Logistic & softmax Regression

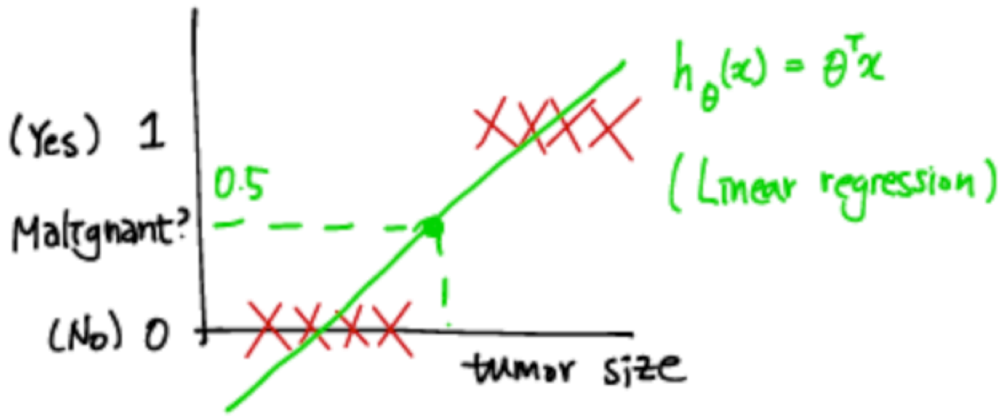
발표자 : 김동휘

logistic

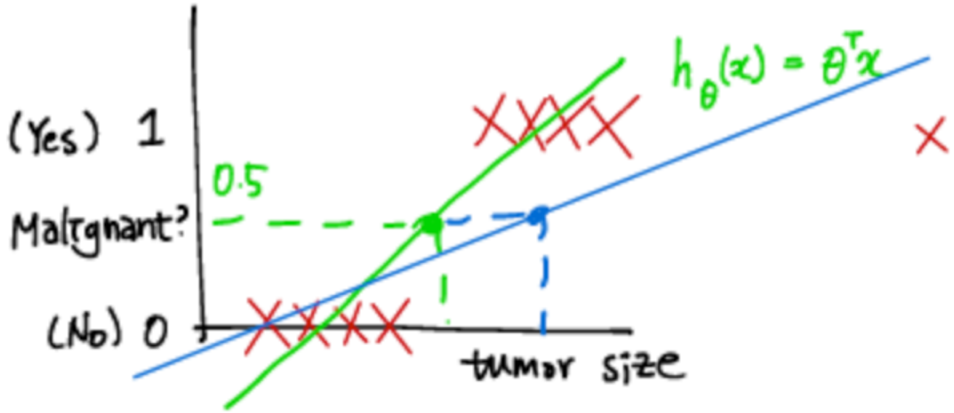
- Logistic regression -> Continoåus
- Logistic classification -> Discrete

- Binary classification(0 , 1)
- Spam, Facebook, Credit card etc..

Linear Regression for Classification



Linear Regression for Classification



When Large x values are received,
Leaving the existing x out of range.

Linear Regression for Classification

$$H(x) = Wx + b$$

X [1, 2, 5, 10, 11]

$W = 0.5$, $b = 0$

$x = 100 \rightarrow \boxed{\times 0.5} \rightarrow 50 \gg 1$

Logistic Regression Hypotheis Model(sigmoid)

$h_\theta(x) = g(\theta^T x)$

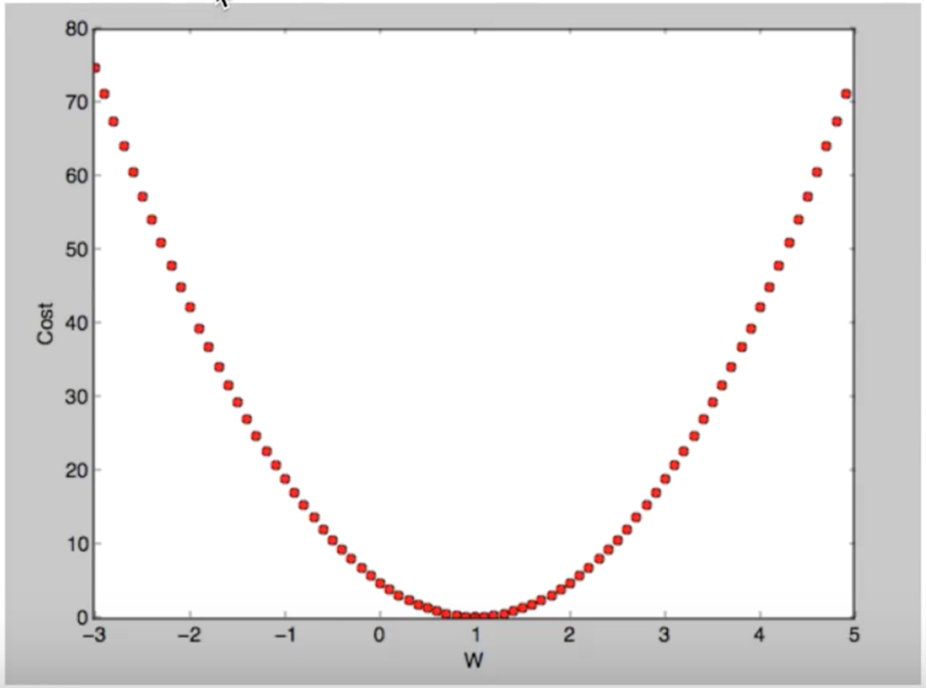
$g(z) = \dfrac{1}{1+e^{-z}}$

(sigmoid fct, or logistic fct)

$\Rightarrow h_\theta(x) = \dfrac{1}{1+e^{-\theta^T x}}$

Linear Regression Cost Function

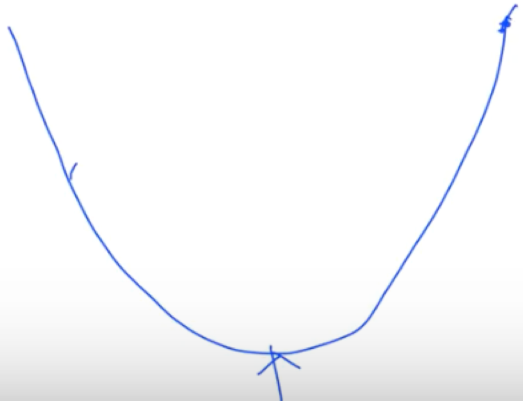
$$cost(W,b) = \frac{1}{m}\sum_{i=1}^{m}(H(x^{(i)}) - y^{(i)})^2 \quad \text{when} \quad H(x) = Wx + b$$

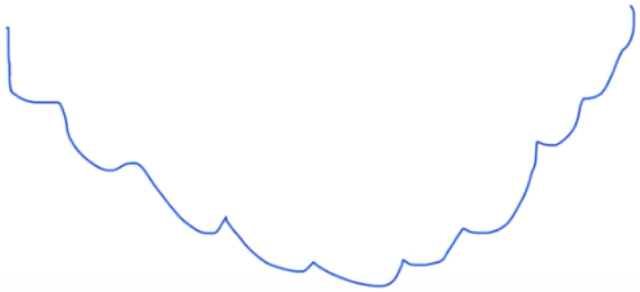


Logistic Regression Cost Function

$$H(x) = Wx + b$$

$$H(X) = \frac{1}{1 + e^{-W^T X}}$$





Fall into local minimum

New Cost Fuction for logistic

New cost function for logistic

$$cost(W) = \frac{1}{m} \sum \quad c(H(x), y)$$

$$c(H(x), y) = \begin{cases} -log(H(x)) & : y = 1 \\ -log(1 - H(x)) & : y = 0 \end{cases}$$

Why log function

Gradient decent algorithm for logistic

$$cost(W) = -\frac{1}{m}\sum y log(H(x)) + (1-y) log(1-H(x))$$

$$W := W - \alpha \frac{\partial}{\partial W} cost(W)$$

Softmax Classification

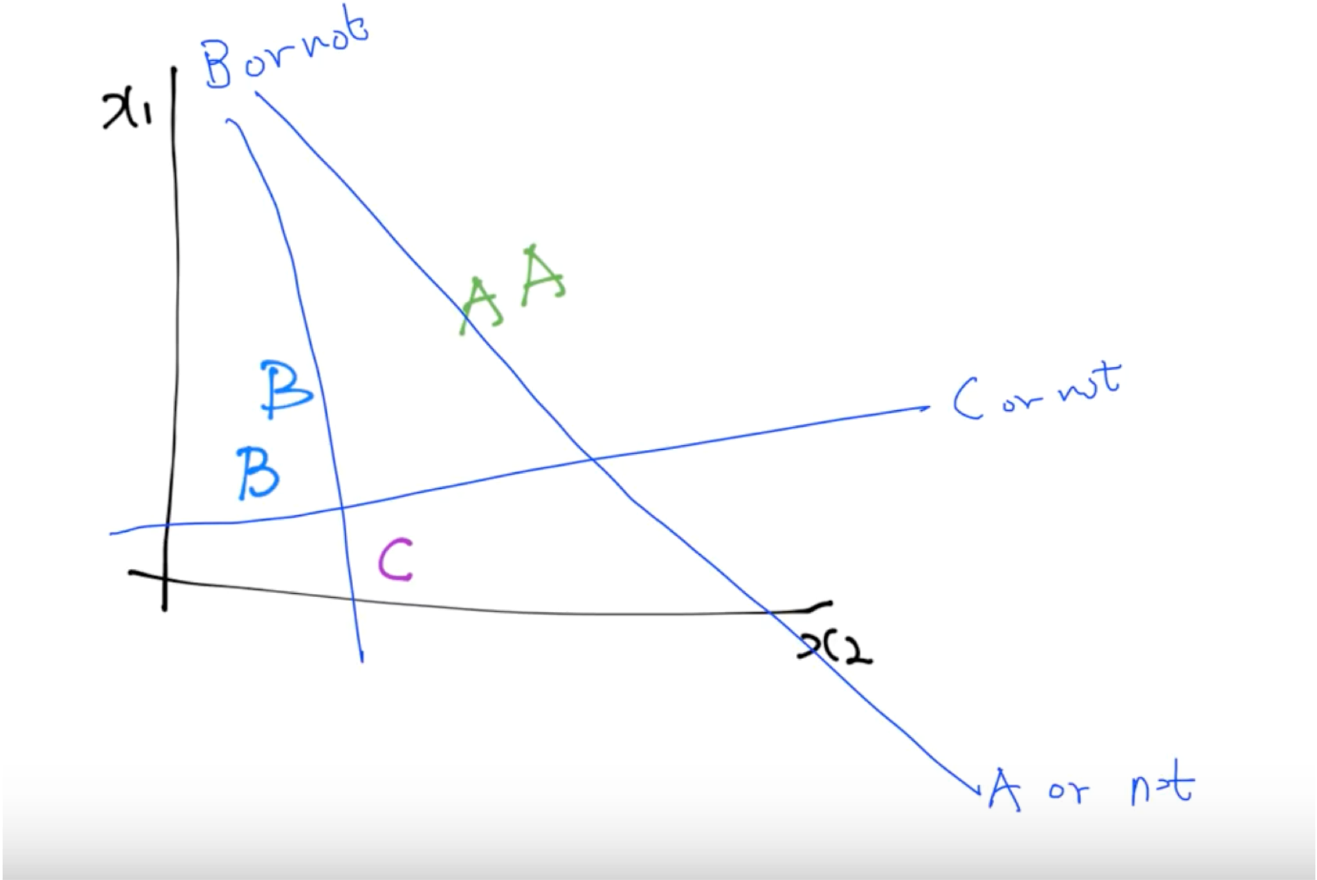
- Multinomial Classification
- model that extends binary organization
- Hardmax = the concept of finding the largest value in a statistic
- Softmax = A concept that looks for the largest value under the new condition

Multinomial Classification

x1 (hours)	x2 (attendance)	y (grade)
10	5	A
9	5	A
3	2	B
2	4	B
11	1	C

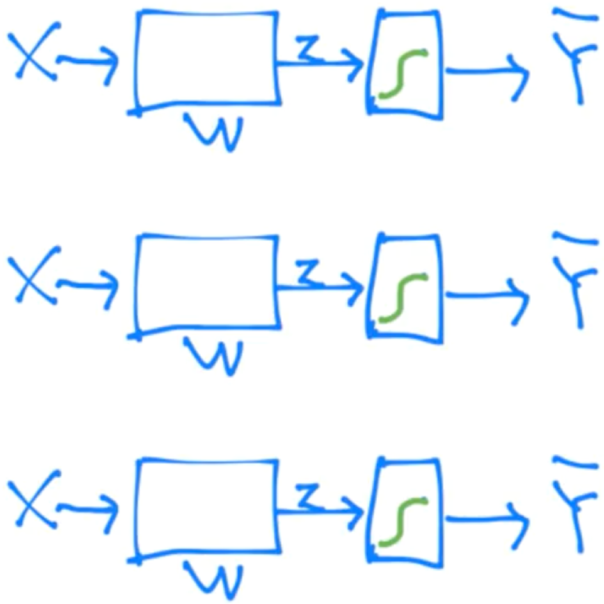
A or B or C

Multinomial Classification



Multinomial Classification

$$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \\ w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \\ w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix} = \begin{bmatrix} \bar{y}_A \\ \bar{y}_B \\ \bar{y}_C \end{bmatrix}$$



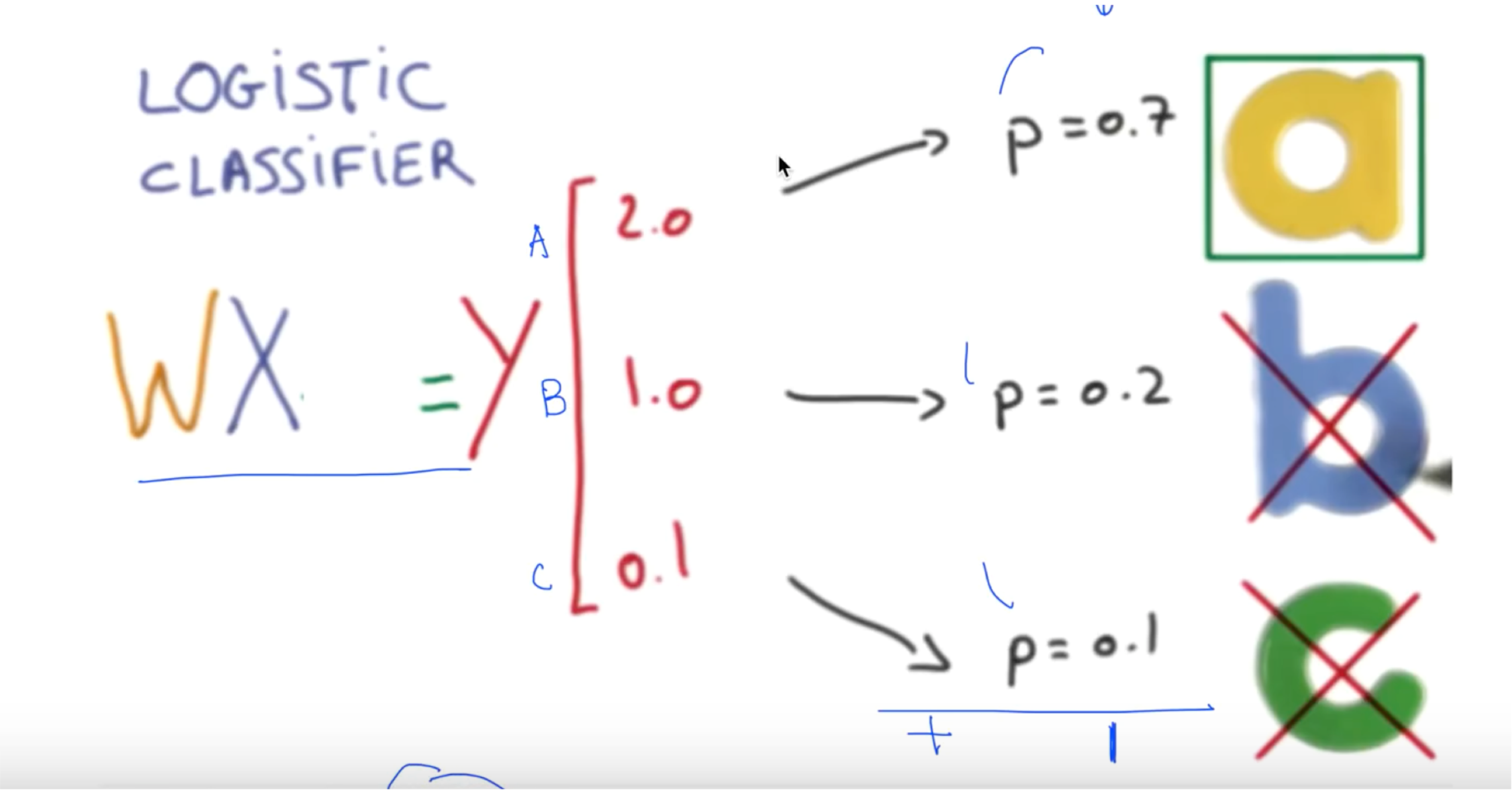
Sigmoid

0~1

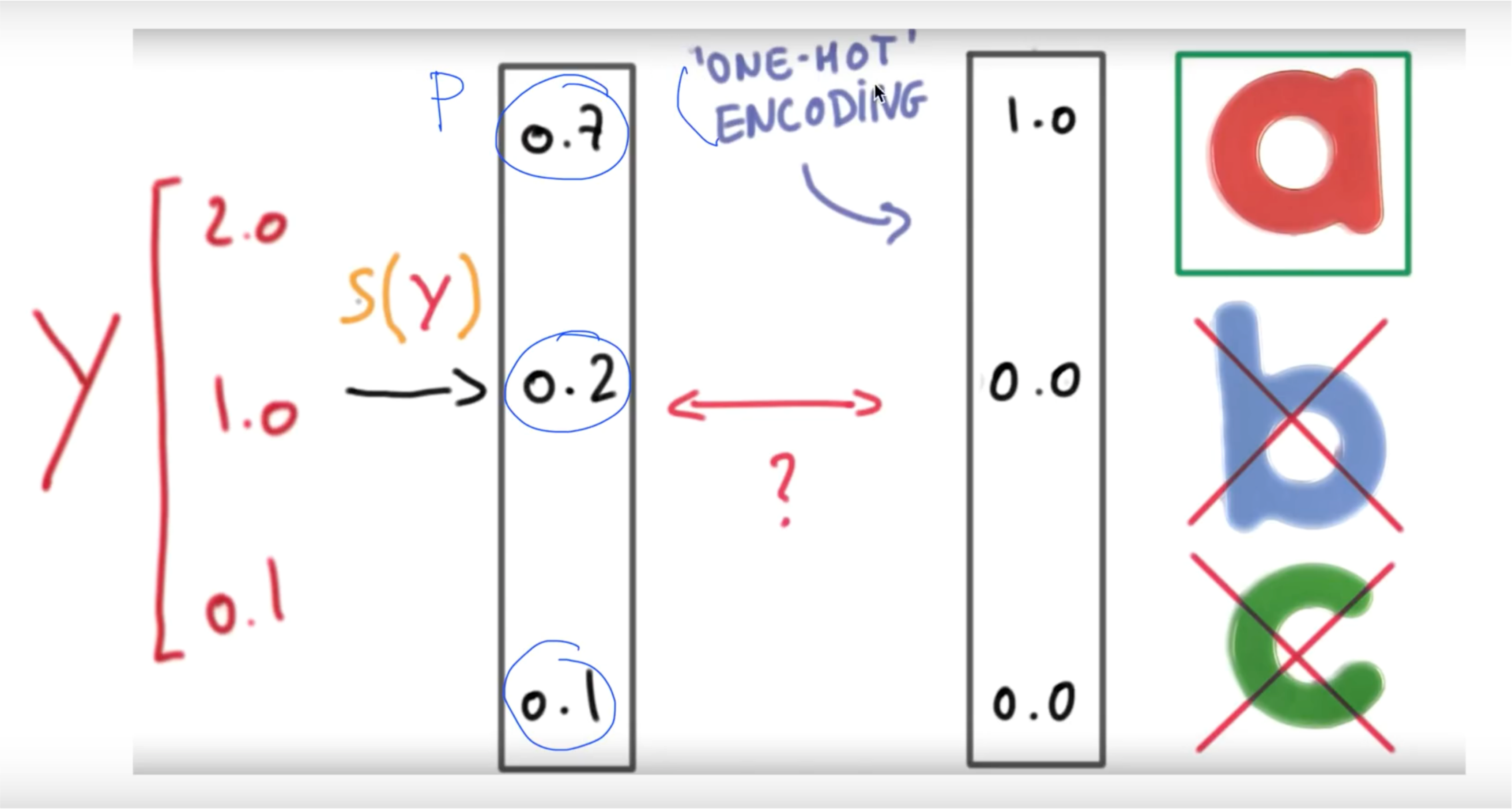
$$= \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \\ w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \\ w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix} = \begin{bmatrix} \bar{y}_A \\ \bar{y}_B \\ \bar{y}_C \end{bmatrix} \quad \begin{bmatrix} 2.0 \\ 1.0 \\ 0.1 \end{bmatrix}$$

We want 0~1

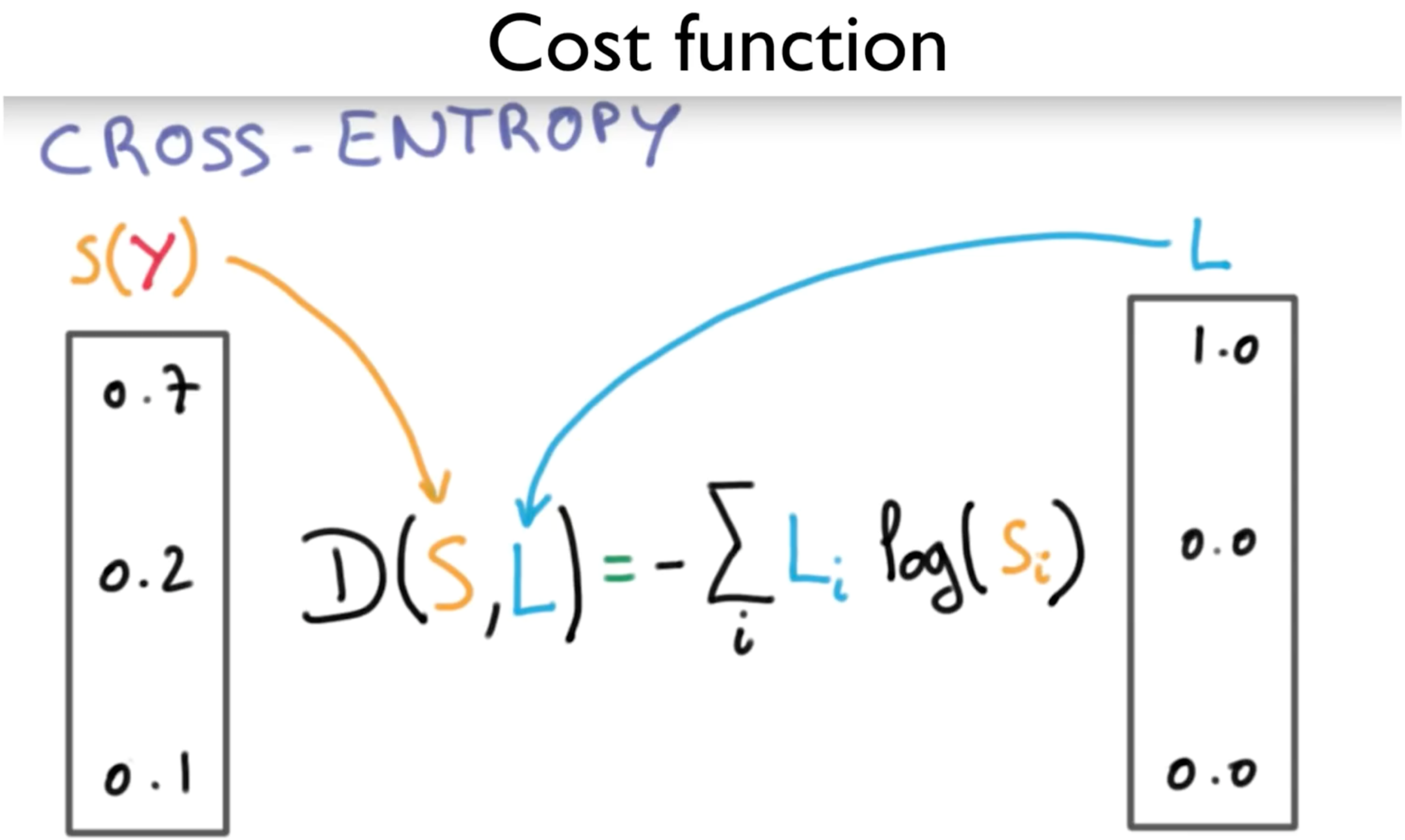
Sigmoid



One Hot encoding(argmax)

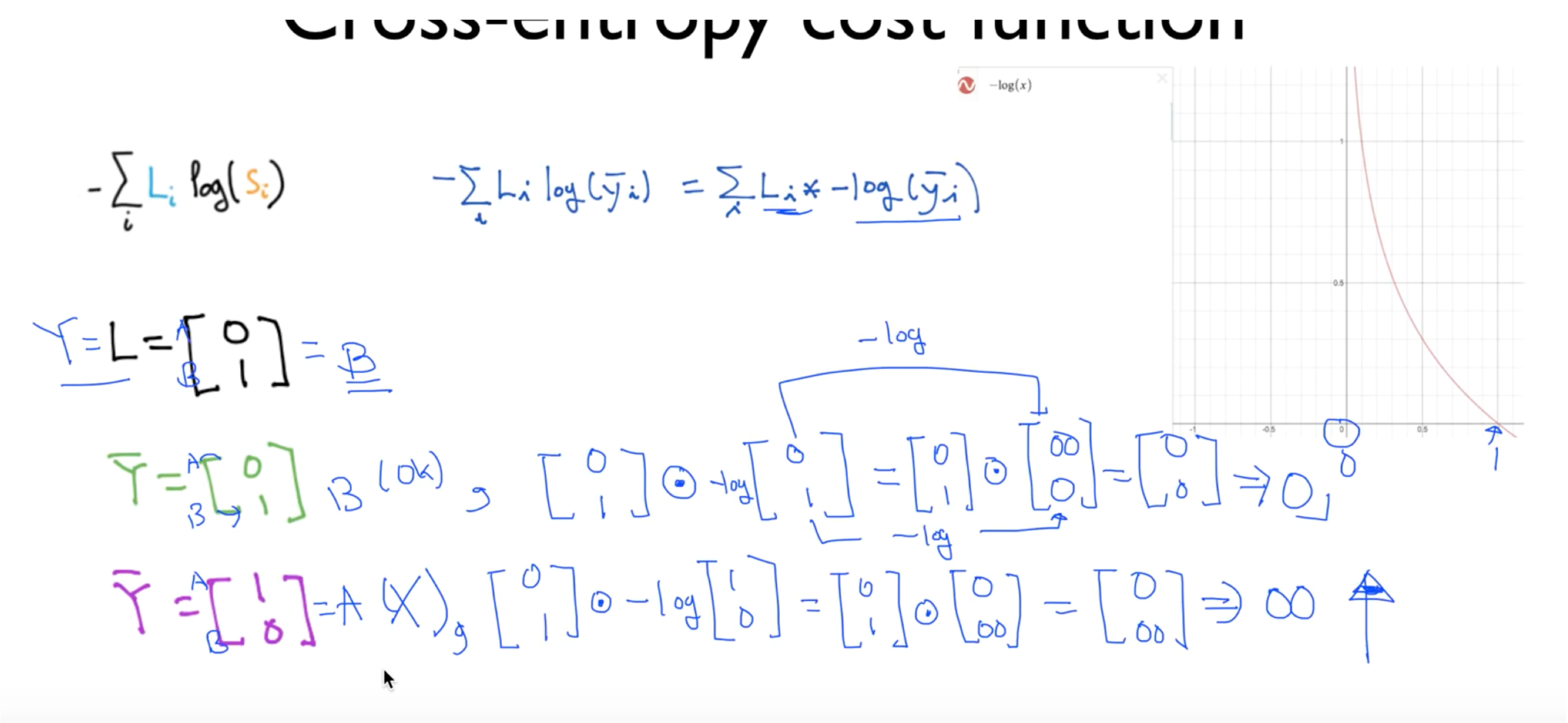


Cost function(cross entropy)



L = real S = expect

Cross entropy



Logistic cost VS Cross entropy

$$C(H(x), y) = ylog(H(x)) - (1-y)log(1-H(x))$$

$$D(S, L) = -\sum_i L_i \log(S_i)$$

Gradient Descent