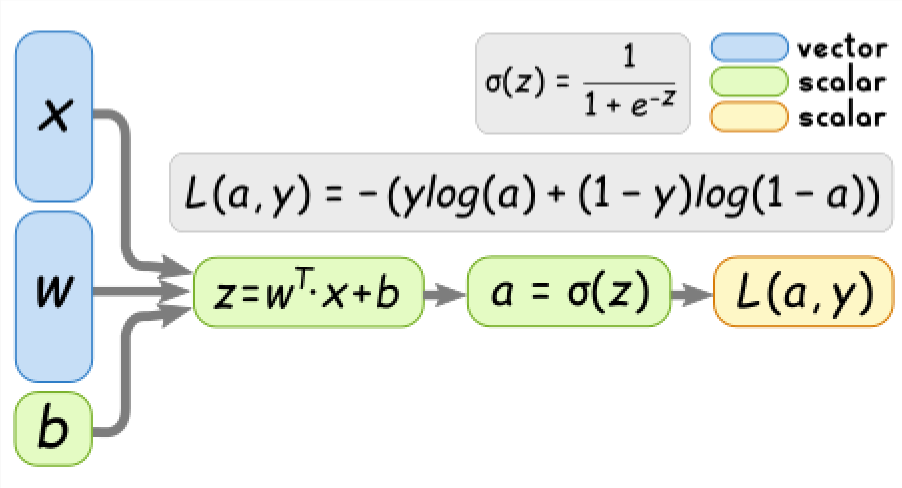
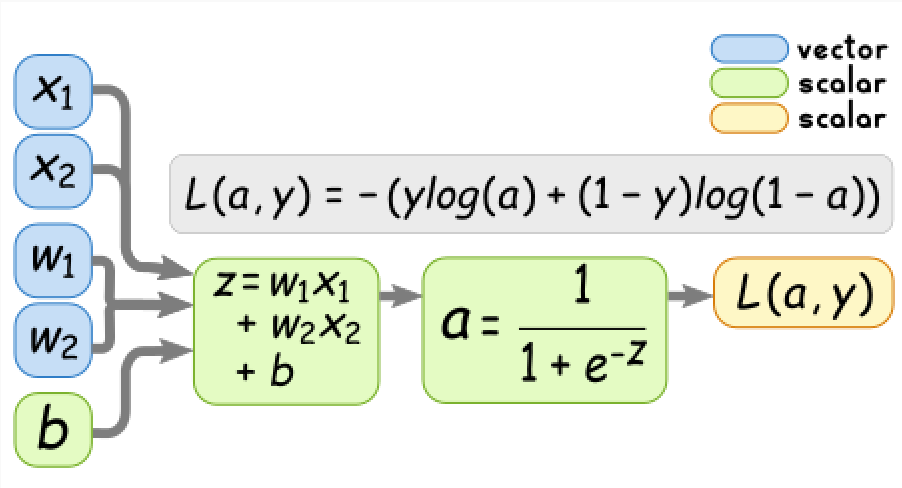
From http://ronny.rest/blog/post_2017_08_12_logistic_regression_derivative/

Logistic Regression computational graph

The computational graph of logistic regression can be visualised as follows:



`w` and `x` are vectors, whose size depend on the number of input features. In order to keep things simple, we will consider the case where we only have two input features. We can therefore represent the computational graph more clearly as follows:



Desired partial derivatives

The partial derivatives we are particularly interested are the following two:

$$\frac{dL(a,y)}{dw_i}$$ Partial derivatives with respect to the weights

$$\frac{dL(a,y)}{db}$$ Partial derivatives with respect to the bias

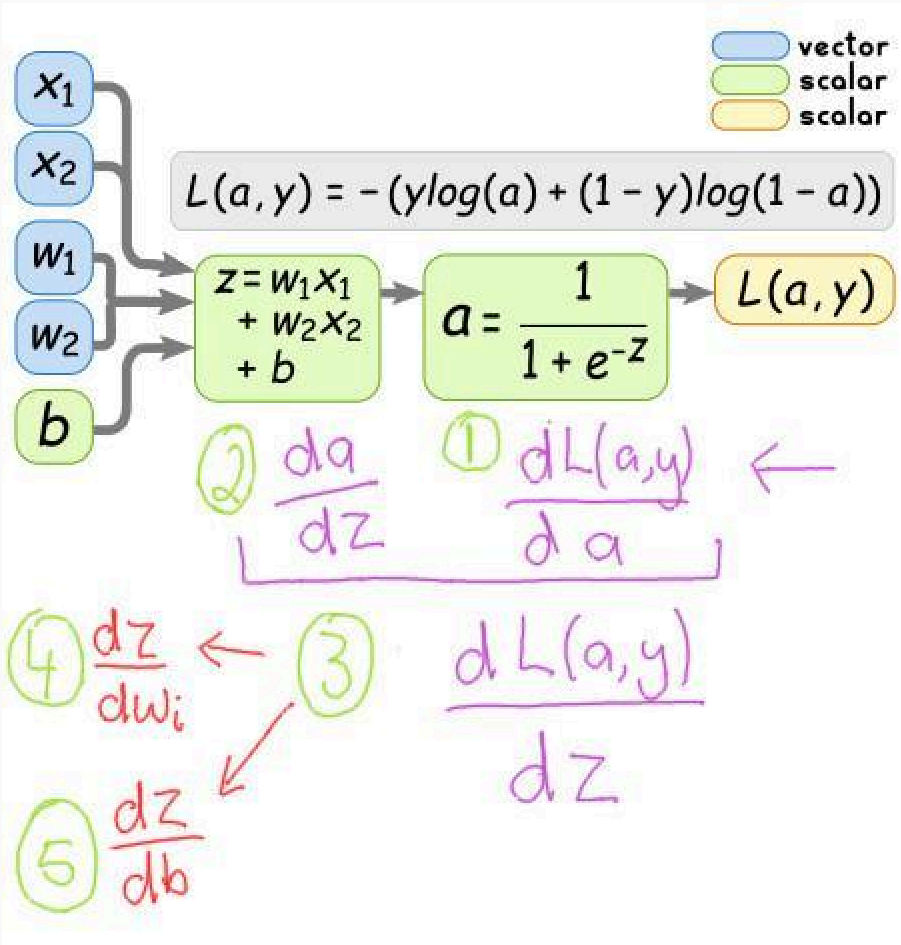
Strategy for solving partial derivatives

In order to solve the partial derivatives, we can make use of the chain rule, which allows us to simplify the process by considering small components at a time.

$$\frac{\delta L(a,y)}{\delta w_i} = \frac{\delta L(a,y)}{\delta a} \times \frac{\delta a}{\delta z} \times \frac{\delta z}{\delta w_1}$$

$$\frac{\delta L(a,y)}{\delta b} = \underbrace{\frac{\delta L(a,y)}{\delta a} \times \frac{\delta a}{\delta z}}_{\frac{\delta L(a,y)}{\delta z}} \times \frac{\delta z}{\delta b}$$

Which can be visualised on the computational graph as follows:



Once we calculate those five smaller components, we can solve the partial derivatives we want more easily. So lets start solving each compoment.

Compoment 1

Be aware that the logs used in the loss function are natural logs, and not base 10 logs.

1. $\dfrac{dL(a,y)}{da}$

$$= \frac{d}{da} - \left[y\log(a) + (1-y)\log(1-a) \right]$$

$$= -\left[\frac{d}{da} y\log(a) + \frac{d}{da}(1-y)\log(1-a) \right]$$

$\boxed{\text{derivative of log rule} \quad \frac{d}{dx} k\log(x) = \frac{k}{x}}$

$$= -\left[\frac{y}{a} + \underbrace{\frac{d}{da}(1-y)\log(\underbrace{1-a}_{g(a)})}_{f(g)} \right]$$

chain rule: $\dfrac{df(a)}{da} = \dfrac{df(g)}{dg} \cdot \dfrac{dg(a)}{da}$

$$\frac{d}{dg}(1-y)\log(g) \cdot \frac{d}{da}(1-a)$$

$$\frac{1-y}{g} \cdot (0-1)$$

$$\underbrace{\frac{1-y}{g} \cdot -1}$$

$$\frac{-(1-y)}{g} \leftarrow \boxed{g = (1-a)}$$

$$-\left[\frac{y}{a} - \frac{(1-y)}{(1-a)} \right]$$

$$= -\frac{y}{a} + \frac{(1-y)}{(1-a)}$$

$$\frac{dL(a,y)}{da} = \frac{1-y}{1-a} - \frac{y}{a}$$

Compoment 2

The full calculation of this component was explained in my previous blog post for calculating the derivative of the sigmoid function. Be sure to check out that post if you want to know how it was calculated.

$$② \quad \frac{da}{dz} = \frac{d}{dz} \frac{1}{1+e^{-z}}$$

$$\boxed{\frac{da}{dz} = a(1-a)}$$

Compoment 3

$$③ \quad \frac{dL(a,y)}{dz} = \frac{dL(a,y)}{da} \cdot \frac{da}{dz}$$

$$= \left(\frac{(1-y)}{(1-a)} - \frac{y}{a} \right) \cdot a(1-a)$$

$$= \frac{a\cancel{(1-a)}(1-y)}{\cancel{(1-a)}} - \frac{y\cancel{a}(1-a)}{\cancel{a}}$$

$$= a(1-y) - y(1-a)$$

$$= a - \underline{ay} - y + \underline{ay}$$

$$= a - y$$

$$\boxed{\frac{dL(a,y)}{dz} = a - y}$$

Compoment 4

4 $$\frac{dz}{dw_i} = \frac{d}{dw_i} w_1x_1 + w_2x_2 + b$$

let us say $i = 1$

$$= \frac{d}{dw_1} w_1x_1 + \frac{d}{dw_1} w_2x_2 + \frac{d}{dw_1} b$$

$$= x_1 + 0 + 0$$

$$\frac{dz}{dw_1} = x_1$$

more generally:

$$\boxed{\frac{dz}{dw_i} = x_i}$$

Compoment 5

5 $$\frac{dz}{db} = \frac{d}{db} w_1x_1 + w_2x_2 + b$$

$$= \frac{d}{db} w_1x_1 + \frac{d}{db} w_2x_2 + \frac{d}{db} b$$

$$= 0 + 0 + 1$$

$$\boxed{\frac{dz}{db} = 1}$$

Putting the components together

$$\frac{dL(a,y)}{dw_i} = \underset{3}{\frac{dL(a,y)}{dz}} \cdot \underset{4}{\frac{dz}{dw_i}}$$

$$\boxed{\frac{dL(a,y)}{dw_i} = (a-y)\,x_i}$$

$$\frac{dL(a,y)}{db} = \underset{3}{\frac{dL(a,y)}{dz}} \cdot \underset{4}{\frac{dz}{db}}$$

$$= (a-y) \cdot 1$$

$$\boxed{\frac{dL(a,y)}{db} = a-y}$$

The final product

To summarize, the derivatives we are interested have been summarized in the following two formulas.

$$\boxed{\frac{dL(a,y)}{dw_i} = (a-y)\,x_i}$$

$$\boxed{\frac{dL(a,y)}{db} = a-y}$$