

2-layer Mean Square error

→

$$z_1 = w_1 \cdot x_1 + b_1$$

$$a_1 = g(z_1) \quad \text{where } g \text{ is the activation function}$$

$g$  can be sigmoid or tanh or RELU...

$$z_2 = w_2 x_2 + b_2$$

$a_2 = z_2 \rightarrow$  This is the difference of Classification and Regression.

Regression uses the identity function  $\phi$  for the output layer, and just needs one neuron as there are no classes present whereas binary classification requires sigmoid function ~~for~~ for the output layer.

$$dz_2 = a_2 - y$$

$$dw_2 = dz_2 \cdot a_1$$

$$db_2 = dz_2$$

$$dz_1 = w_2 \cdot dz_2 \cdot g'(z_1)$$

$$dw_1 = dz_1 \cdot x_1$$

$$db_1 = dz_1$$

$$w_1 = w_1 - \alpha dw_1$$

$$w_2 = w_2 - \alpha dw_2$$

$$b_1 = b_1 - \alpha db_1$$

$$b_2 = b_2 - \alpha db_2$$

④