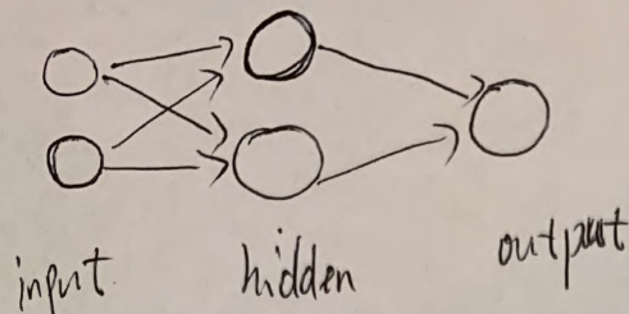

Back prop

$$W^{[l]} = W^{[l]} - \alpha \frac{\partial \mathcal{L}}{\partial W^{[l]}}$$

$$b^{[l]} = b^{[l]} - \alpha \frac{\partial L}{\partial b^{[l]}}$$

$$\frac{\partial L}{\partial W^{[l]}}$$

Let X : input Features $[n \times f]$

y : output labels $[n \times 1]$

Define the weights [parameters]

W_1, b_1 (bias) First Layer

W_2, b_2 (bias) Second Layer.

The output of layer 1:

$z_1 = W_1 X + b_1$

$a_1 = g(z_1)$, where "g" is the activation function.

here, sigmoid $s(x) = \frac{1}{1+e^{-x}}$

The output of layer 2:

$z_2 = W_2 \cdot a_1 + b_2$

$a_2 = (z_2)$ ⟶ Identity activation function.

Final prediction

$\hat{y} = a_2$

Task: Regression

Loss Function: MSE.

$$L = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

① values of weights and bias

W_1, W_2, b_1, b_2

② update the weights using gradient descent

$$W_i = W_i - \alpha \frac{\partial L}{\partial W_i}$$

$$b_i = b_i - \alpha \frac{\partial L}{\partial b_i}$$

α : Learning rate.

i, ith layer.

③ repeat until convergence

$$\frac{\partial L}{\partial W_2} = \frac{\partial [\frac{1}{N}(y-\hat{y})^2]}{\partial W_2} = \frac{(y-\hat{y})\,\partial(\frac{2}{N}(y-\hat{y}))}{\partial W_2} = \frac{-\partial \hat{y}}{\partial W_2}(y-\hat{y})$$

$$\hat{y} = z_2 = a_2$$

$$\frac{\partial \hat{y}}{\partial W_2} = \frac{\partial a_2}{\partial W_2} = \frac{\partial z_2}{\partial W_2} = \frac{\partial (W_2 \cdot a_1 + b_2)}{\partial W_2} = a_1$$

$$\frac{\partial L}{\partial W_2} = -a_1(y-\hat{y}) = -a_1(y - a_2) = (a_2 - y)a_1$$

similarly, $$\frac{\partial L}{\partial b_2} = \frac{\partial (\frac{1}{N}(y-\hat{y})^2)}{\partial b_2} = \frac{(y-\hat{y})\,\partial(y-\hat{y})}{\partial b_2} = \frac{-\partial \hat{y}}{\partial b_2}(y-\hat{y})$$

$$= -\frac{\partial (W_2 \cdot a_1 + b_2)}{\partial b_2}(y-\hat{y}) = -(y-\hat{y}) = -(y-a_2) = a_2 - y$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial a_2} \cdot \frac{\partial a_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial a_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1}$$ (chain rule)

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial a_2} \cdot \frac{\partial z_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial a_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1}$$

$$= \frac{\partial \frac{1}{N}(y-\hat{y})^2}{\partial a_2} \cdot 1 \cdot \frac{\partial (w_2 \cdot a_1 + b_2)}{\partial a_1} \cdot \frac{\partial g(z_1)}{\partial z_1} \cdot \frac{\partial w_1 x + b_1}{\partial w_1}$$

$$g(x) = \frac{1}{1+e^{-x}}, \quad g'(x) = g(x)[1-g(x)]$$

$$\frac{\partial g(z_1)}{\partial z_1} = g(z_1)[1-g(z_1)] = a_1(1-a_1)$$

$\Rightarrow$

$$\frac{\partial L}{\partial w_1} = -(y-\hat{y}) \cdot w_2 \cdot a_1(1-a_1) \cdot x = (a_2-y) \cdot w_2 \cdot a_1(1-a_1) \cdot x$$

similarly,

$$\frac{\partial L}{\partial b_1} = (a_2-y)\, w_2 \cdot a_1(1-a_1)\, x$$

$$y = \tanh = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$\tanh' = 1 - y^2$$

$$\Rightarrow \frac{\partial L}{\partial w_1} = (a_2 - y) \cdot w_2 \cdot (1 - y^2)\ X$$

$$\frac{\partial L}{\partial b_1} = (a_2 - y)\ w_2 \cdot (1 - y^2)$$