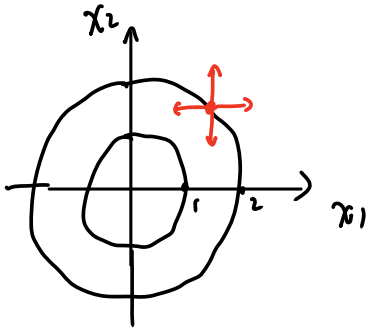
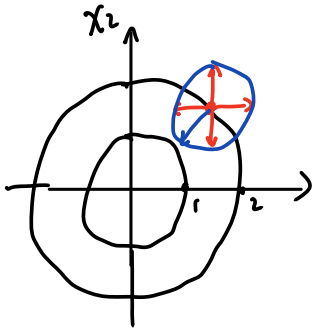
1. Goal : $\min_{x} f(x)$

Eg $f(x) = x_1^2 + x_2^2$



Contour line

x_1	x_2	f	
2	2	8	
1	2	5	☆
3	2	13	
2	1	5	☆
2	3	13	



$(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$ $\Rightarrow (-1, -1)$

$$f'(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \Rightarrow (1, 1)$$

$$x = x - \alpha f'(x) \qquad \Rightarrow (-\alpha, -\alpha)$$

2. $f(x)$

$$J(w, b \mid X, y)$$

Logistic regression

$$\hat{y} = \sigma(w_1 x_1 + w_2 x_2 + b)$$

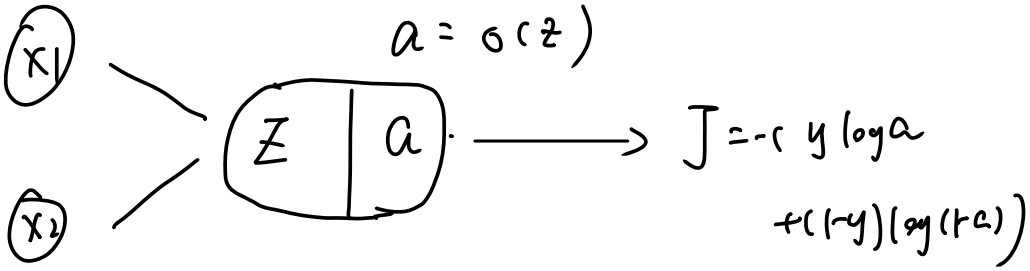
$$\text{cost} = J = -[\, y \log \hat{y} + (1-y) \log(1-\hat{y}) \,]$$

Goal: $\min_{w,b} J(w, b \mid X, y)$

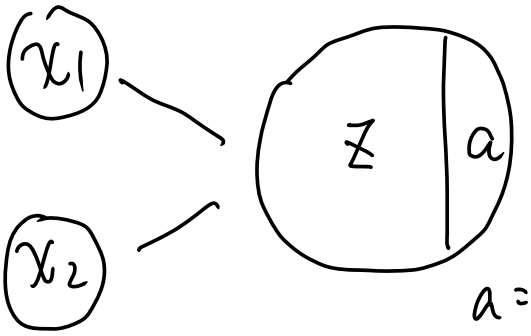
$$\Rightarrow \quad w_1 = w_1 - \alpha \cdot \frac{\partial J}{\partial w_1}$$

$$w_2 = w_2 - \alpha \cdot \frac{\partial J}{\partial w_2}$$

$$b = b - \alpha \cdot \frac{\partial J}{\partial b}$$



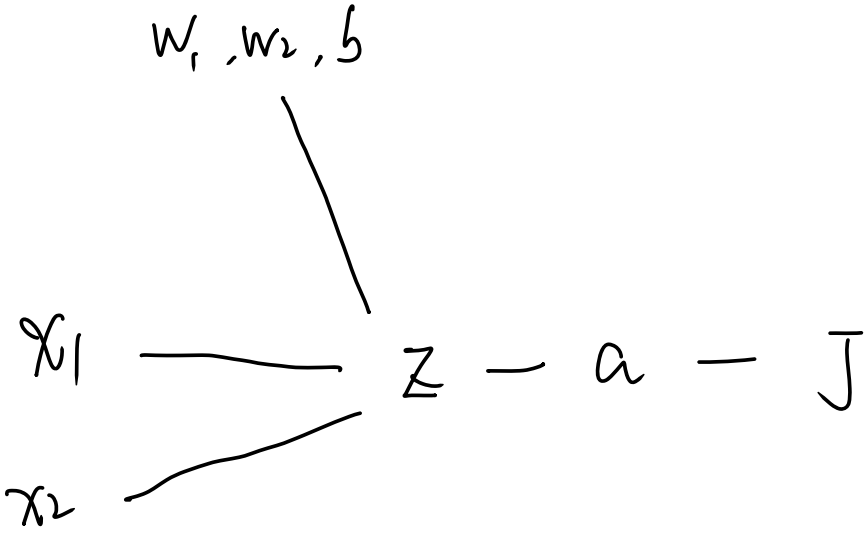
$$z = w_1 x_1 + w_2 x_2 + b$$



$$J = -(y \log a + (1-y)\log(1-a))$$

$$a = \sigma(z)$$

$$z = w_1 x_1 + w_2 x_2 + b$$



$$\frac{\partial J}{\partial w_1} = \frac{\partial J}{\partial a} \quad \frac{\partial a}{\partial z} \quad \frac{\partial z}{\partial w_1}$$

$$\frac{\partial J}{\partial a} = -\frac{y}{a} + \frac{1-y}{1-a}$$

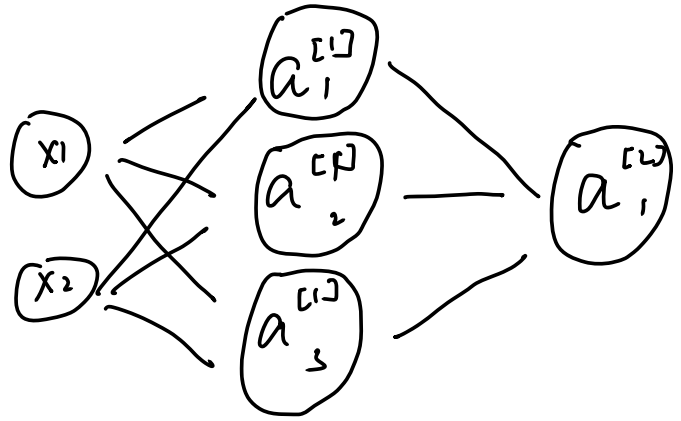
$$\frac{\partial a}{\partial z} = a(1-a)$$

$$\frac{\partial z}{\partial w_1} = x_1$$

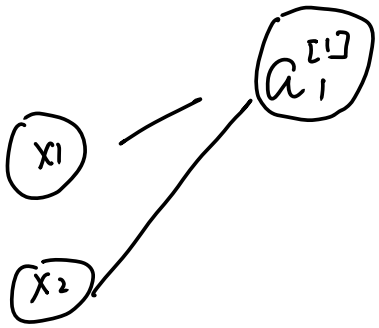
$$\left.\begin{aligned}\frac{\partial J}{\partial z} &= -(1-a)y + a(1-y) \\ &= a - y\end{aligned}\right]$$

$$\begin{cases} \frac{\partial J}{\partial w_1} = (a-y)\, x_1 \\ \frac{\partial J}{\partial w_2} = (a-y)\, x_2 \\ \frac{\partial J}{\partial b} = (a-y) \end{cases} \Rightarrow \frac{\partial J}{\partial w} = (a-y)\, x$$

3. Shallow Neural Net



v. y



Logistic Regression

$z = w^T x + b$

$a = g(z)$ ← Relu, tanh

PCA · $x_1\ x_2\ x_3\ x_4$

$\Rightarrow$ $y_1\ y_2\ y_3\ y_4$

$y_4 = w^T x$

$\sigma(y_4)$

W. b

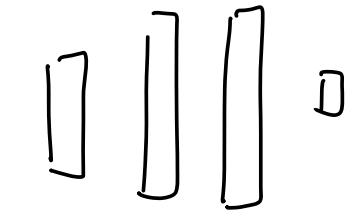
$w^{[1]}$ 2×3 $\begin{bmatrix} w_{1,1} & w_{2,1} \\ w_{1,2} & w_{2,2} \\ w_{1,3} & w_{2,3} \end{bmatrix}$ $\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$

x_1 a_1

a_2

x_2 a_3

$w^{[2]}$ $a_1^{[1]}$, $a_2^{[1]}$, $a_3^{[1]}$ $\rightarrow$ $a_1^{[2]}$ $\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$ 3×1 $[b]$



x

$a^{[0]}$ $a^{[1]}$ $a^{[2]}$ $a^{[3]}$

$$z^{[1]} = W^{[1]T} a^{[0]} + b^{[1]}$$

$$a^{[1]} = \sigma(z^{[1]})$$

$W^{[1]}$: 第0层 # nodes × 第1层 # nodes

$a^{[0]}$: 第0层 # nodes × m

$z^{[1]}, b^{[1]}$: 第1层 # nodes × m