Derivation of DL/dz

<https://www.coursera.org/learn/neural-networks-deep-learning/discussions/weeks/2/threads/ysF-gYfISSGBfoGHyLkhYg>

This is optional material that you can read after the week 2 video "Gradient descent on m examples." You don't need to know calculus in order to complete this course (or the other courses in the specialization), so this derivation is optional. This is for those who are curious about where the "dz = a - y" comes from.

This can be more fun and easier to digest if you follow along with a pencil and paper!

Text, letter

Description automatically generated

We're taking the derivative with respect to a.

Remember that there is an additional −1 in the last term when we take the derivative of (1-a) with respect to a (remember the Chain Rule). Also note that the notational conventions are different in the ML world than the math world: here log always means the natural log.

A picture containing table

Description automatically generated

You can look up why this derivation is of this form. For example, google "derivative of a sigmoid", and you can see the derivation in detail.

Recall that *σ*(*z*)=*a*, because we defined "a", the activation, as the output of the sigmoid activation function.

So we can substitute into the formula to get:



A picture containing logo

Description automatically generated

We'll multiply step 1 and step 2 to get the result.

Text

Description automatically generated