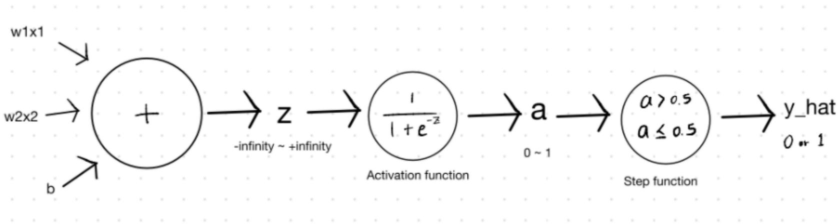
In [2]:

```
# classification model ex) classification positive or negative cancer
# Logistic regression is one of the classification algorithms.
```



Now we will go to use several features to classify samples. z is the linear function that has several variables, weights, and an intercept, so the value of z will be from negative infinity to positive infinity.

Z is the input of the sigmoid function and the output will be a. a will have the value from 0 to 1, which we can interpret as a probability. (ex, if a is bigger than 0.5, we can classify a sample as a case of positive cancer. if a is smaller than 0.5, we can classify a sample as the case of negative cancer.)

The purpose of Logistic regression is to increase the ratio of corrected classified samples.

$$Logistic\ loss\ function:$$
$$L = -(ylog(a) + (1-y)\log(1-a))$$

$$\text{if } y = 1\ (positive\ cancer),\quad L = -\log(a)$$
$$\text{if } y = 0\ (negative\ canver),\quad L = -\log(1-a)$$

$$we\ have\ \text{to}\ minimize\ the\ result\ of\ loss\ function\ \text{in}\ both\ cases.$$

$$In\ first\ case,\ absolute\ value\ of\ -\log(a)\ will\ be\ smallest$$
$$when\ a = 1.$$
$$(when\ we\ are\ trying\ \text{to}\ find\ smallest\ value\ of\ L,$$
$$a\ will\ be\ getting\ closer\ \text{to}\ 1.(positive\ case))$$

$$In\ second\ case,\ absolute\ value\ of\ -\log(a-1)\ will\ be\ smallest$$
$$when\ a = 0.$$
$$(when\ we\ are\ trying\ \text{to}\ find\ smallest\ value\ of\ L,$$
$$a\ will\ be\ getting\ closer\ \text{to}\ 0.(negative\ case))$$

Like we did before in the squared error part, we will find a derivative of the Logistic loss function in terms of weight and intercept. And we will subtract the value of the derivative of the Logistic loss function from the old weight.

$$\frac{dL}{dw} = \frac{dL}{da}\cdot\frac{da}{dz}\cdot\frac{dz}{dw}$$

$$\begin{aligned}
\frac{dL}{da} &= \frac{d}{da}(-(ylog(a) \ + \ (1-y)\log(1-a))). \\
&= -\left(y\frac{d}{da}\log(a) + (1-y)\frac{d}{da}\log(1-a)\right). \\
&= -\left(y\cdot\left(\frac{1}{a}\right) + (1-y)\left(\frac{1}{1-a}\right)\frac{d}{da}(1-a)\right). \\
&= -\left(y\left(\frac{1}{a}\right) + (1-y)\left(\frac{1}{1-a}\right)(-1)\right) \\
&= -\left(\frac{y}{a} - (1-y)\left(\frac{1}{1-a}\right)\right) \\
&= -\frac{y}{a} + \frac{(1-y)}{(1-a)}
\end{aligned}$$

$$\begin{aligned}
\frac{da}{dz} = \frac{d}{dz}\left(\frac{1}{1+e^{-z}}\right) &= \frac{d}{dz}(1+e^{-z})^{-1} \\
&= -(1+e^{-z})^{-2}\cdot\frac{d}{dz}(1+e^{-z}) \\
&= -(1+e^{-z})^{-2}\cdot -e^{-z} \\
&= \frac{e^{-z}}{(1+e^{-z})^2} \\
&= \frac{1}{(1+e^{-z})}\cdot\frac{e}{(1+e^{-z})} = \frac{1}{1+e^{-z}}\left(1-\frac{1}{1+e^{-z}}\right) \\
&= a(1-a)
\end{aligned}$$

$$\begin{aligned}
\frac{dz}{dw} &= \frac{d(w_1x_1 + w_2x_2 + \ldots w_ix_i + \ldots w_nx_n + b)}{dw_i} \\
&= x_i
\end{aligned}$$

$$\begin{aligned}
\frac{dL}{dw} = \frac{dL}{da}\cdot\frac{da}{dz}\cdot\frac{dz}{dw} &= \left(-\frac{y}{a} + \frac{(1-y)}{(1-a)}\right)\cdot(a(1-a))\cdot(x_i) \\
&= (-y(1-a) + a(1-y))\cdot(x_i) \\
&= (a-y)(x_i) = -(y-a)(x_i)
\end{aligned}$$

In similar way, derivative of L in *terms of b is:* $-(y-a)\cdot 1$

The process of differentiate L using chain rule is called 'back propagation'.

Therefore, we can update weight and *intercept* in *this way:*

- $w_i = w_i - \frac{dL}{dw_i} = w_i + \ (y-a)x_i$
- $b = b - \frac{dL}{db} = b + \ (y-a)\cdot 1$