x = input Features
y = labels

First layer w_1, b_1
second layer w_2, b_2

output of layer 1
$z_1 = w_1 \cdot x + b_1$
$a_1 = g(z_1)$ g is activation function

Layer 2:

$z_2 = w_1 \cdot a_1 + b_2$
$a_2 = g(z_2)$

the o/p = $\hat{y} = a_2$

task: regression
loss = mse mean squared error

$$L = \frac{1}{n} \Sigma (y - \hat{y})^2$$

1. start initil guess for w_1, w_2, b_1, b_2
2. $w_i = w_i - \alpha \cdot \frac{dL}{dw_i}$ $b_i = b_i - \alpha \cdot \frac{dL}{dL_i}$
3. reapeat

CS CamScanner

②

update for second layer $\frac{dL}{dw_2}$

$$\frac{dL}{dw_2} = \frac{d}{dw_2}\left[\frac{1}{J}\sum_j (y-\hat{y})^2\right]$$

$$= 1 \cdot \frac{d}{dw_2}(y-\hat{y})^2$$

$$= 1 \cdot 2\cdot(y-\hat{y})^{2-1} \cdot \frac{d}{dw_2}\left[y-\hat{y}\right]$$

$$= -2\cdot(y-\hat{y})$$

$\hat{y} = a_2 = g(z_2)$

$g =$ nothing

$$\boxed{\therefore \frac{\partial L}{\partial w_2} = -2\cdot(y-a_2)\cdot a_1^T}$$

$$\boxed{\therefore \frac{dL}{db_2} = -2\cdot(y-a_2)}$$

③

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial a_2} \cdot \frac{\partial a_2}{?} \cdot \frac{?}{\partial w_1}$$

$$\therefore \frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial a_2} \cdot \frac{\partial a_2}{\partial z_2} \cdot \frac{dz_2}{da_1} \cdot \frac{da_1}{dz_1} \cdot \frac{dz_1}{dw_1}$$

● rewrite using chain rule

$$\frac{\partial L}{\partial w_1} = -2(y-a_2)\, \frac{dz_2}{da_1} \cdot \frac{da_1}{dz_1} \cdot \frac{dz_1}{dw_1} \quad \left(\frac{dz_2}{da_1} = w_2\right.$$

$a_1 = g(z_1)$ then $\frac{da_1}{dz_1} = g'(z_1)$

$$\therefore \frac{\partial L}{\partial w_1} = -2(y-a_2) \cdot w_2 \cdot g'(z_1) \cdot x$$

$g =$ Sigmod $\frac{1}{1+e^{-x}}$

$g' = g(z)(1-g(z))$

$$\boxed{\therefore \frac{\partial L}{\partial w_1} = w_2^T \cdot g'(z_1) \cdot (-2(y-a_2)) \cdot x^T}$$

$$\boxed{\therefore \frac{\partial L}{\partial b_1} = w_2^T \cdot g'(z_1) \cdot (-2(y-a_2))}$$

4

question for question 1:

the difference bewteen regression for mean squared error loss + binary classification log loss, is the loss function it's self,

instead of having logloss in the update rule you now have mean squared error

CS CamScanner