1 Gradient Descent

As the ϵ decrease in Finite Difference approach result is derivative of cost function.

$$C'(w) = \lim_{\epsilon \to 0} \frac{C(w+\epsilon) - C(w)}{\epsilon} \tag{1}$$

1.1 Cost function

Cost function is used to calculate the output of the network and using the output we can calculate how accurate the function is?
Cost function for a single layer can be written as:

$$C = \sum_{i=1}^{n} a_i w_i + b \tag{2}$$

So for a layer cost function becomes:

$$C^l = \sum_{i=1}^{n} a_i^l w_i^{l-1} + b^l \tag{3}$$

1.2 Linier Model