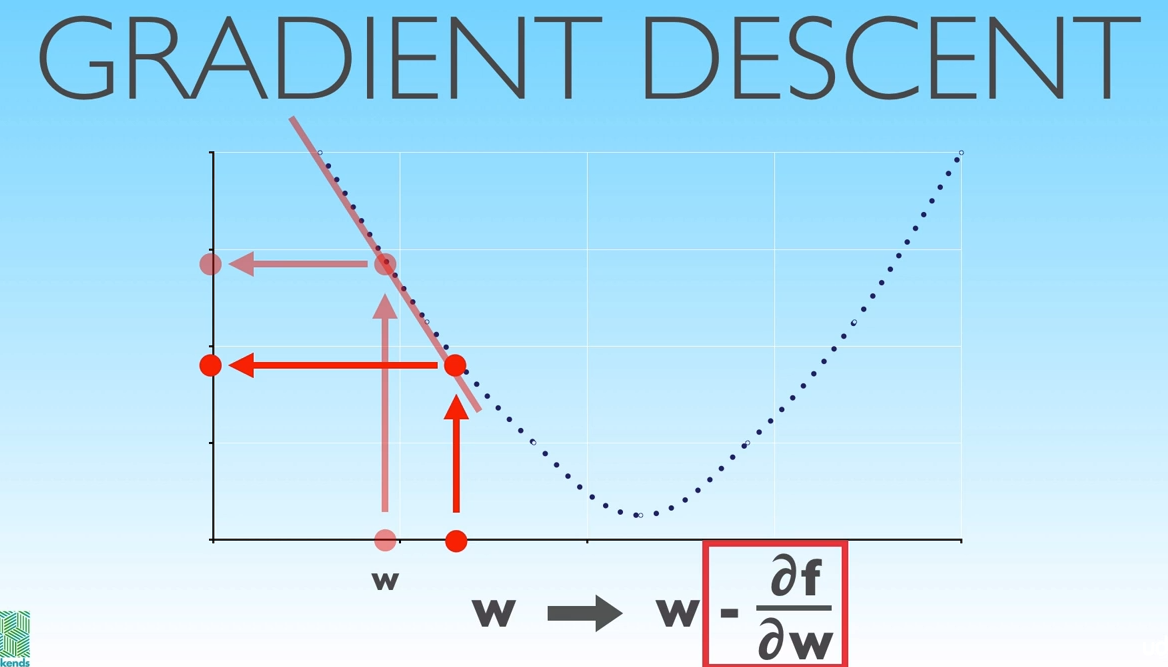
Derivatives: slope, rate of change

Gradient: extends derivative to multivariate functions.

* + 1. Back Propagation

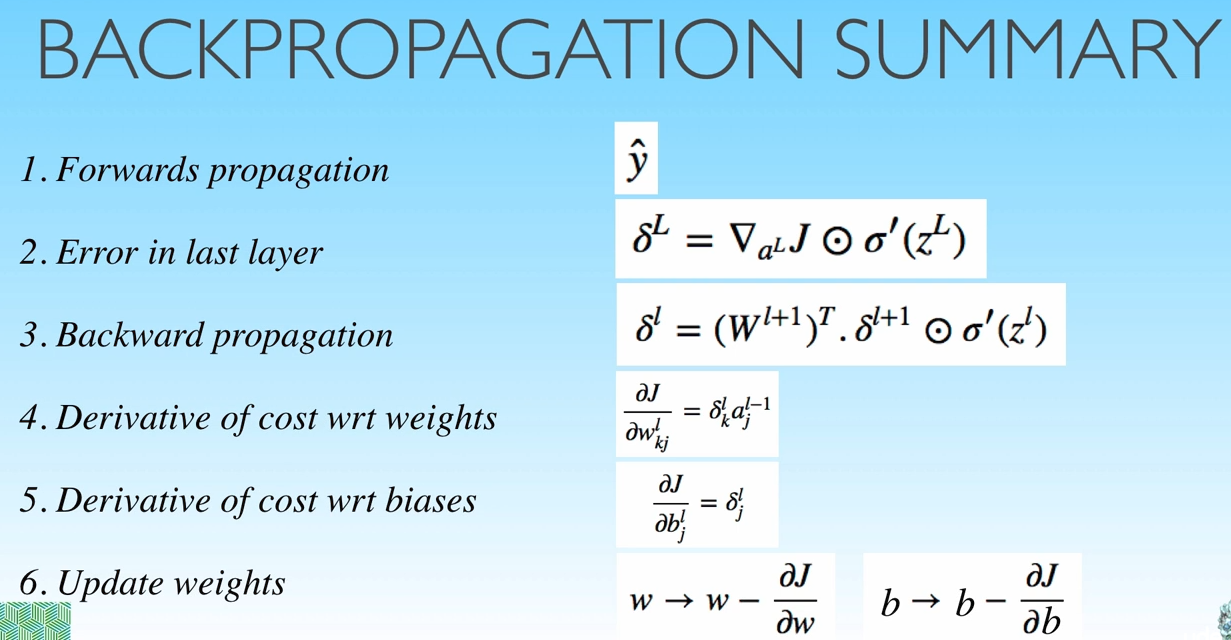
Csokkeno fuggveny - derivacio negative

Novekvo fuggveny – derivalt pozitiv



For a neural network we define cost function, that depends on the values of the parameters. We find the values of the parameters by minimalizing the cost by gradient descent. All we are really doing is taking the cost function, calculating its partial derivatives with respect to each parameter and then using the update rule (x := x – df/dx) we decrease the cost by updating the paramters we do this by **subtracting the value of the negative gradient from each of the parameters.**

* + 1. The Complete Mathematical Notation of a Fully Connected Neural Network



1. Technologies