Logistic regression

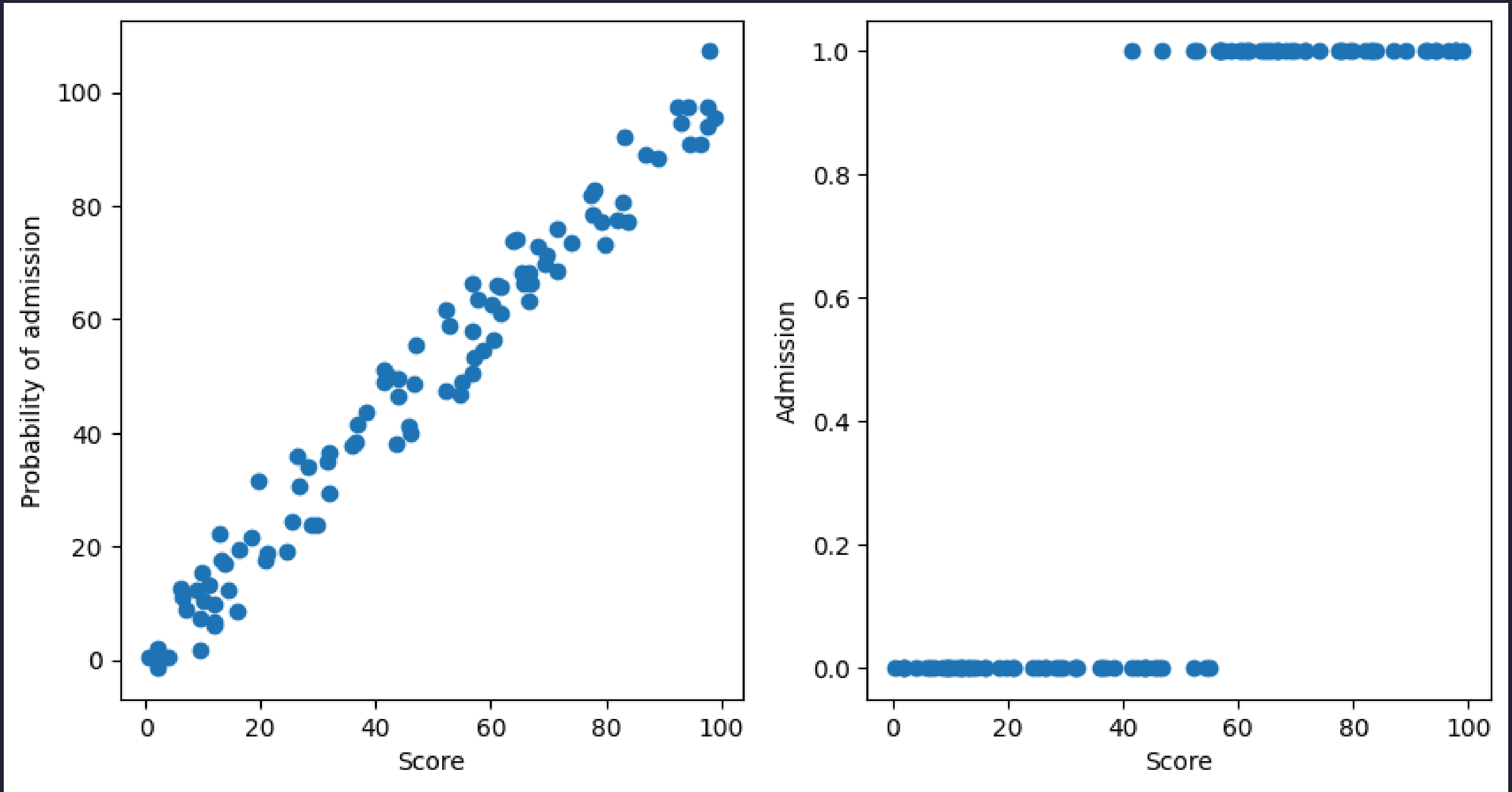
Regression

predict a number. Many possible outputs.

Classification

predict categories. Small number of possible outputs.

College admission

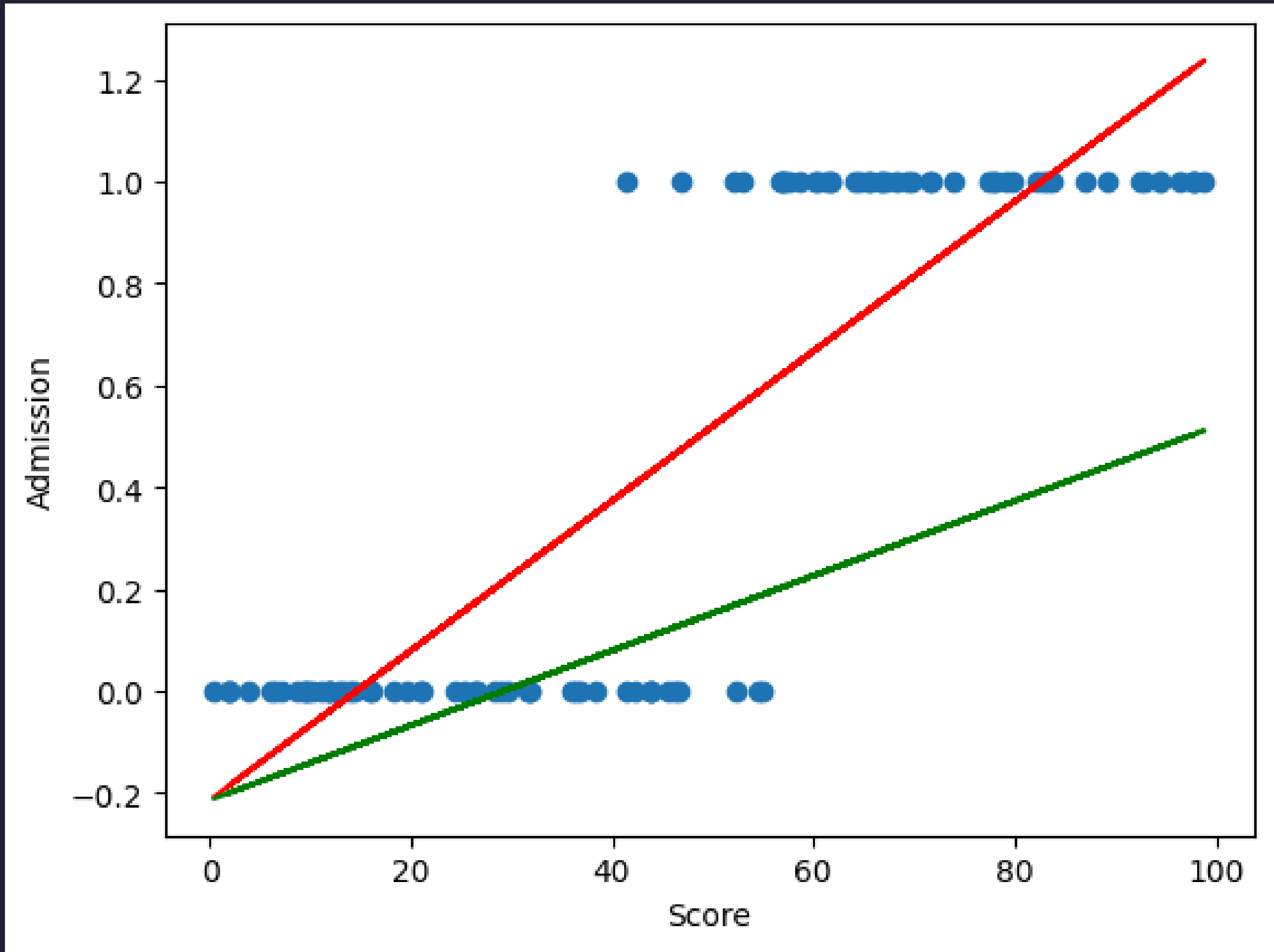


Why not use linear regression for classification?

$f(x) = wx + b$

if $f(x) \geq 0.5$, predict class 1

if $f(x) < 0.5$, predict class 0

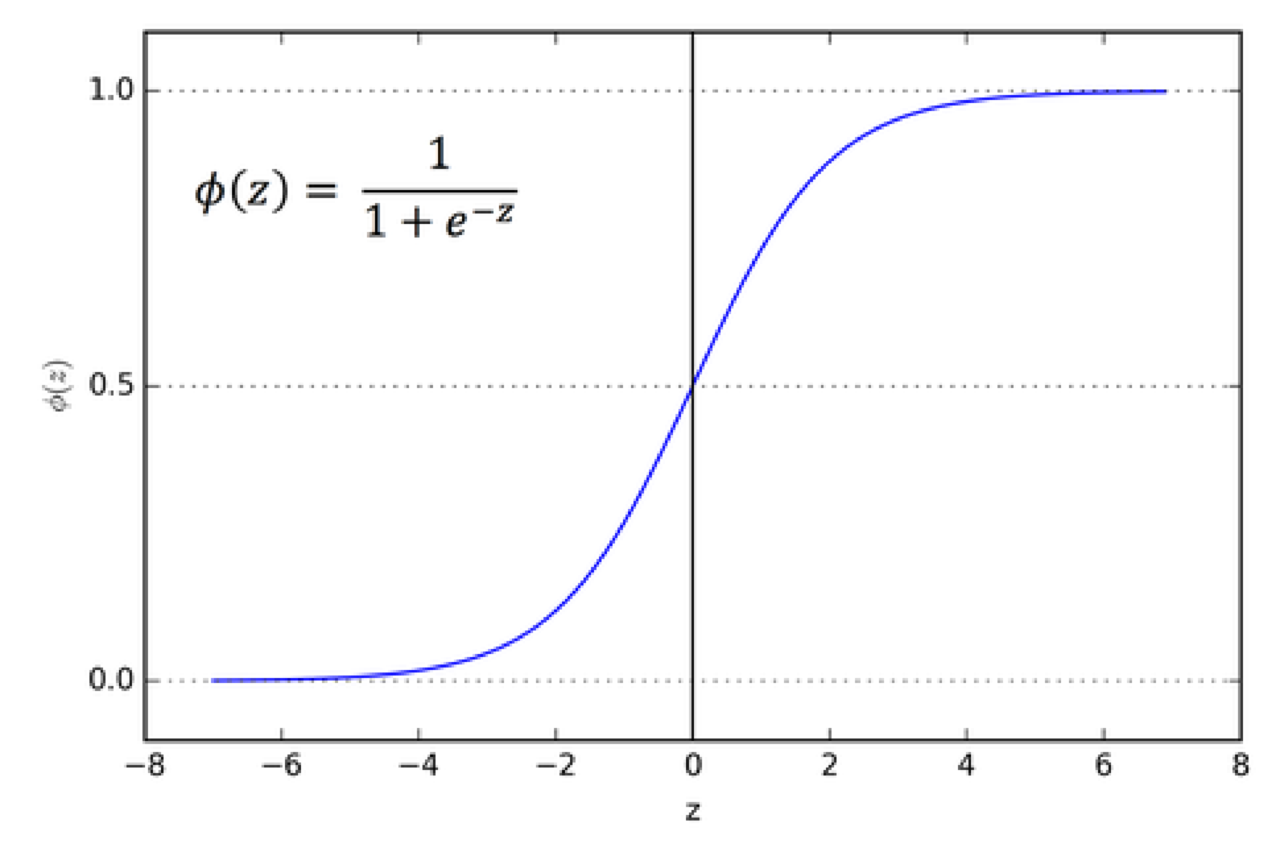


Logistic (sigmoid) function

$$g(z) = \frac{1}{1 + e^{-z}}$$

$$z \to \infty, g(z) = \frac{1}{1+e^{-\infty}} = 1$$
$$z \to \infty, g(z) = \frac{1}{1+e^{\infty}} = 0$$
$$z = 0, g(z) = \frac{1}{1+e^{0}} = 0.5$$

$$0 \leq g(z) \leq 1$$



Logistic regression: fitting logistic curve

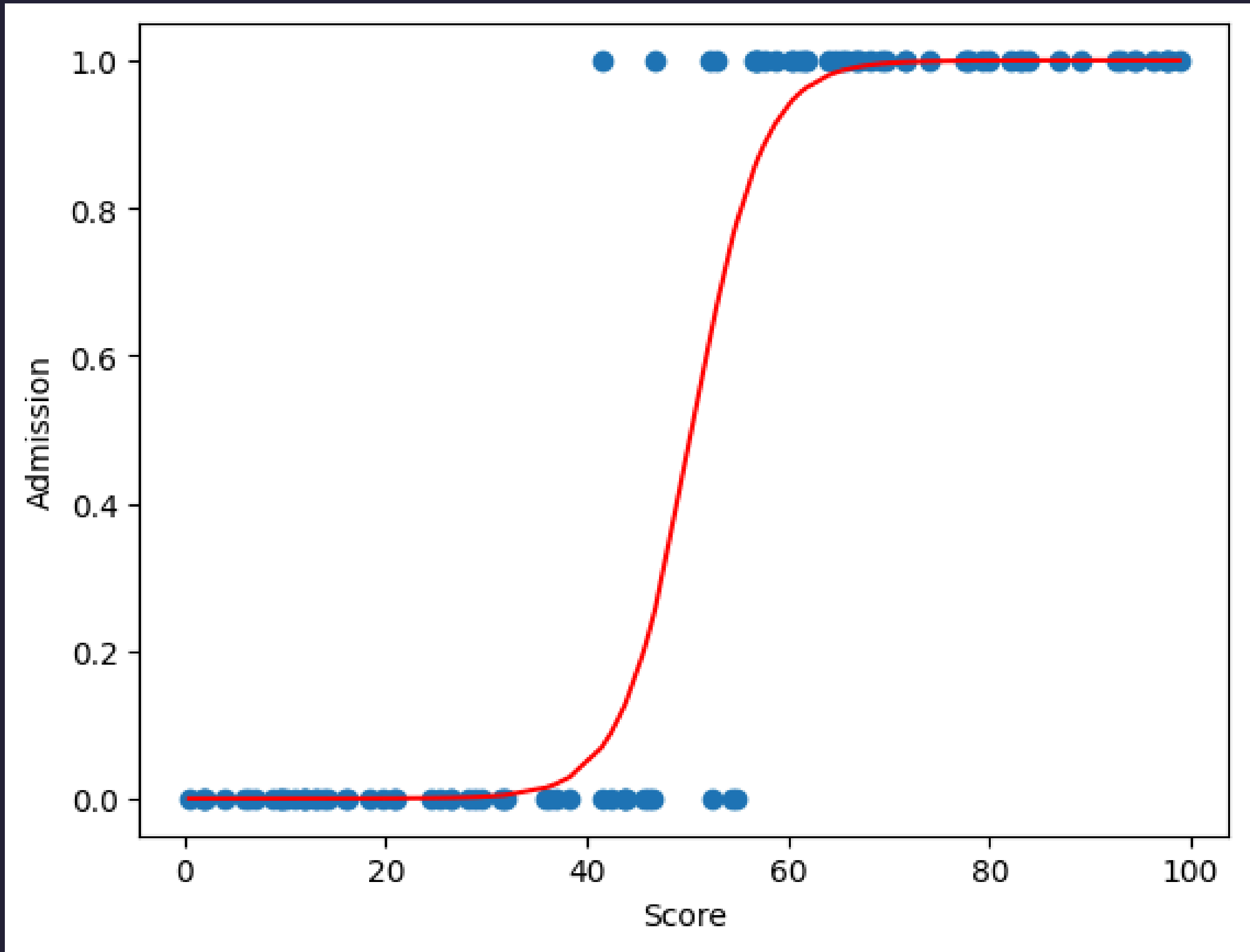
$$z = wx + b$$

$$f(x) = g(z) = \frac{1}{1 + e^{-z}} = \frac{1}{1 + e^{-(wx+b)}}$$

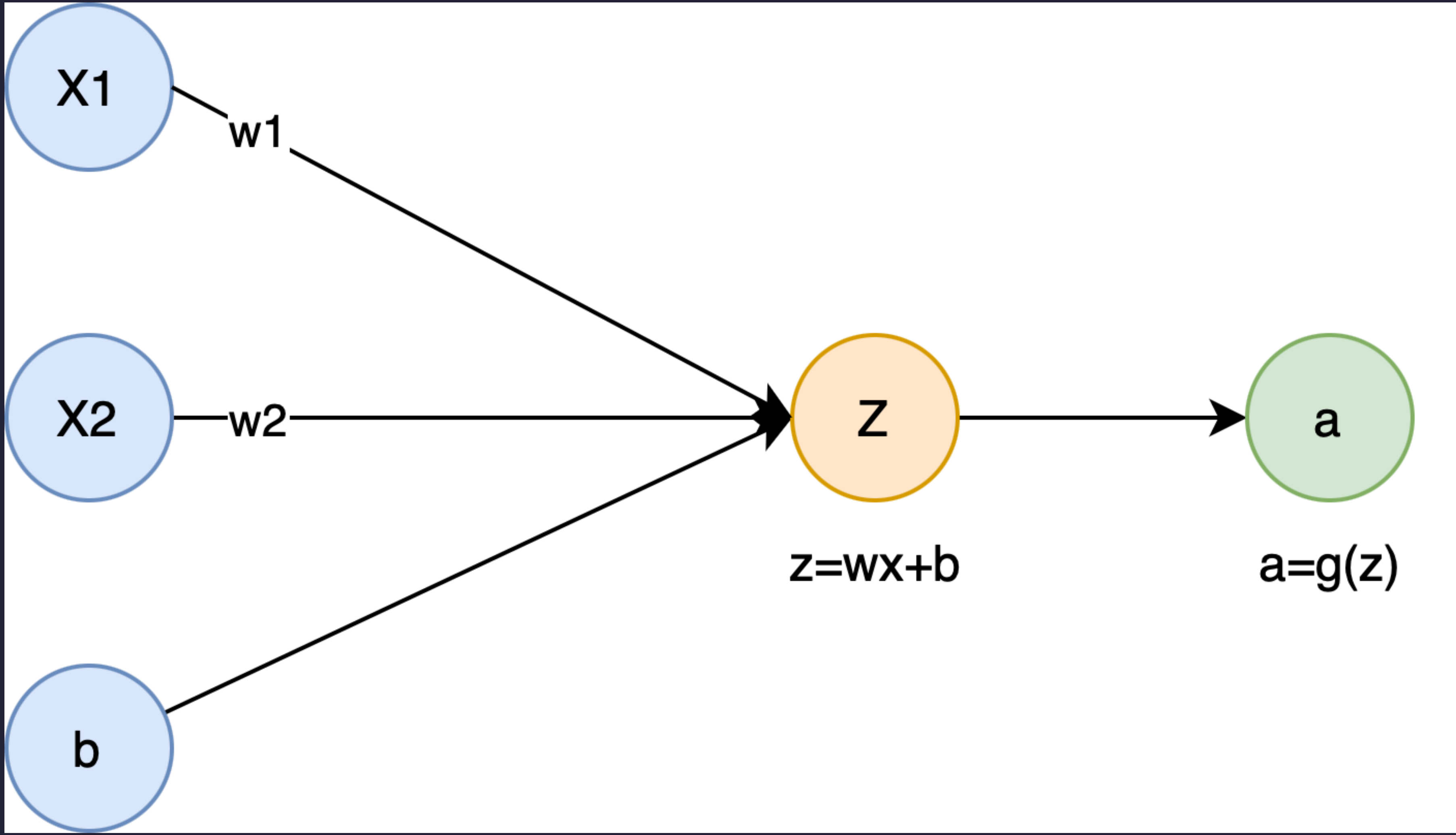
$$f(55) = 0.7$$

probability for 55 to be admitted: 0.7

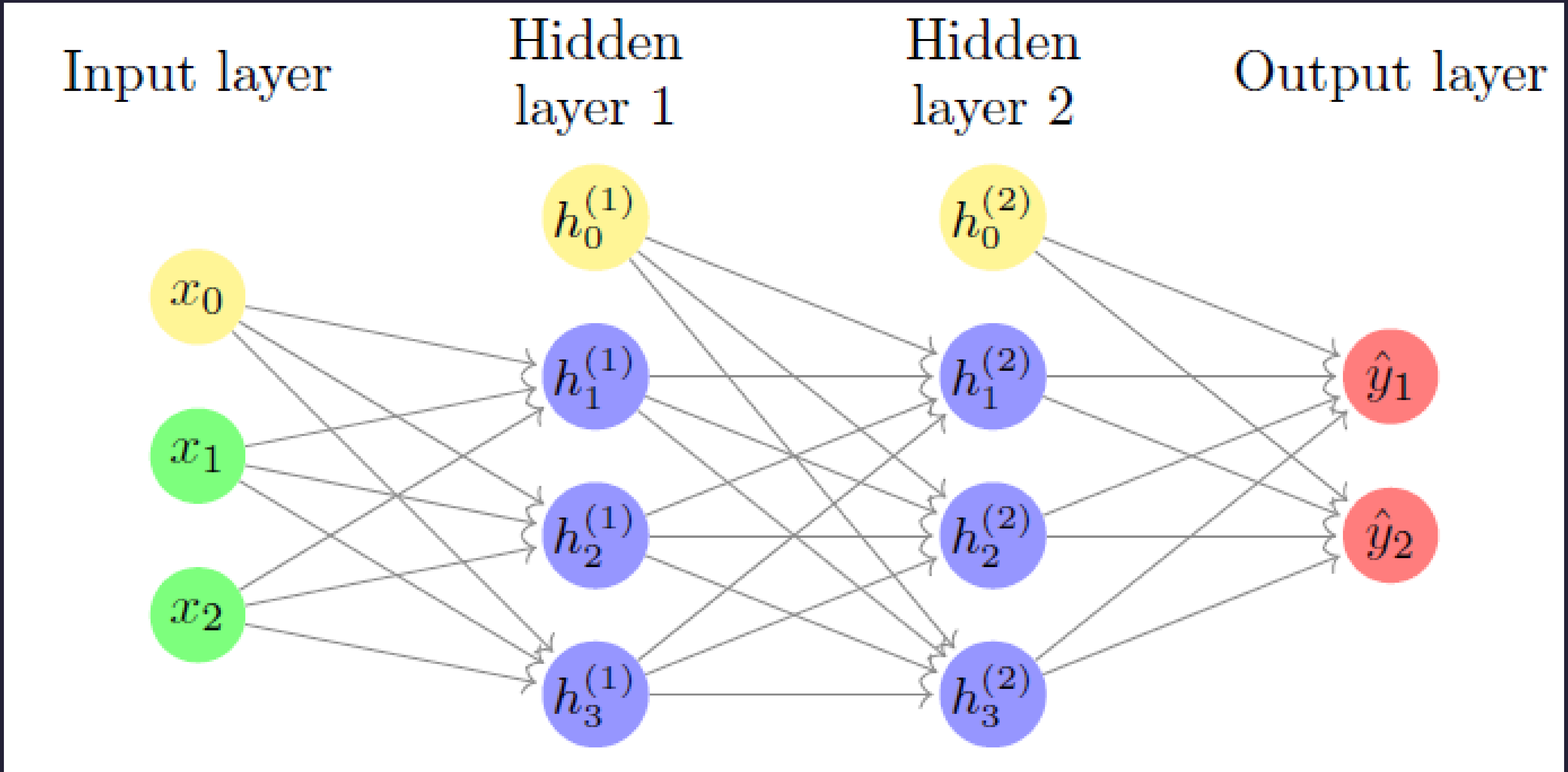
probability for 55 to be rejected: 0.3



Computation graph: forward pass



Computation graph: neural network



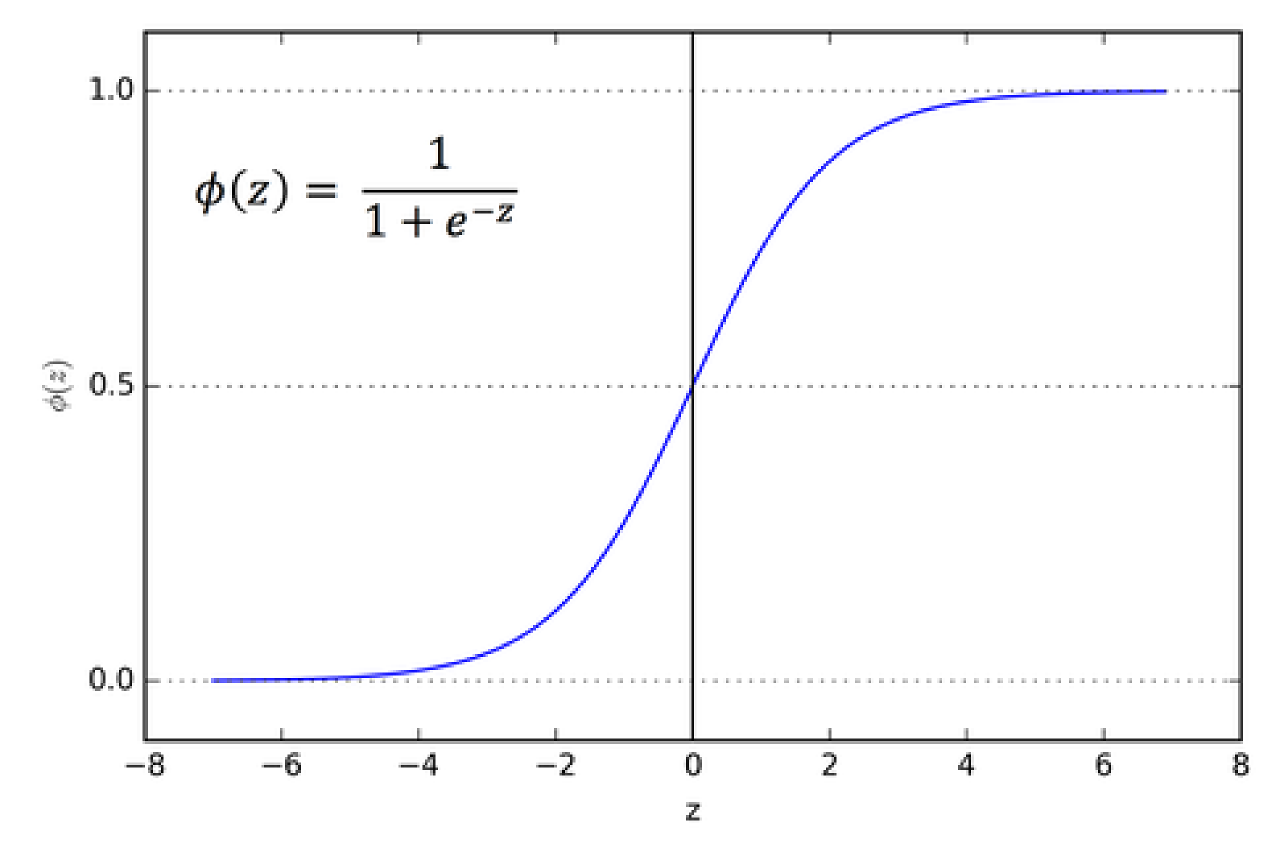
From probability to class

$$f(x) = g(z) = P(y = 1|x; w, b) = 0.7$$

- $g(z) \geq 0.5$
- $z \geq 0$

Decision boundary: threshold that separates the two classes

- $z = wx + b = 0$



Linear decision boundary

Two features: x_1 and x_2

w=[1,1], b=-1

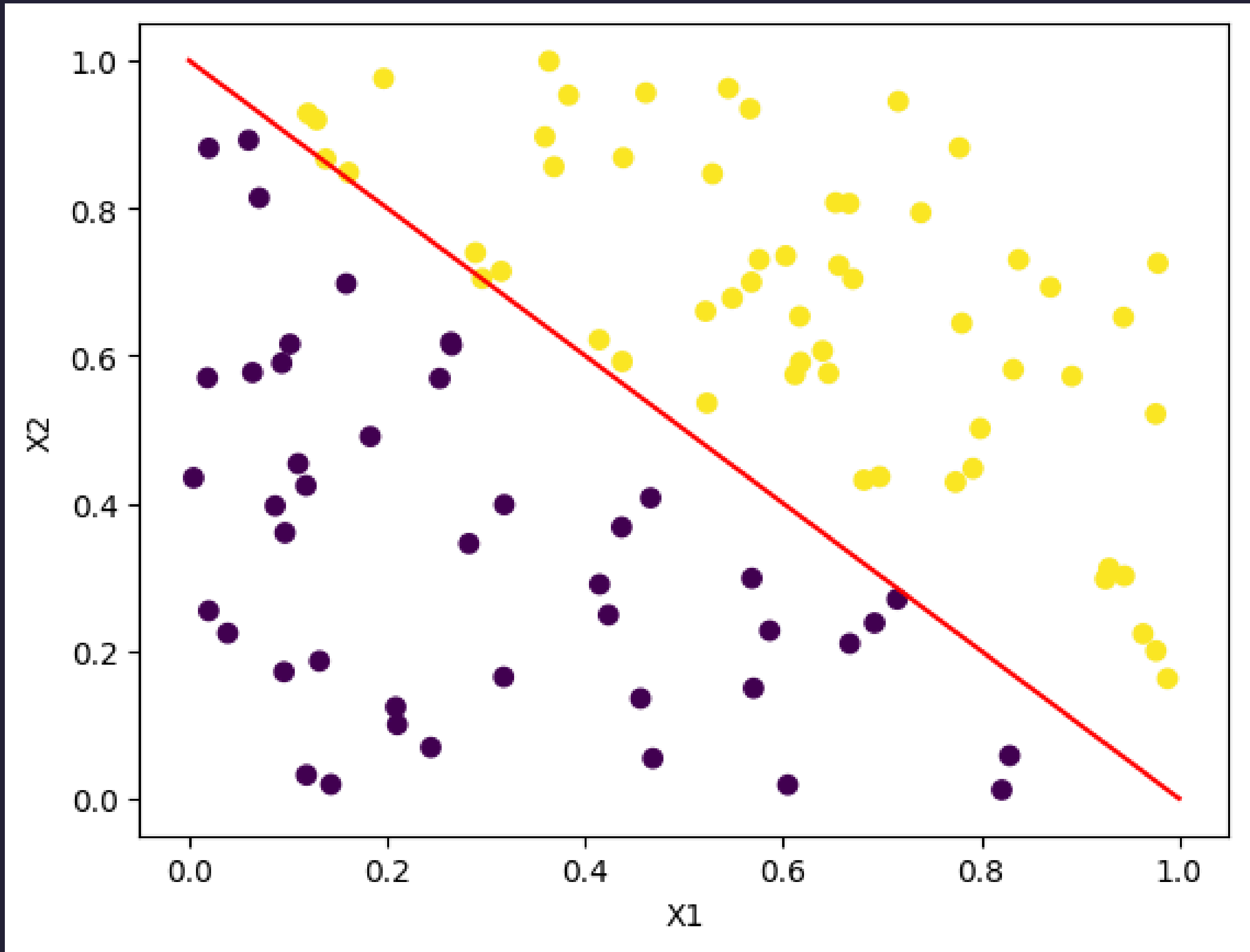
$$f(x) = g(w_1 x_1 + w_2 x_2 + b)$$
$$= g(x_1 + x_2 - 1) = 0.5$$

$$z = x_1 + x_2 - 1 = 0$$

$$x_1 + x_2 = 1$$

$x_1 + x_2 \geq 1$, predict class 1 (yellow)

$x_1 + x_2 < 1$, predict class 0 (blue)



Non-linear decision boundary

Two features: x_1 and x_2

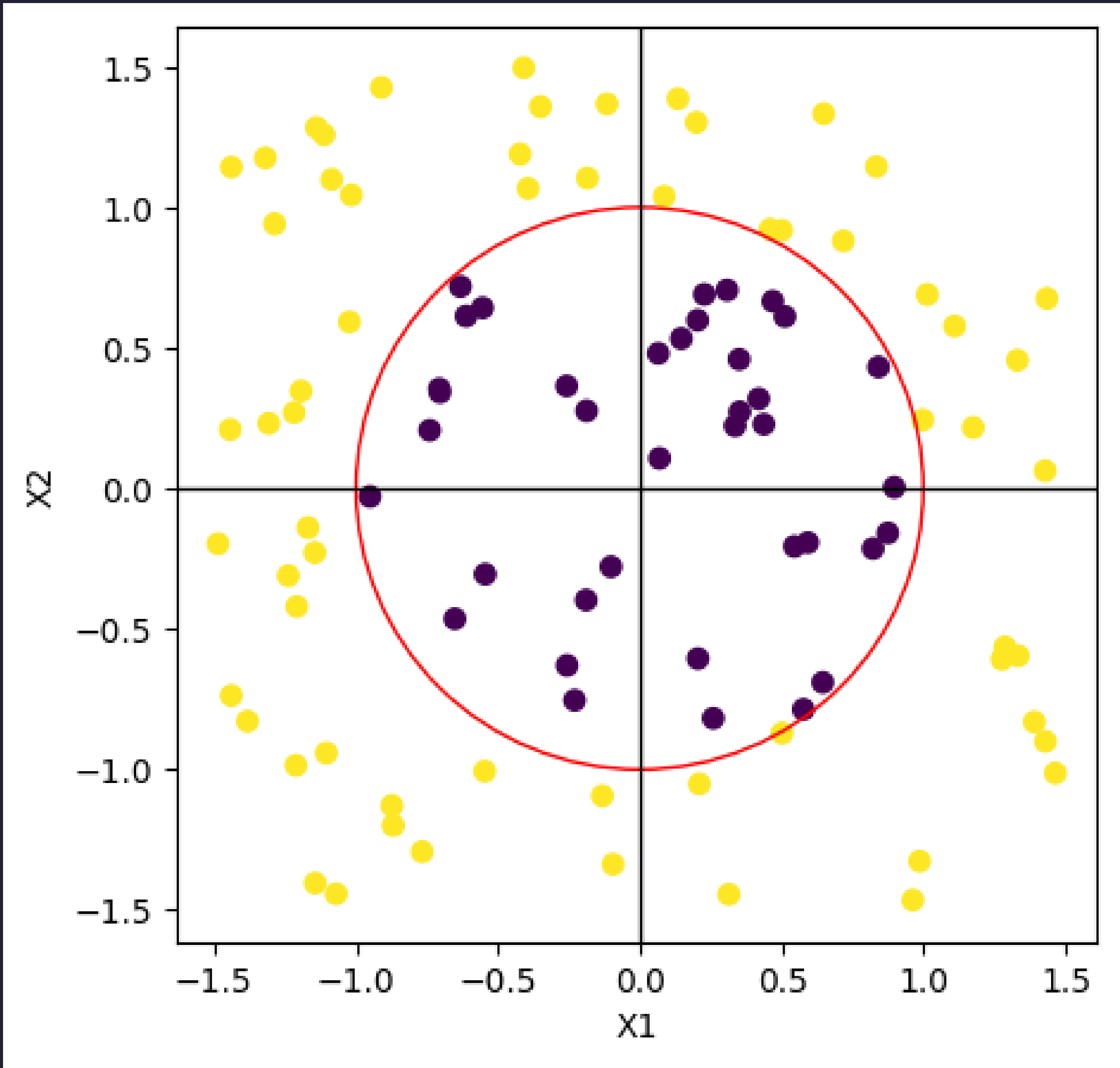
w=[1,1], b=-1

$$f(x) = g(w_1 x_1^2 + w_2 x_2^2 + b)$$
$$= g(x_1^2 + x_2^2 - 1) = 0.5$$

$$z = x_1^2 + x_2^2 - 1 = 0$$

$$x_1^2 + x_2^2 = 1$$

$x_1^2 + x_2^2 \geq 1$, predict class 1 (yellow)
$x_1^2 + x_2^2 < 1$, predict class 0 (blue)



💻 Logistic regression

Loss for logistic regression

$L(y, f(x))$

if $y = 1$,

- $L(1, f(x))$ should be small when $f(x)$ is close to 1
- $L(1, f(x))$ should be large when $f(x)$ is close to 0

if $y = 0$,

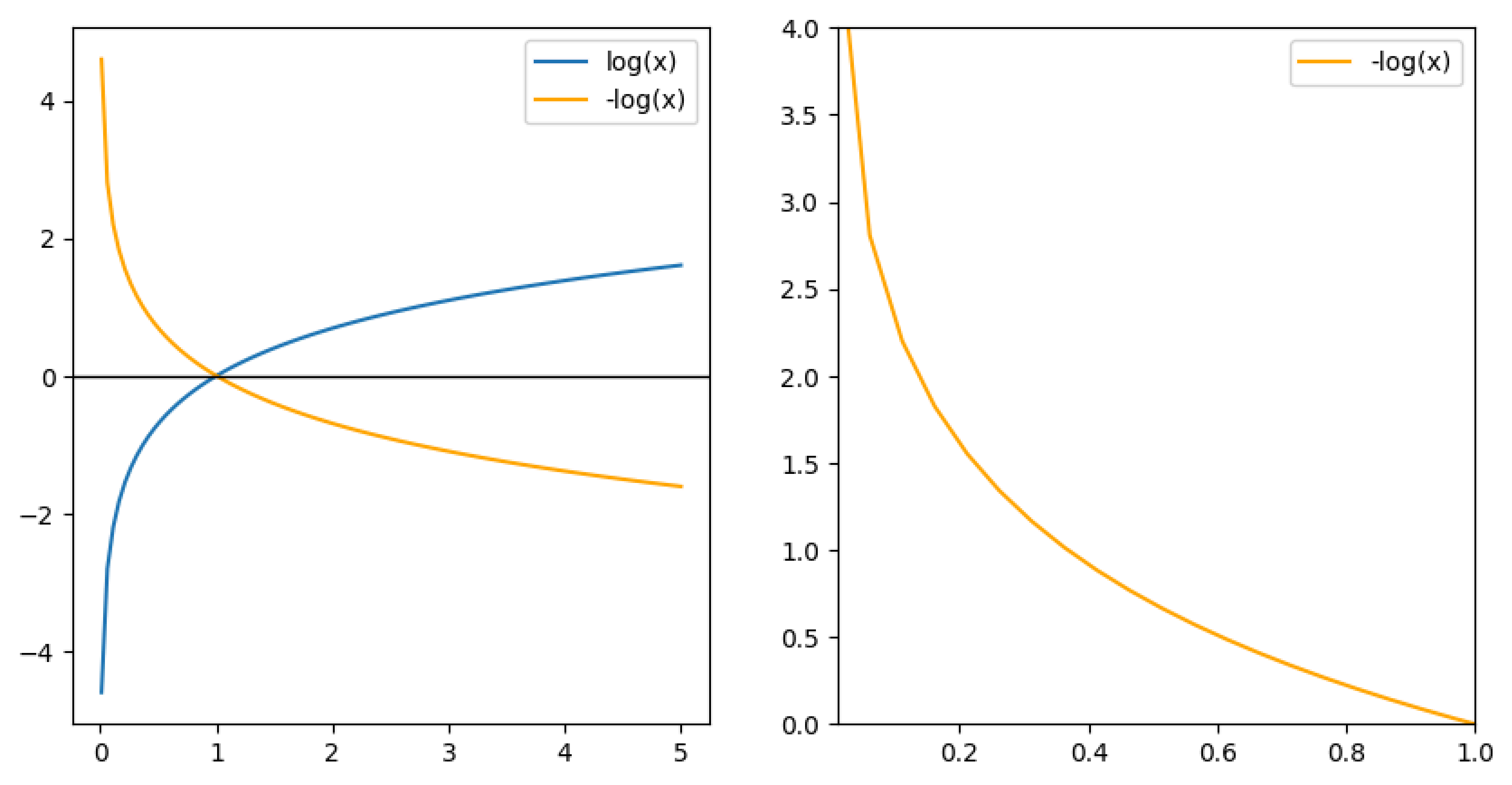
- $L(0, f(x))$ should be small when $f(x)$ is close to 0
- $L(0, f(x))$ should be large when $f(x)$ is close to 1

Log(istic) loss function

$$L(y, f(x)) = \begin{cases} -\log(f(x)) & \text{if } y = 1 \\ -\log(1 - f(x)) & \text{if } y = 0 \end{cases}$$

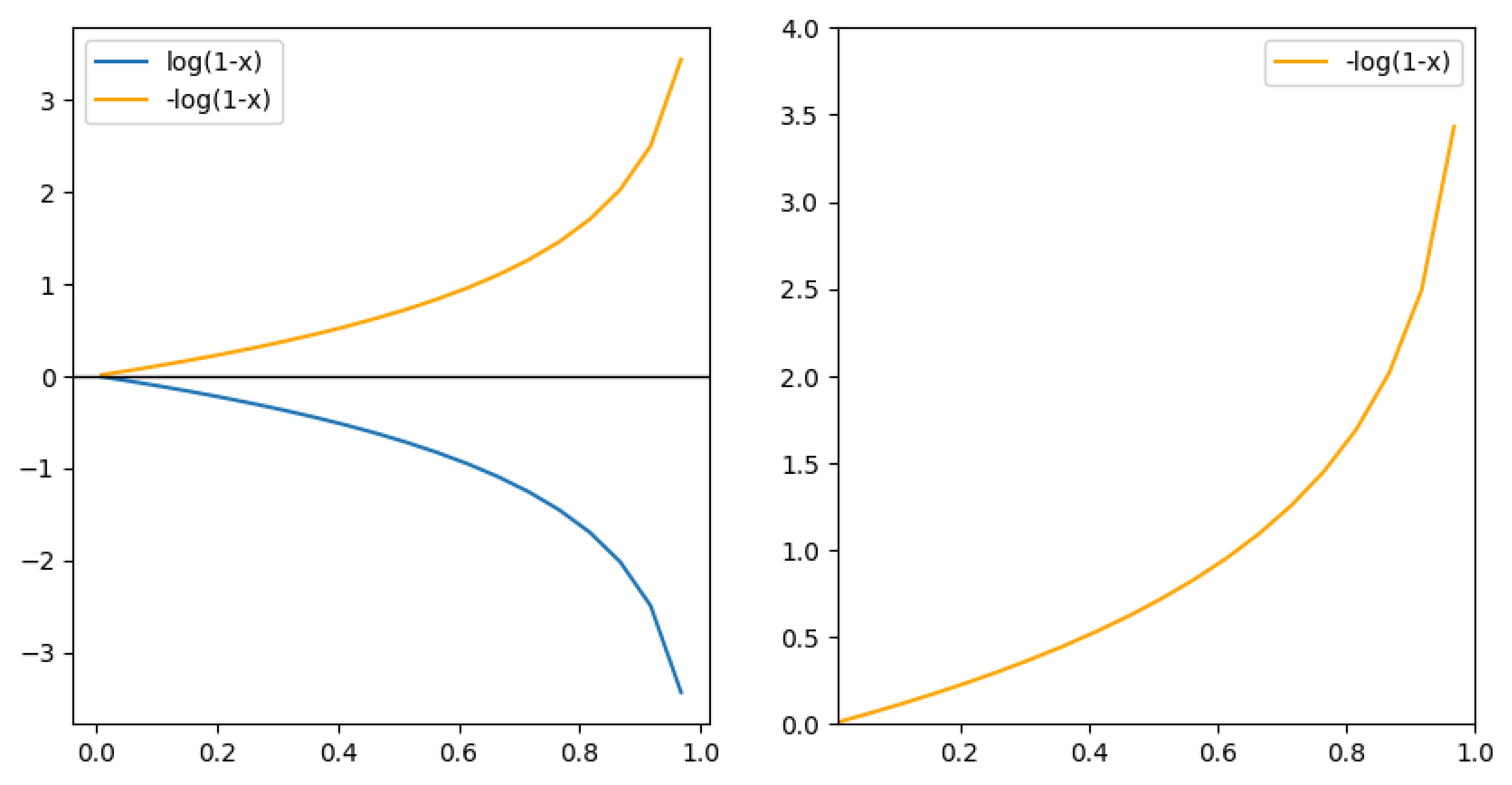
Log(istic) loss function (y=1)

$$\text{if } y = 1, L(1, f(x)) = -\log(f(x))$$



Log(istic) loss function (y=0)

if $y = 0, L(1, f(x)) = -\log(f(x))$



Combining the two cases

$$L(y, f(x)) = \begin{cases} -\log(f(x)) & \text{if } y = 1 \\ -\log(1 - f(x)) & \text{if } y = 0 \end{cases}$$

$$L(y, f(x)) = -y\log(f(x)) - (1 - y)\log(1 - f(x))$$

$$\text{if } y = 1, L(1, f(x)) = -1 * \log(f(x)) - (1 - 1)\log(1 - f(x)) = -\log(f(x))$$
$$\text{if } y = 0, L(1, f(x)) = -0 * \log(f(x)) - (1 - 0)\log(1 - f(x)) = -\log(1 - f(x))$$

Cost function: average loss

$$J(w, b) = \frac{1}{m} \sum_{i=1}^{m} L(y^{(i)}, f(x^{(i)})) = -\frac{1}{m} \sum_{i=1}^{m} [y^{(i)} \log(f(x^{(i)})) + (1 - y^{(i)}) \log(1 - f(x^{(i)}))]$$

Gradient descent for logistic regression

$$w_j = w_j - \alpha \frac{\partial J(\vec{w},b)}{\partial w_j}, j = 1, 2, \ldots, k$$

$$w_1 = w_1 - \alpha \frac{\partial J(\vec{w},b)}{\partial w_1}$$

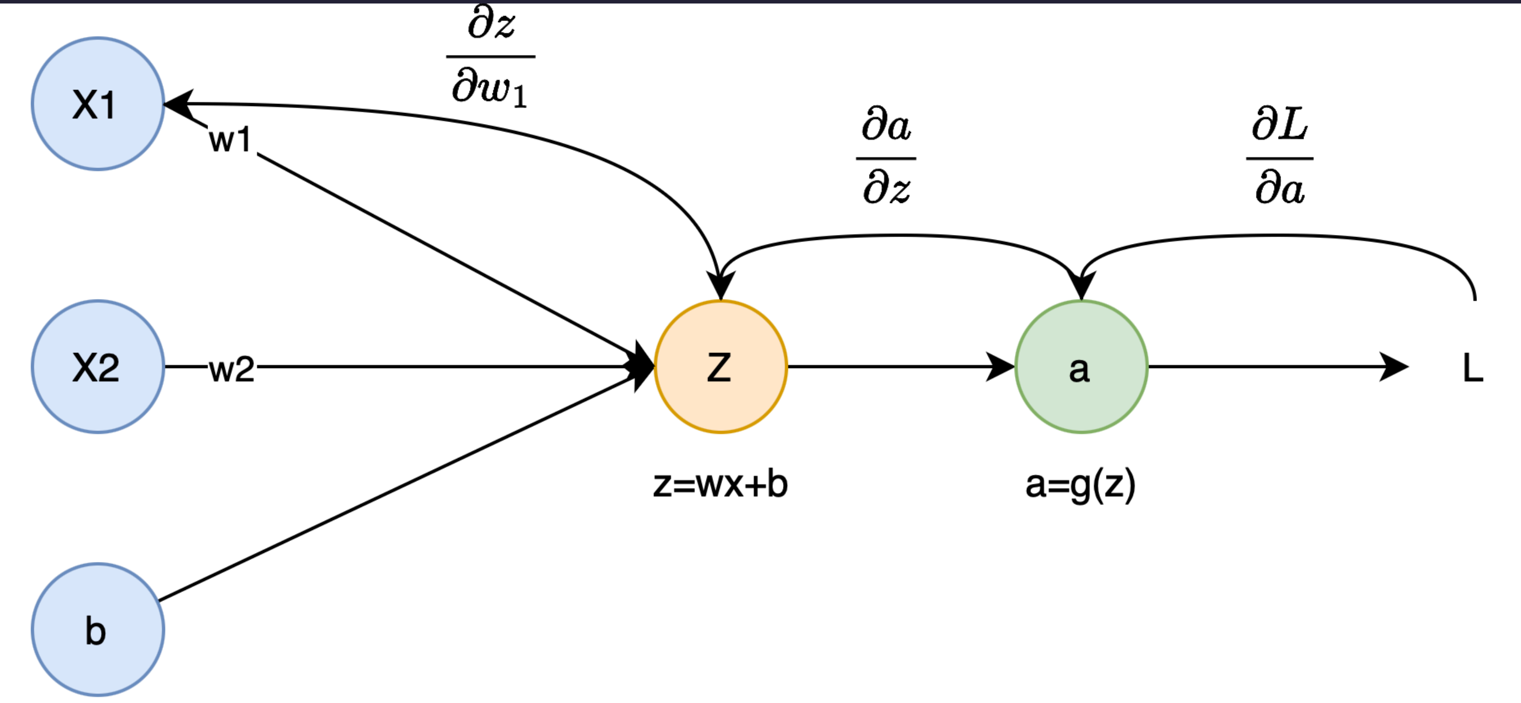
$$w_2 = w_2 - \alpha \frac{\partial J(\vec{w},b)}{\partial w_2}$$

...

$$w_k = w_k - \alpha \frac{\partial J(\vec{w},b)}{\partial w_k}$$

$$b = b - \alpha \frac{\partial J(\vec{w},b)}{\partial b}$$

Computation graph: backward pass



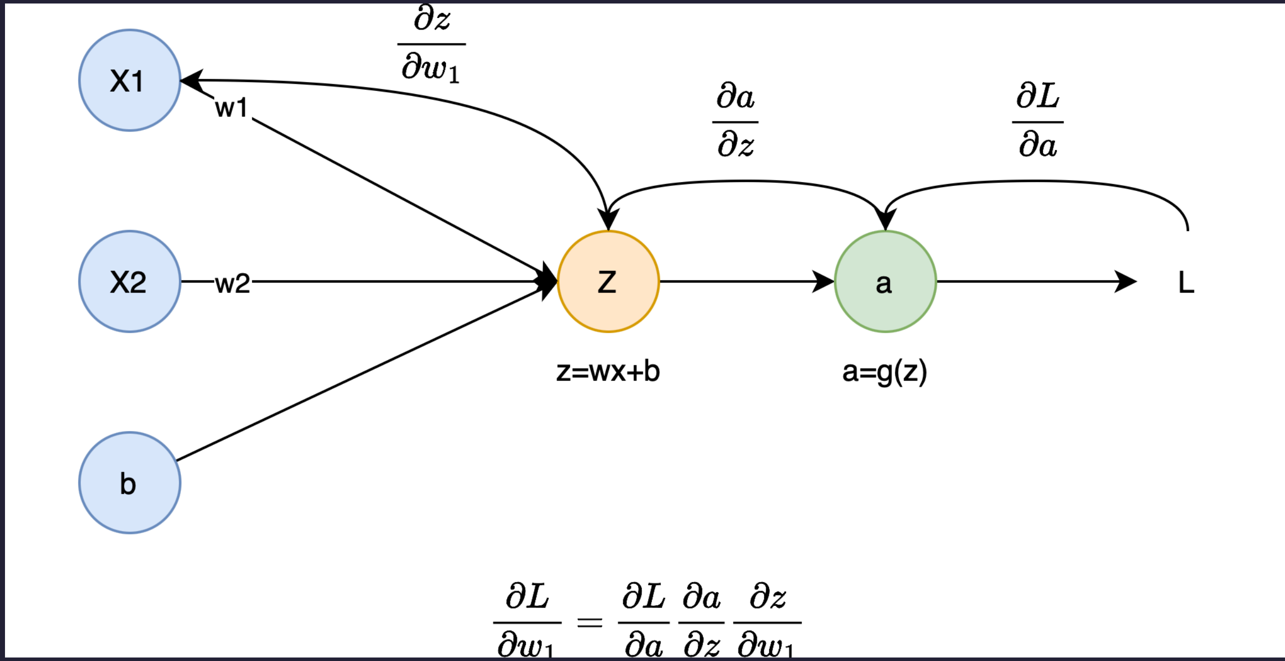
$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial a}\frac{\partial a}{\partial z}\frac{\partial z}{\partial w_1}$$

Gradient for weight w_j

$$L = -y\log(a) - (1-y)\log(1-a)$$
$$a = f(\vec{x}) = g(z) = \frac{1}{1+e^{-z}}$$
$$z = \vec{w}\cdot\vec{x} + b$$

$$\frac{\partial L}{\partial a} = -\frac{y}{a} + \frac{1-y}{1-a} = \frac{a-y}{a(1-a)}$$
$$\frac{\partial a}{\partial z} = a(1-a)$$
$$\frac{\partial z}{\partial w_j} = x_j$$

$$\frac{\partial L}{\partial w_j} = \frac{\partial L}{\partial a}\frac{\partial a}{\partial z}\frac{\partial z}{\partial w_j}$$
$$= \frac{a-y}{a(1-a)}\cdot a(1-a)\cdot x_j = (a-y)x_j$$

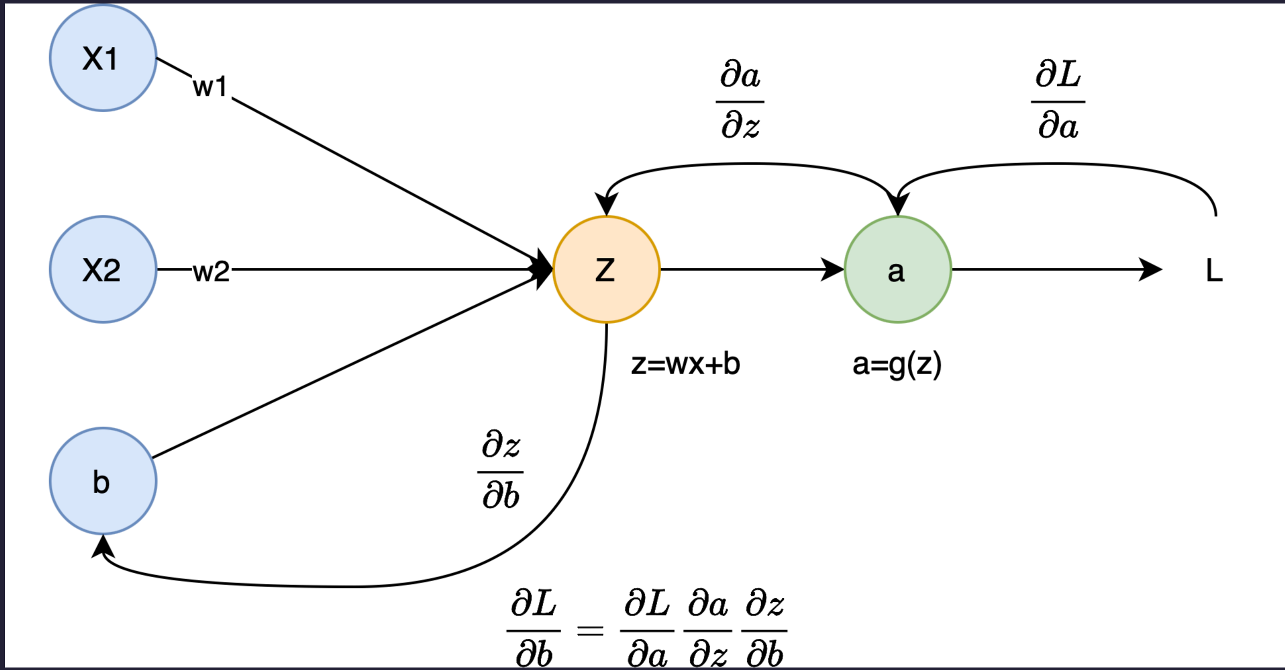


Gradient for bias b

$$z = \vec{w} \cdot \vec{x} + b$$
$$a = f(\vec{x}) = g(z) = \frac{1}{1+e^{-z}}$$
$$L = -y\log(a) - (1-y)\log(1-a)$$

$$\frac{\partial L}{\partial a} = -\frac{y}{a} + \frac{1-y}{1-a} = \frac{a-y}{a(1-a)}$$
$$\frac{\partial a}{\partial z} = a(1-a)$$
$$\frac{\partial z}{\partial b} = 1$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial a}\frac{\partial a}{\partial z}\frac{\partial z}{\partial b}$$
$$= \frac{a-y}{a(1-a)} \cdot a(1-a) \cdot 1 = (a-y)$$



Gradient for logistic regression

$z = \vec{w} \cdot \vec{x} + b$
$a = f(\vec{x}) = g(z) = \frac{1}{1+e^{-z}}$
$L = -y \log(a) - (1 - y) \log(1 - a)$
$J = \frac{1}{m} \sum_{i=1}^{m} L$

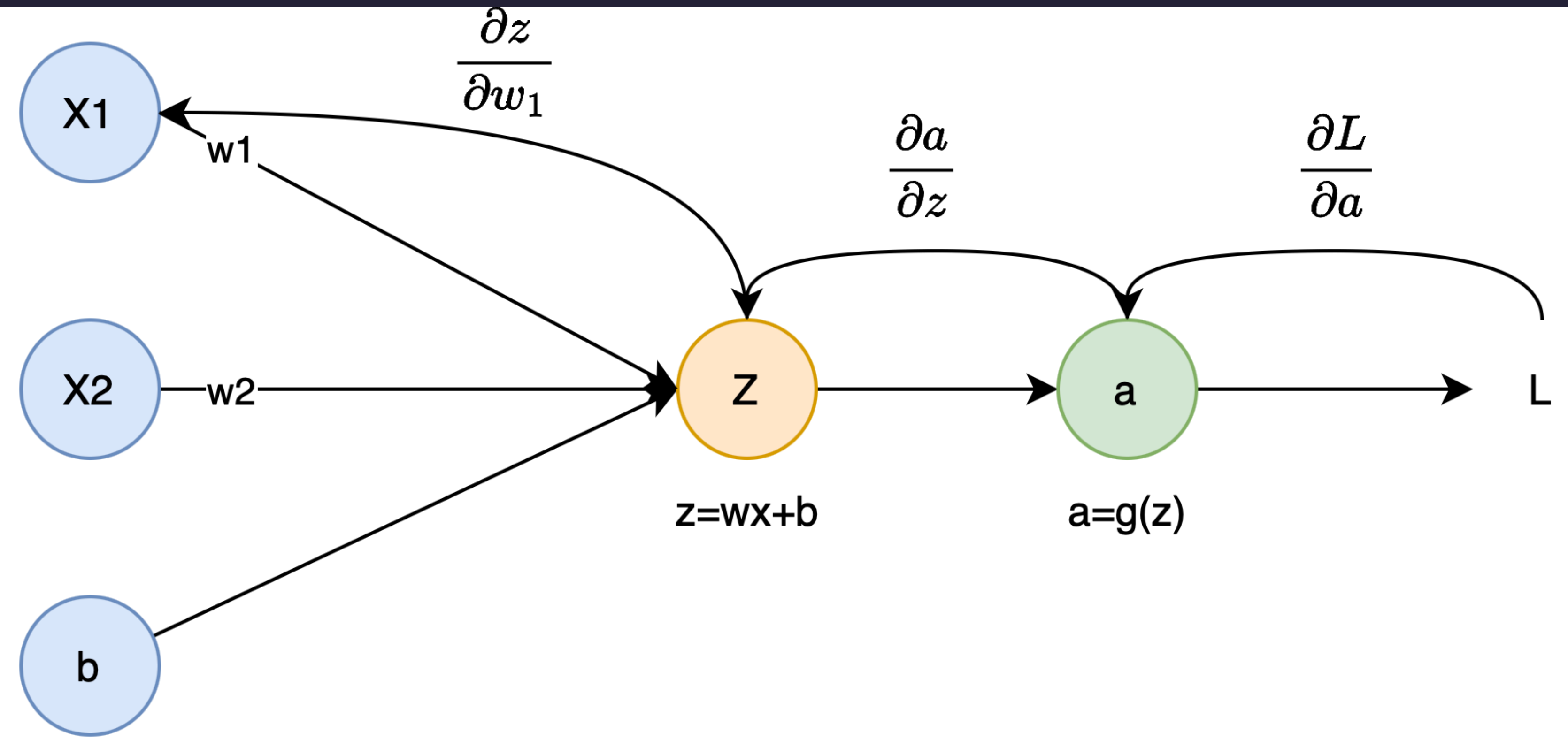
Partial derivative of the cost function with respect to w_j

$$\frac{\partial J(\vec{w}, b)}{\partial w_j} = \frac{1}{m} \sum_{i=1}^{m} (f(\vec{x}^{(i)}) - y^{(i)}) x_j^{(i)}$$

Partial derivative of the cost function with respect to b

$$\frac{\partial J(\vec{w}, b)}{\partial b} = \frac{1}{m} \sum_{i=1}^{m} (f(\vec{x}^{(i)}) - y^{(i)})$$

🤔 a, $\frac{\partial a}{\partial z}$ for linear regression?



$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial a}\frac{\partial a}{\partial z}\frac{\partial z}{\partial w_1}$$

💻 Gradient descent for logistic regression