

Cost Function

Imagine we wanna know how well our neural network works. That's the purpose of this function.

$$C(w, b) = \frac{1}{n} \sum_x \| y(x) - a \|$$
$$a = \sigma(z) = \frac{1}{1 + e^{-z}}$$

The cost function formula is intuitive, innit? Let's define the terms of the formula:

- $y(x)$ is the ideal value,
- a is the output of our neural network (our prediction)
- n is the quantity of predictions we have done (the array's size)

We are just subtracting these two values to find an error $\|y(x) - a\|$, and then dividing by the number of terms " n " (finding an average).

Why should we squaring the error?

Because our objective is to find an average, if we don't square the error some terms could be negative and our average would not be representative.

$$C(w, b) = \frac{1}{2n} \sum_x \| y(x) - a \|^2$$

Why are we multiplying by 1/2 the cost function?

Because it will make some calculus easier later.

The output of C is a measure that defines how well our network works and it's calculated using the weights and biases from the last layer, but we will figure out later that it affects the whole net. The same set of parameters w and b is shared by all the predictions (a), during the training these parameters are globally updated to minimize the average error on the set of data.