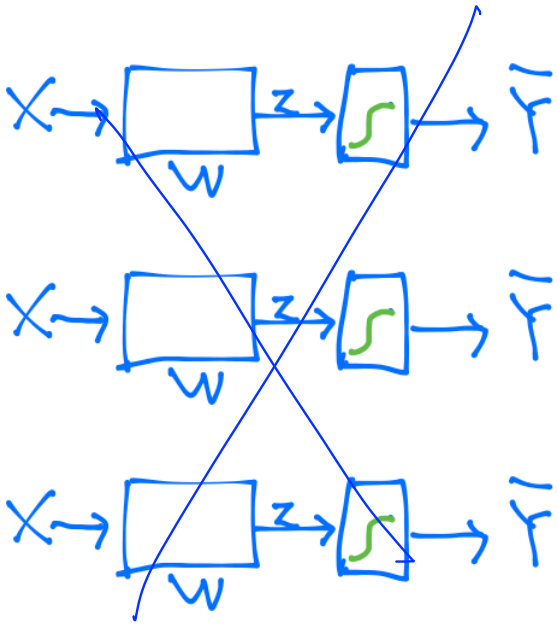
Lecture 6-2

Softmax classification: softmax and cost function

Sung Kim <hunkim+mr@gmail.com>

Where is sigmoid?

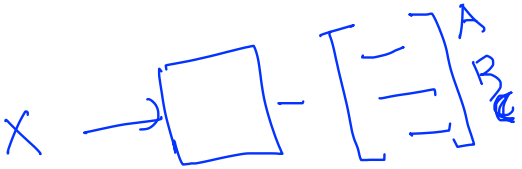
$$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \\ w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \\ w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix} = \begin{bmatrix} y_A \\ y_B \\ y_C \end{bmatrix}$$



Where is sigmoid?

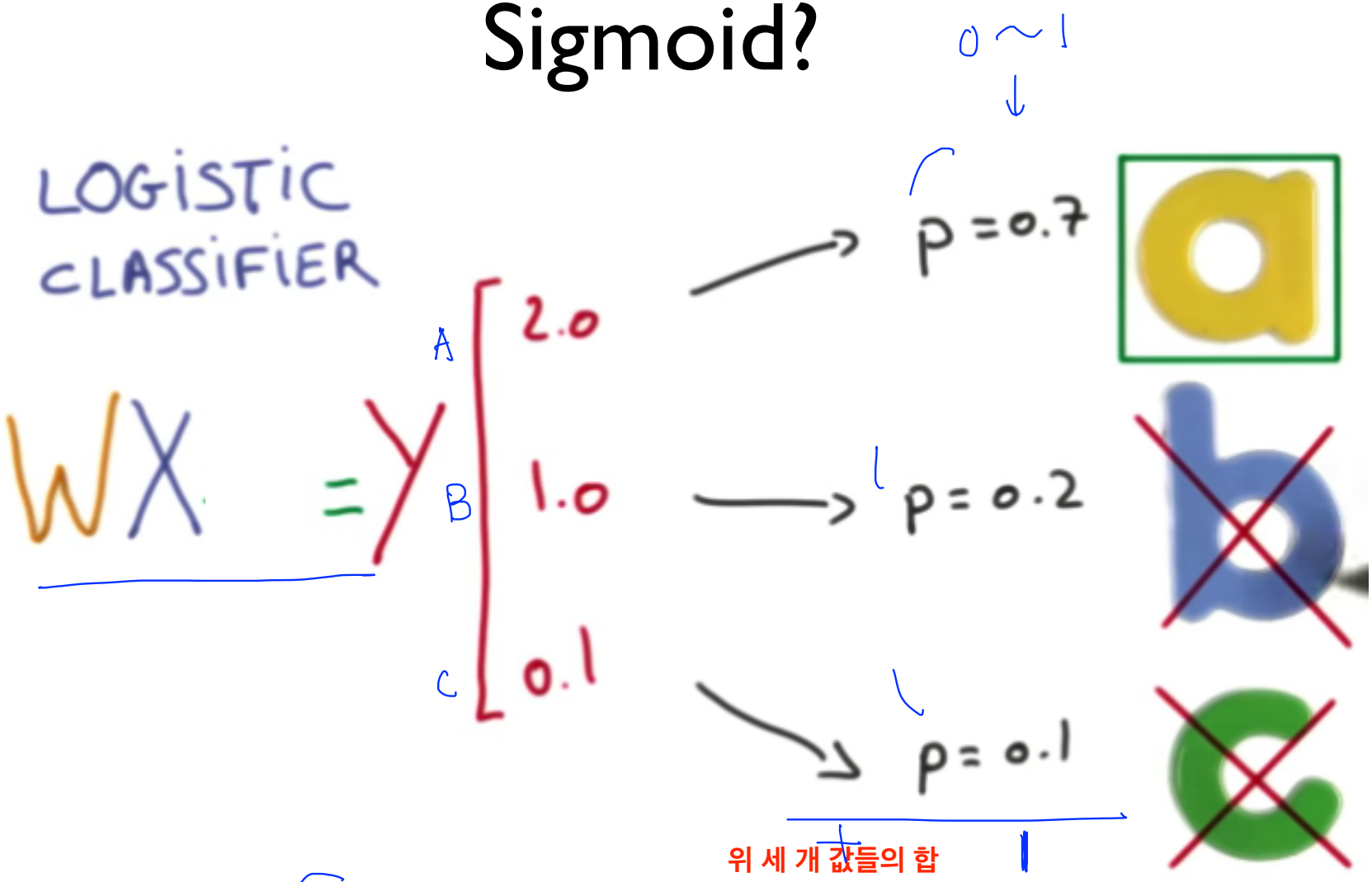
$$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \\ w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \\ w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix} = \begin{bmatrix} y_A \\ y_B \\ y_C \end{bmatrix} \begin{bmatrix} 2.0 \\ 1.0 \\ 0.1 \end{bmatrix}$$

0~1

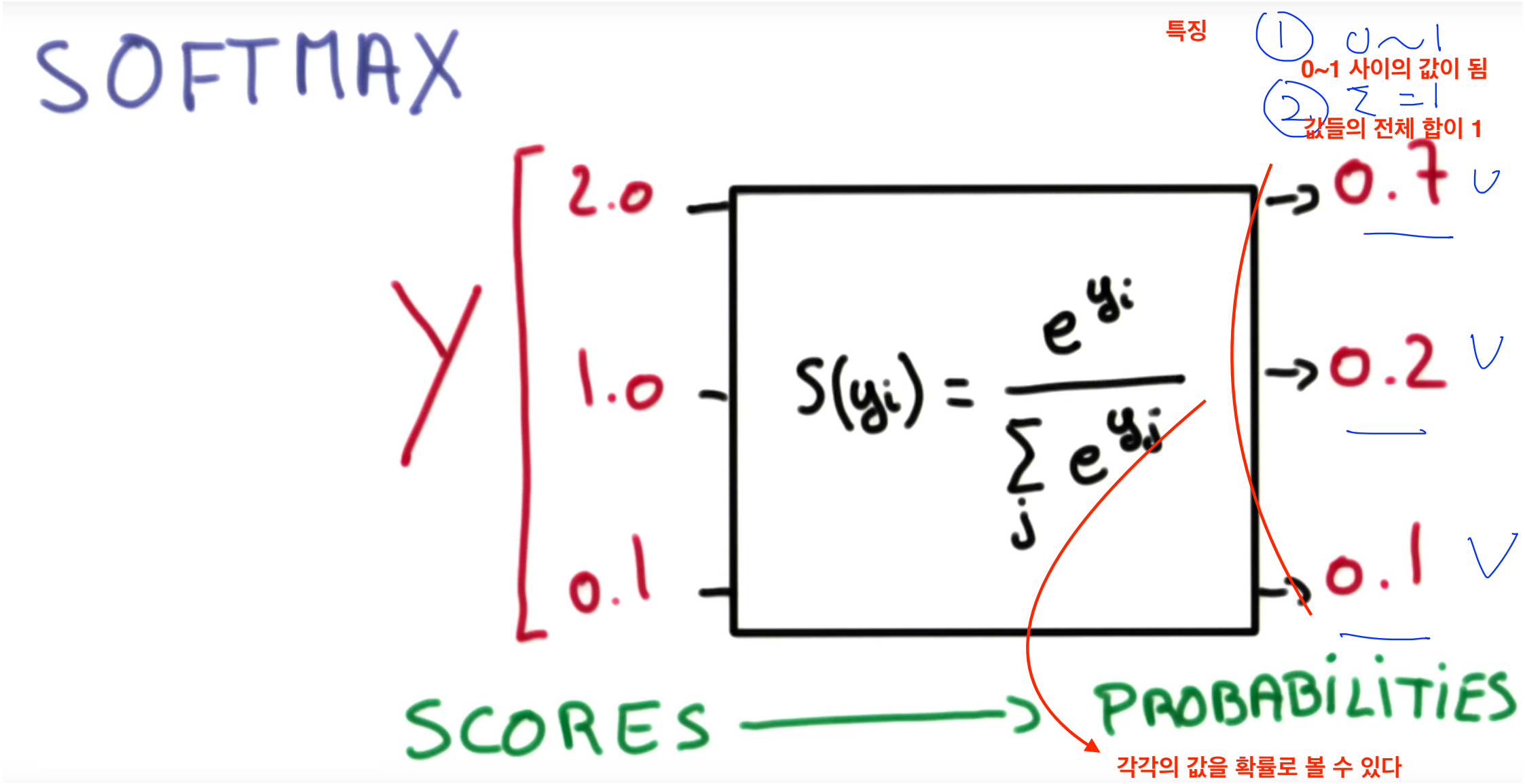


sigmod란 여기 있는 값들이 0~1 사이의 값들로 나왔으면 해서 사용하는 것

Sigmoid?



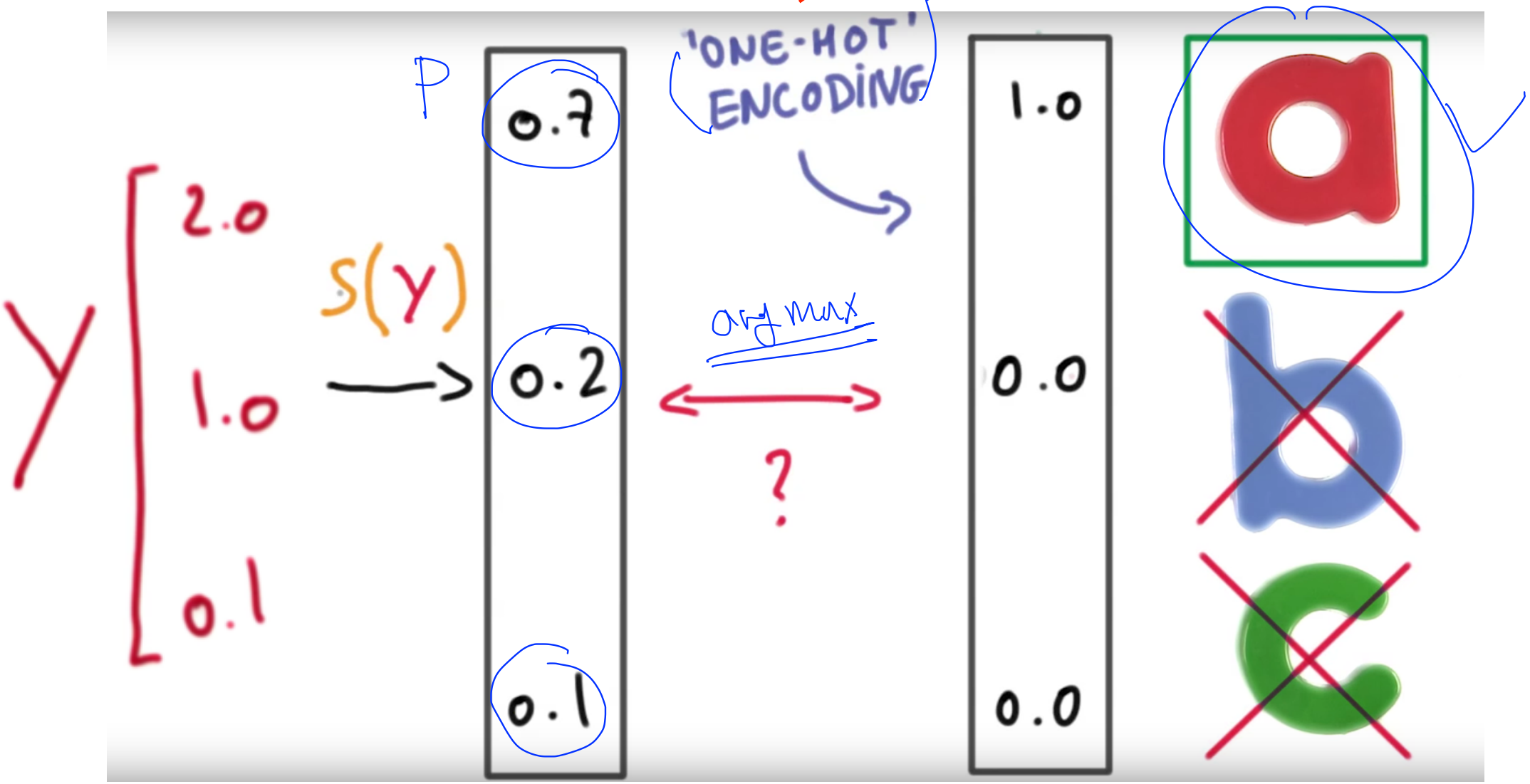
https://www.udacity.com/course/viewer#!/c-ud730/l-6370362152/m-6379811817



https://www.udacity.com/course/viewer#!/c-ud730/l-6370362152/m-6379811817

이 값들을 확률로 볼 수 있다
그래서 a가 나올 확률은
0.7, b는 0.2, c는 0.1

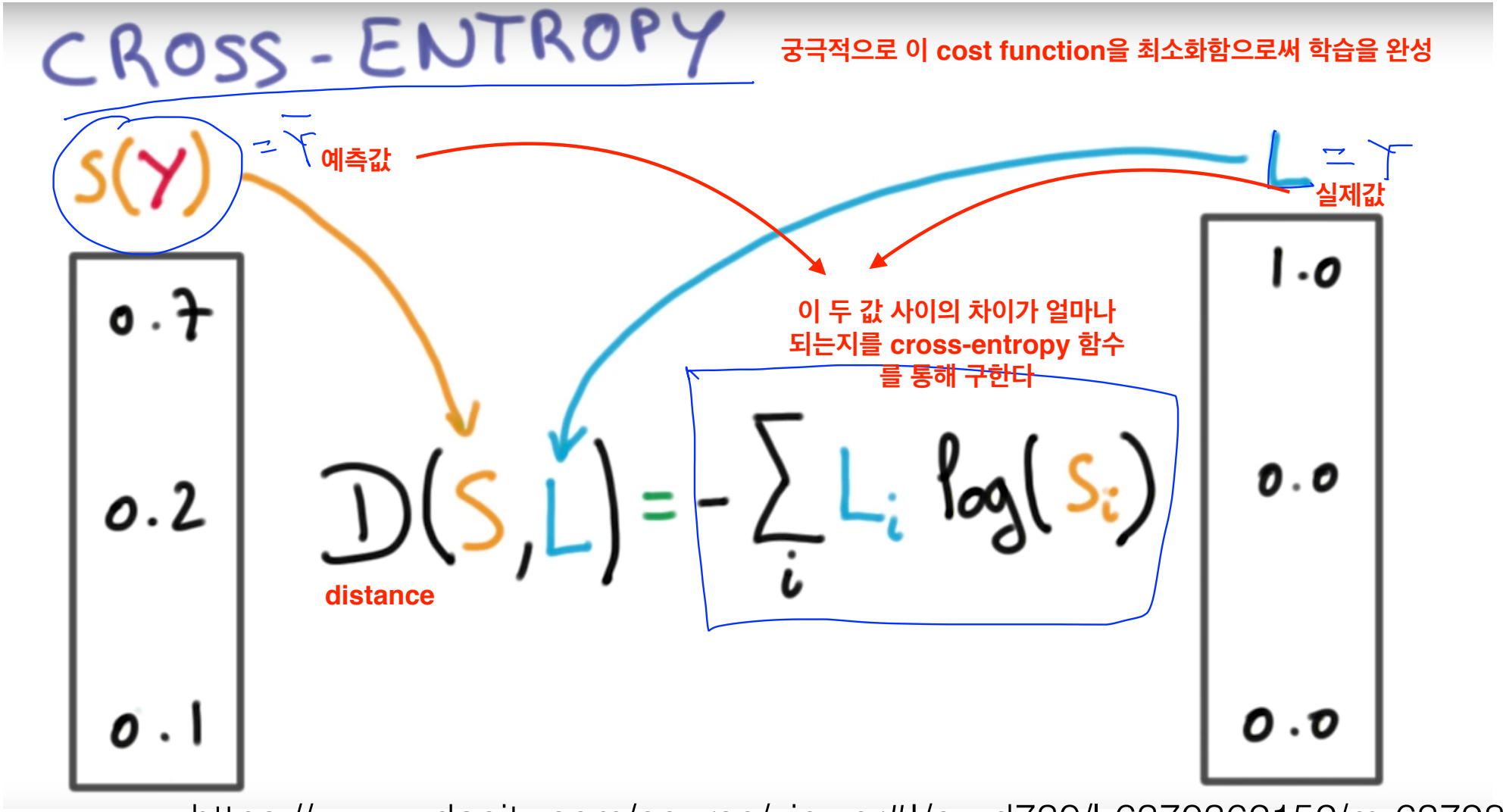
가장 큰 값만 1로 하고 나머지는 0으로



https://www.udacity.com/course/viewer#!/c-ud730/l-6370362152/m-6379811817

예측한 값과 실제의 값이 얼마나 차이나는지 나타내는

Cost function

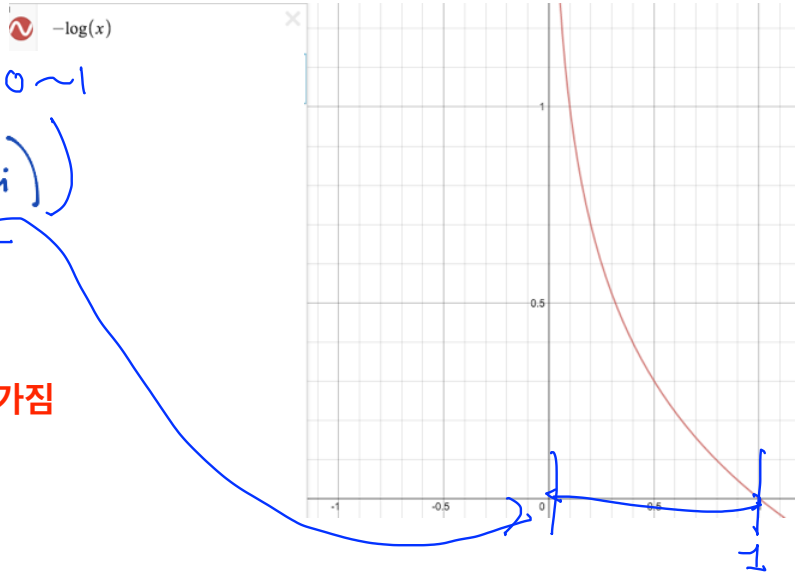


https://www.udacity.com/course/viewer#!/c-ud730/l-6370362152/m-6379811817

Cross-entropy cost function

$$-\sum_i L_i \log(S_i) = -\sum_i L_i \log(\bar{y}_i) = \sum_i (L_i) * (-\log(\bar{y}_i))$$

0 ~ 1



softmax를 통과한 값이라 항상 0에서 1사이의 값을 가짐

Cross-entropy cost function

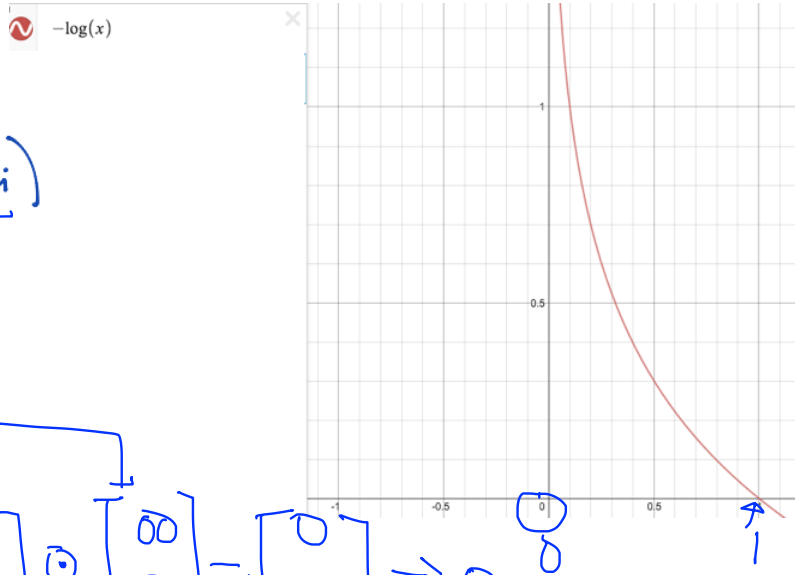
matrix의 곱이 아니라 element의 곱

$$-\sum_i L_i \log(S_i) \qquad -\sum_i L_i \log(\bar{y}_i) = \sum_i L_i * -\log(\bar{y}_i)$$

[0
1]
의 의미는 'B의 값을 갖고 있다.'

실제의 y값 (정답)

$Y = L = A\begin{bmatrix} 0 \\ 1 \end{bmatrix} = B$



예측 1. $\bar{Y} = A\begin{bmatrix} 0 \\ 1 \end{bmatrix} \; B$ (OK)

맞은 예측

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot -\log\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{bmatrix} 00 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow 0$$

$\bar{Y} = A\begin{bmatrix} 1 \\ 0 \end{bmatrix} = A \; (X) \; B$

틀린 예측

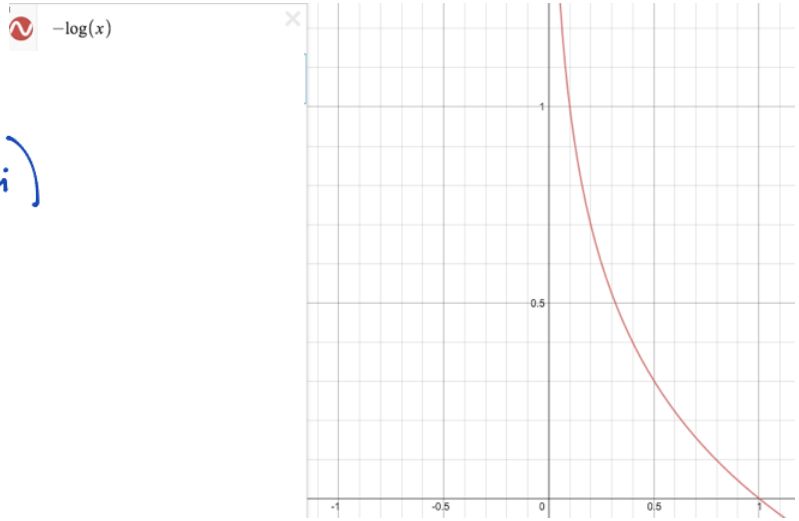
$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot -\log\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{bmatrix} 0 \\ 00 \end{bmatrix} = \begin{bmatrix} 0 \\ 00 \end{bmatrix} \Rightarrow \infty$$

cost function의 의미는 맞은 예측의 경우
작게 되고 틀린 경우 크게 되는 것인데 되는
지 확인하는 과정

Cross-entropy cost function

$$-\sum_i L_i \log(S_i)$$

$$-\sum_i L_i \log(\bar{y}_i) = \sum_i L_i * -\log(\bar{y}_i)$$



$L = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = A$

A의 값을 갖고 있다

$\bar{Y} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = A$ (O) $\begin{bmatrix} 1 \\ 0 \end{bmatrix} \odot \begin{bmatrix} 0 \\ 00 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow 0$

$\bar{Y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = B$ (X) $\begin{bmatrix} 1 \\ 0 \end{bmatrix} \odot \begin{bmatrix} 00 \\ 0 \end{bmatrix} = \begin{bmatrix} 00 \\ 0 \end{bmatrix} \Rightarrow 00$

sum을 취하면

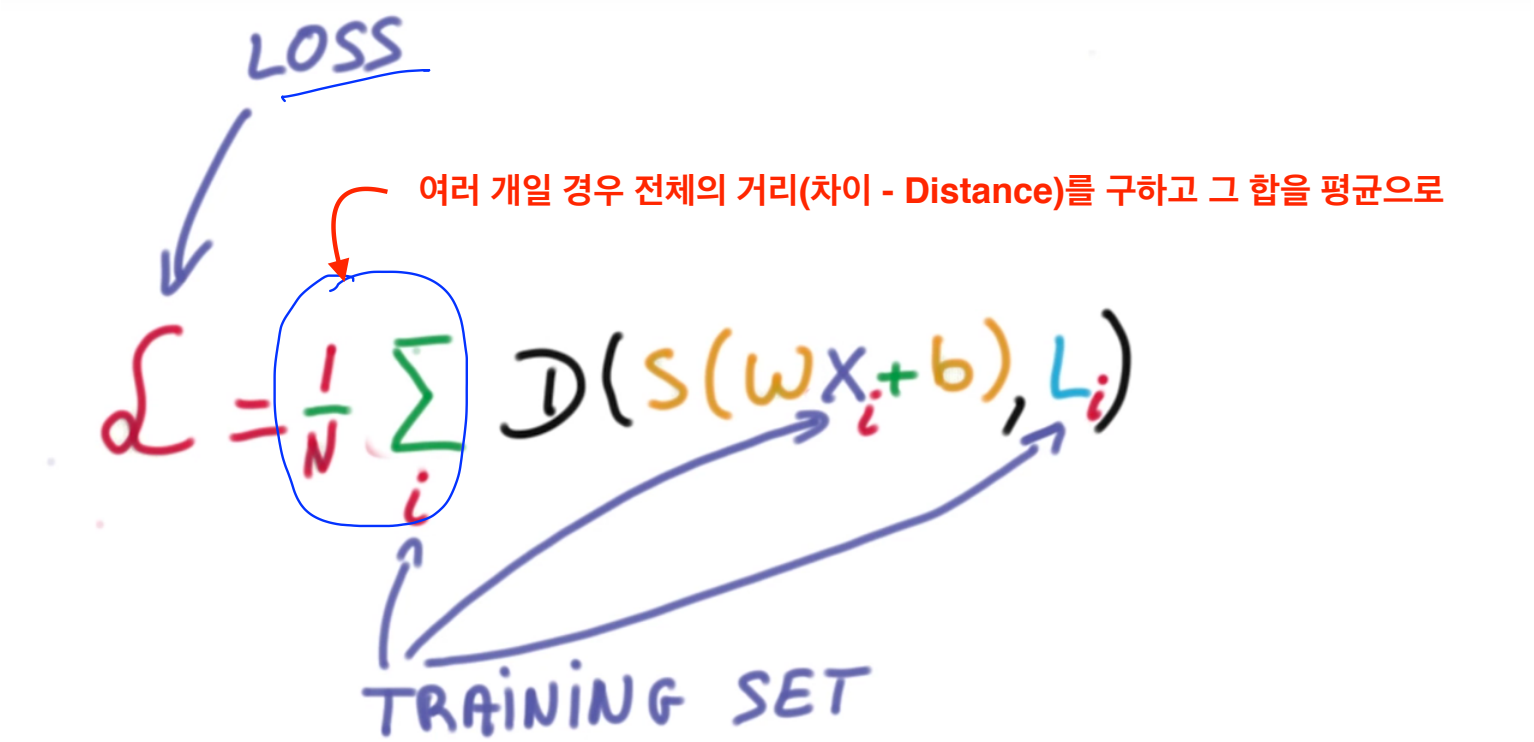
Logistic cost VS cross entropy

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

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

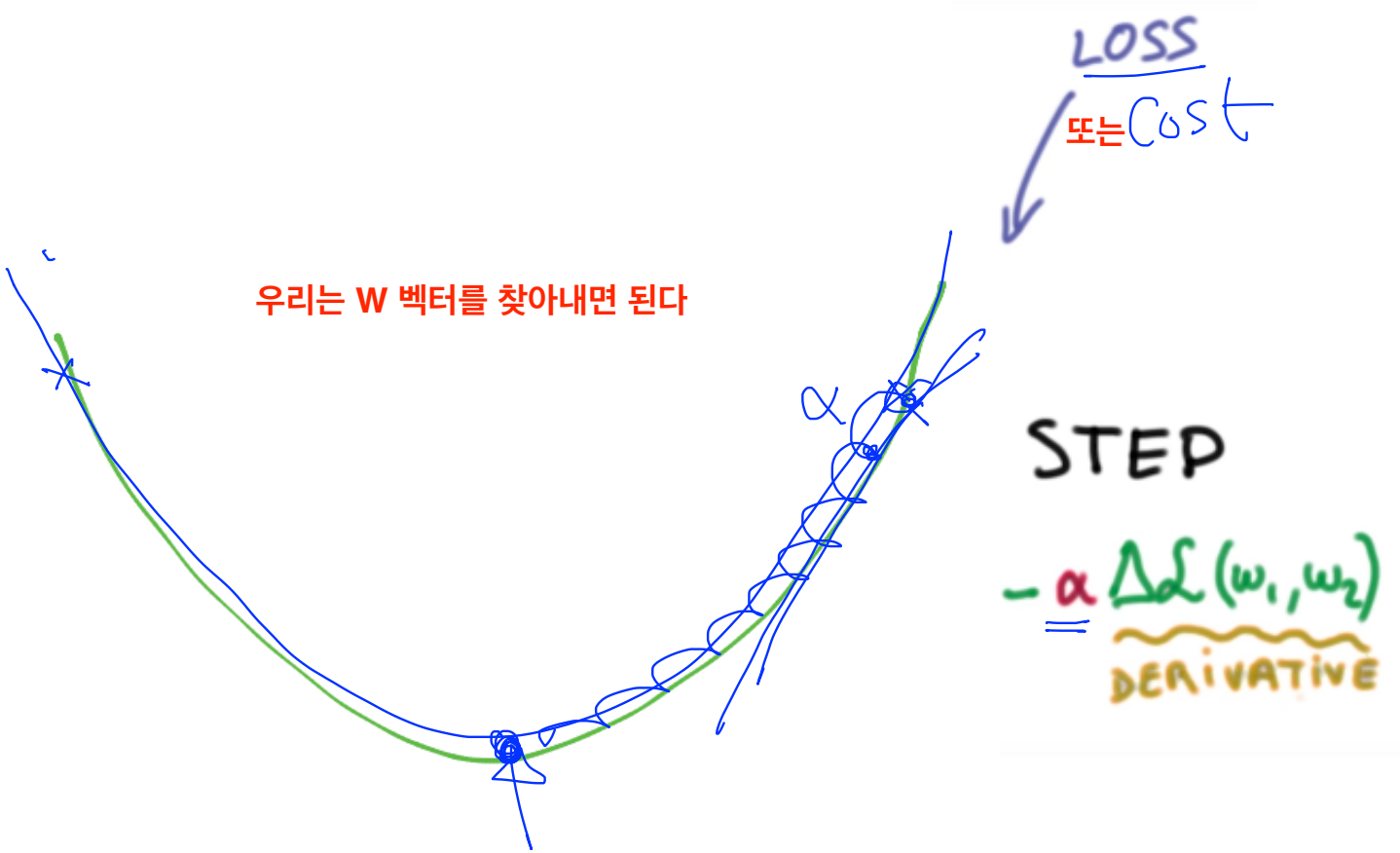
두 가지 방식이 실제로 같은 것을 의미

Cost function



https://www.udacity.com/course/viewer#!/c-ud730/l-6370362152/m-6379811817

Gradient descent



https://www.udacity.com/course/viewer#!/c-ud730/l-6370362152/m-6379811827

Next
Applications & Tips