Lecture 6-1

Softmax classification: Multinomial classification

Sung Kim <hunkim+mr@gmail.com>

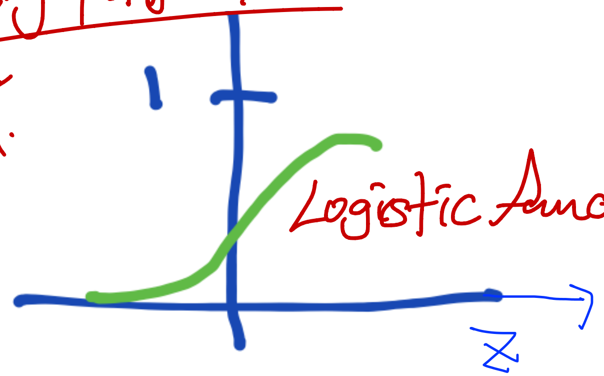
Logistic regression

기본 Base.

$H_L(x) = Wx$ — $\{100, 200, -10\}$ output : 실수.

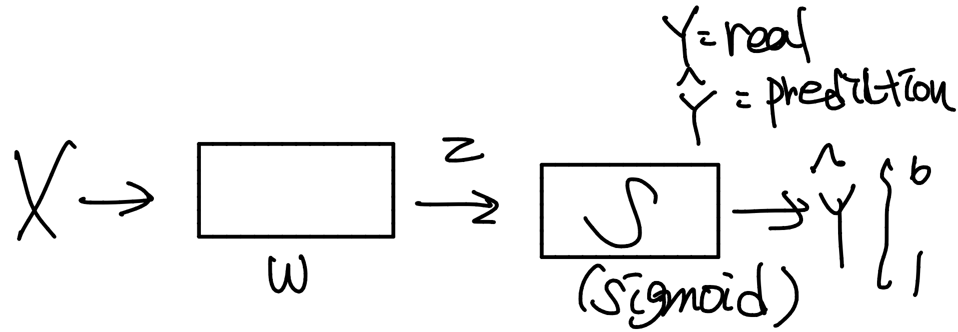
Binary에 적합X.

$Z = H_L(x)$, 대입 $g(z) = \{0, 1\}$

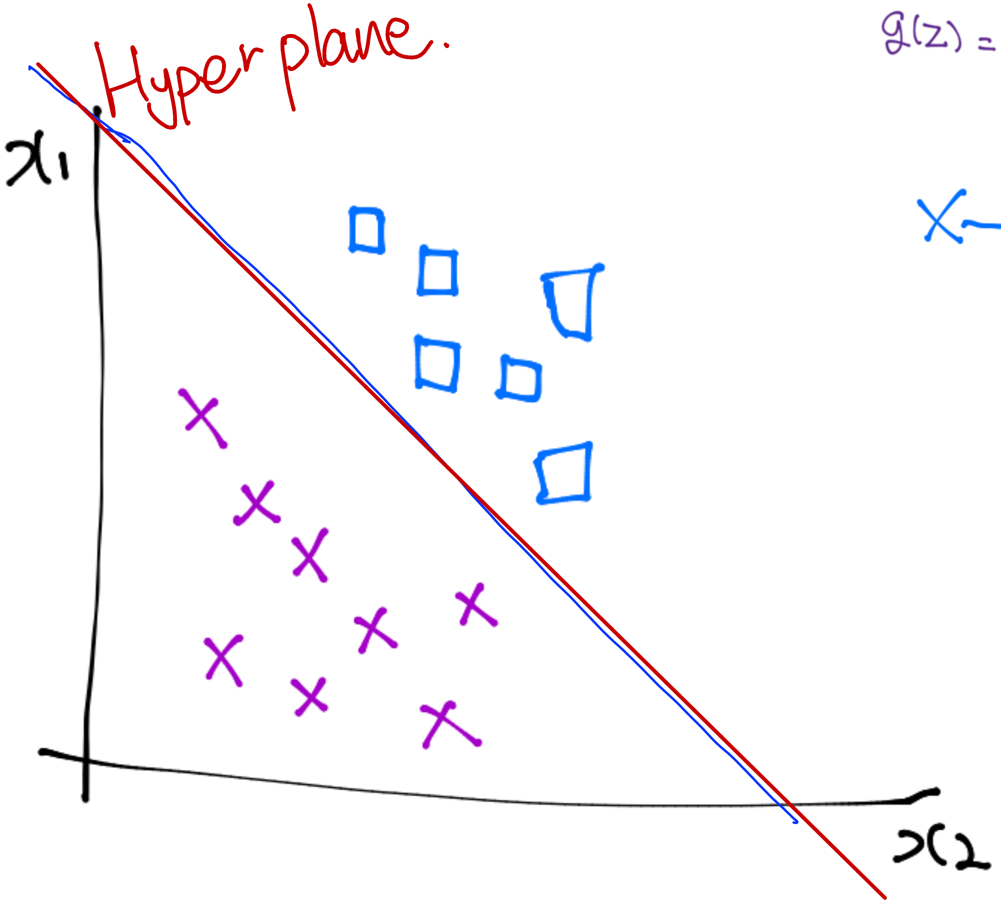


$$g(z) = \frac{1}{1+e^{-z}}$$ (Sigmoid) Logistic.

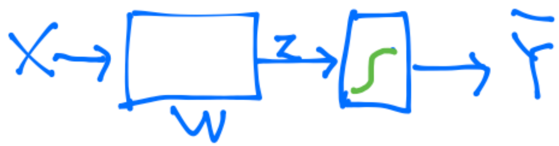
$$H_R(x) = g(H_L(x))$$



Logistic regression



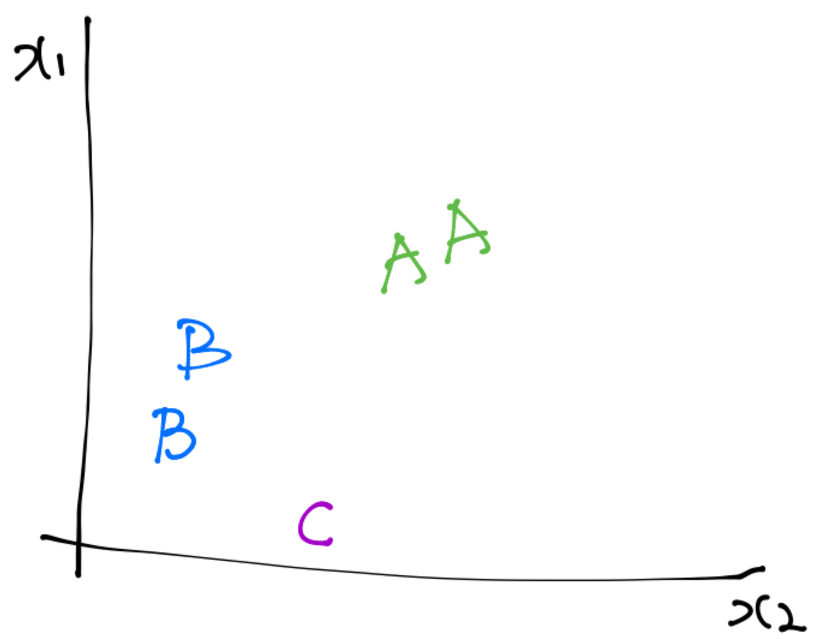
$$g(z) = \frac{1}{1+e^{-z}} \qquad H_R(x) = g(H_L(x))$$



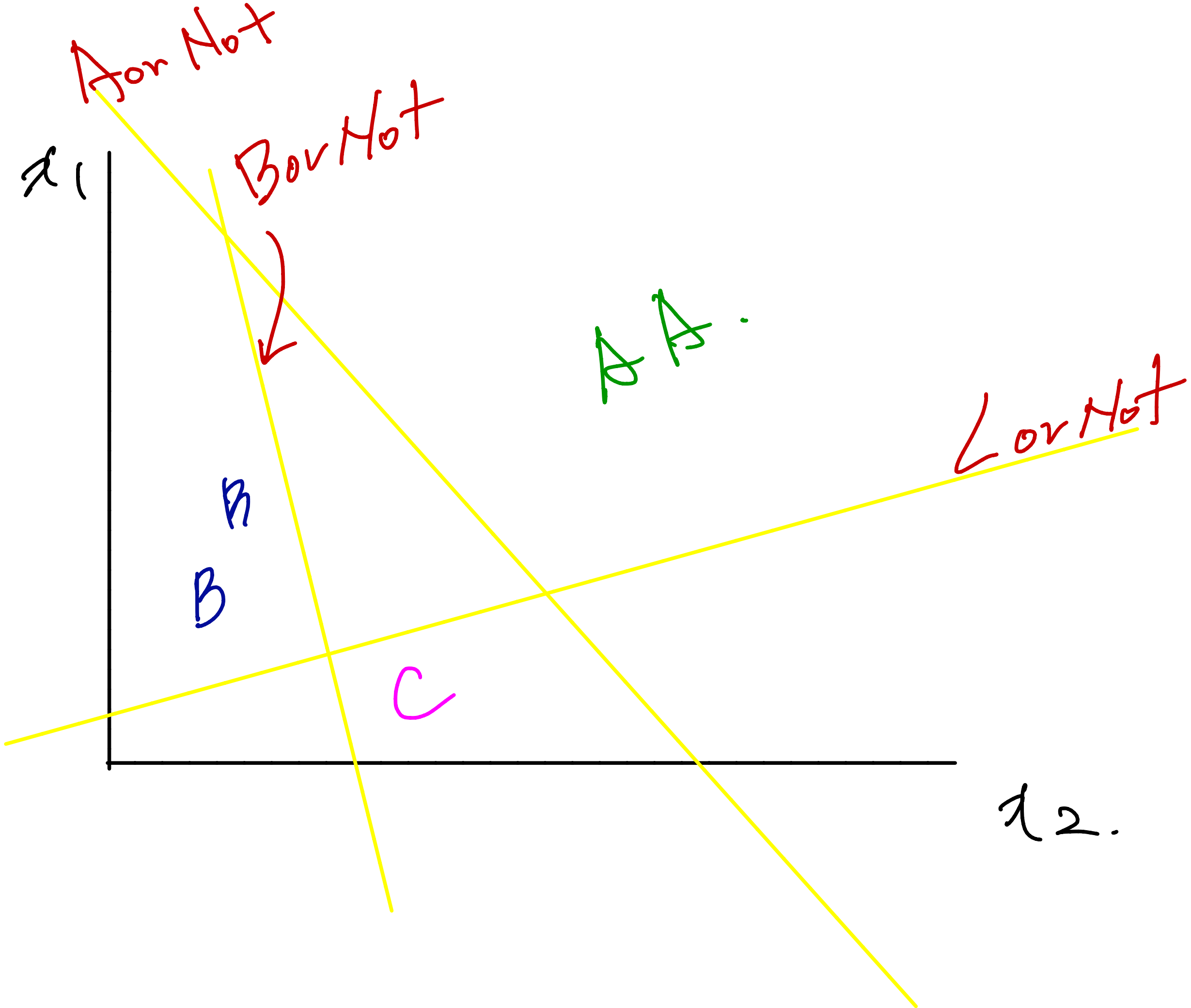
Multinomial classification

(여러가지 class)

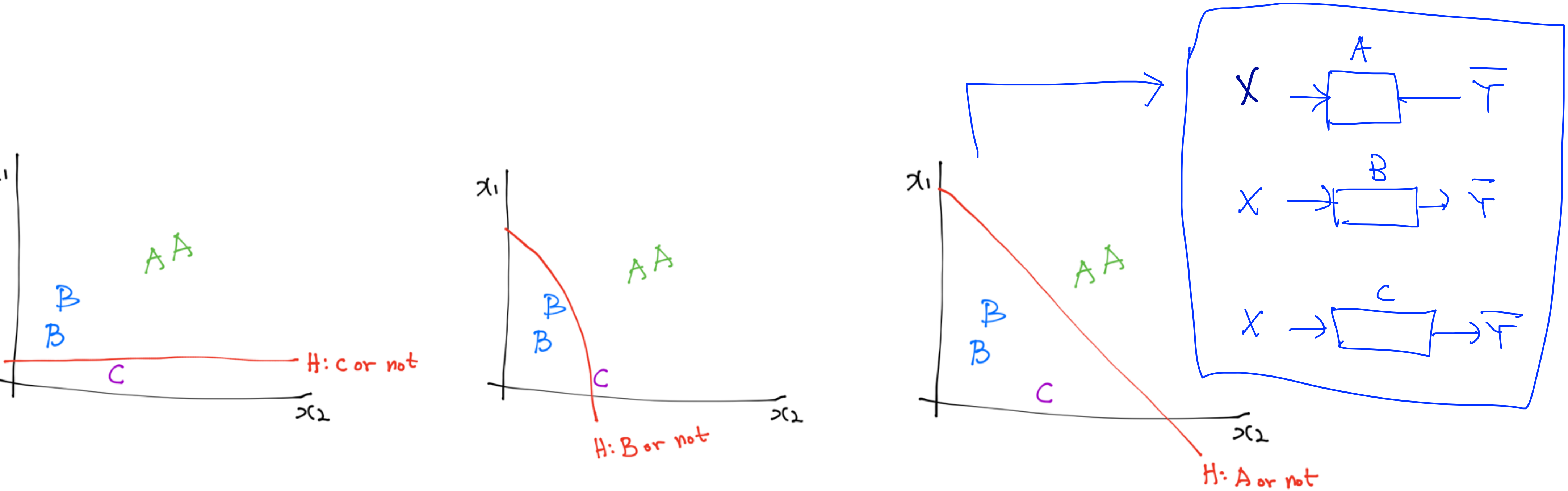
x1 (hours)	x2 (attendance)	y (grade)
10	5	A
9	5	A
3	2	B
2	4	B
11	1	C



Multinomial classification



Multinomial classification



Multinomial classification

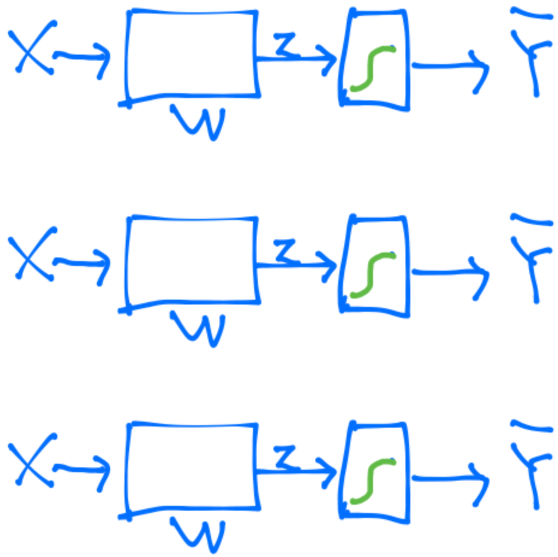
독립 계산도 힘들고 구현도 힘들다.

$$\begin{array}{cc} W & X \end{array} \qquad H(X)$$

$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_1 x_1 + w_2 x_2 + w_3 x_3 \end{bmatrix}$$

$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_1 x_1 + w_2 x_2 + w_3 x_3 \end{bmatrix}$$

$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_1 x_1 + w_2 x_2 + w_3 x_3 \end{bmatrix}$$

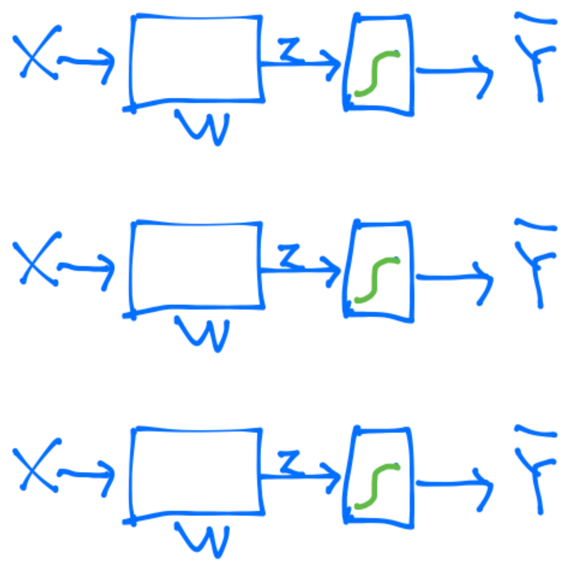


Multinomial classification

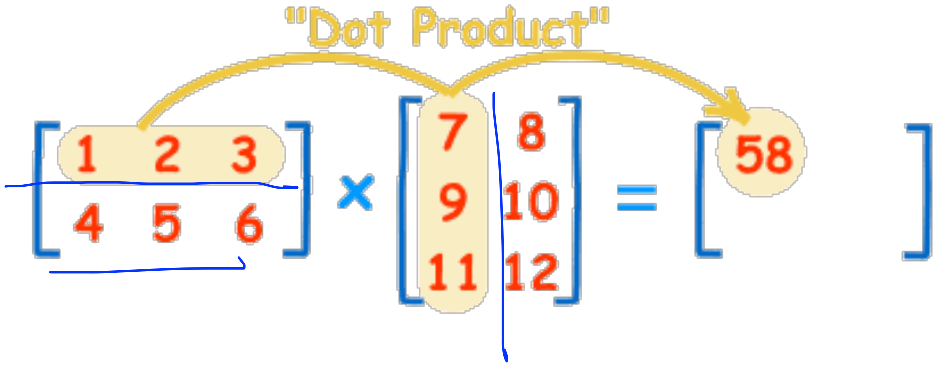
$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_2x_2 + w_3x_3 \end{bmatrix}$$

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

Mat Mul



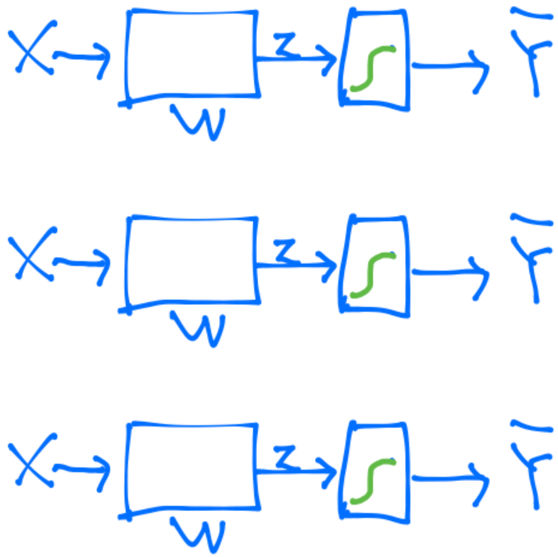
Matrix multiplication



https://www.mathsisfun.com/algebra/matrix-multiplying.html

Multinomial classification

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



Multinomial classification

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

$H_A(x)$

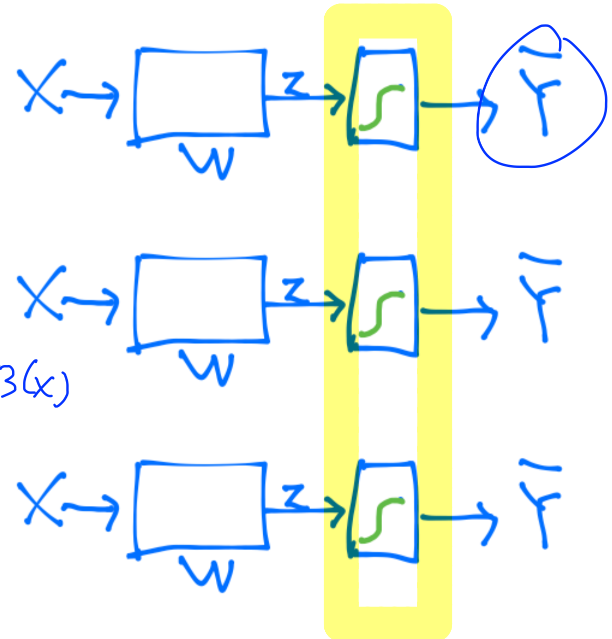
$H_B(x)$

$H_C(x)$

독립적인 형태로

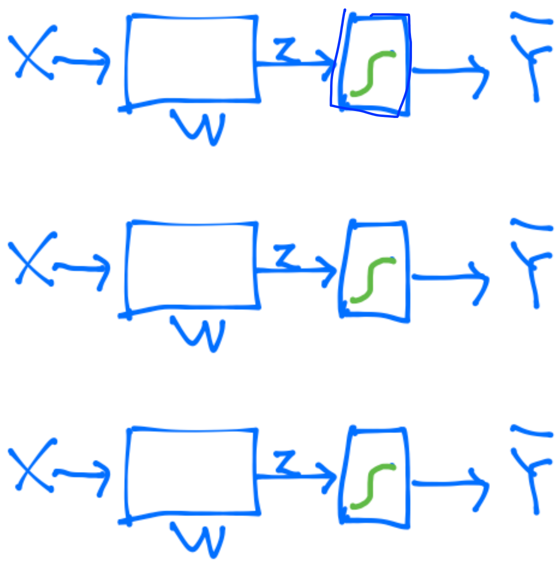
Matrix 계산

⇒ 간단해짐



Where is sigmoid?

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



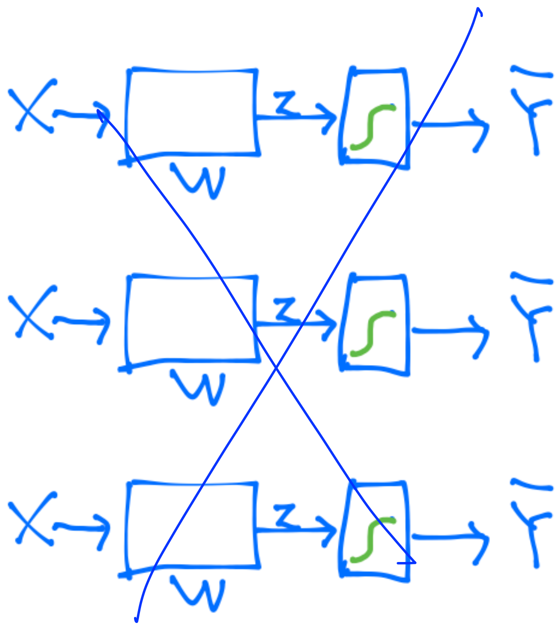
Lecture 6-2

Softmax classification: softmax and cost function

Sung Kim <hunkim+mr@gmail.com>

Where is sigmoid?

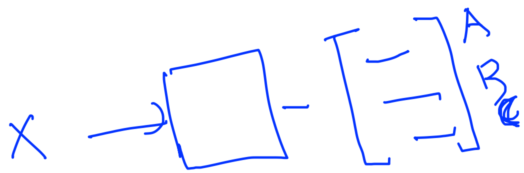
$$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \\ w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \\ w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix}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



Where is sigmoid?

0~1

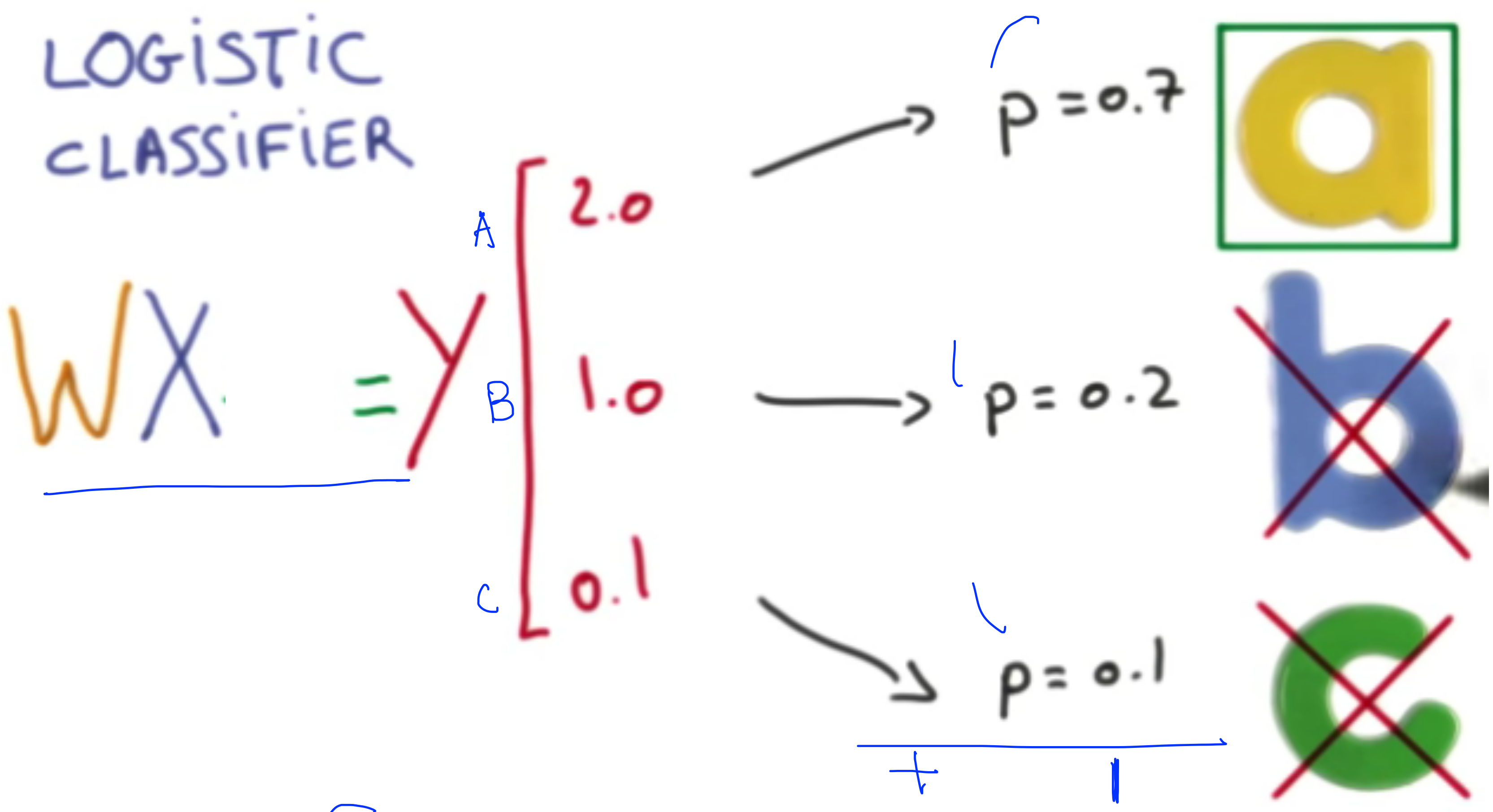
$$\begin{bmatrix} W_{A1} & W_{A2} & W_{A3} \\ W_{B1} & W_{B2} & W_{B3} \\ W_{C1} & W_{C2} & W_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \\ w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \\ w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix}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



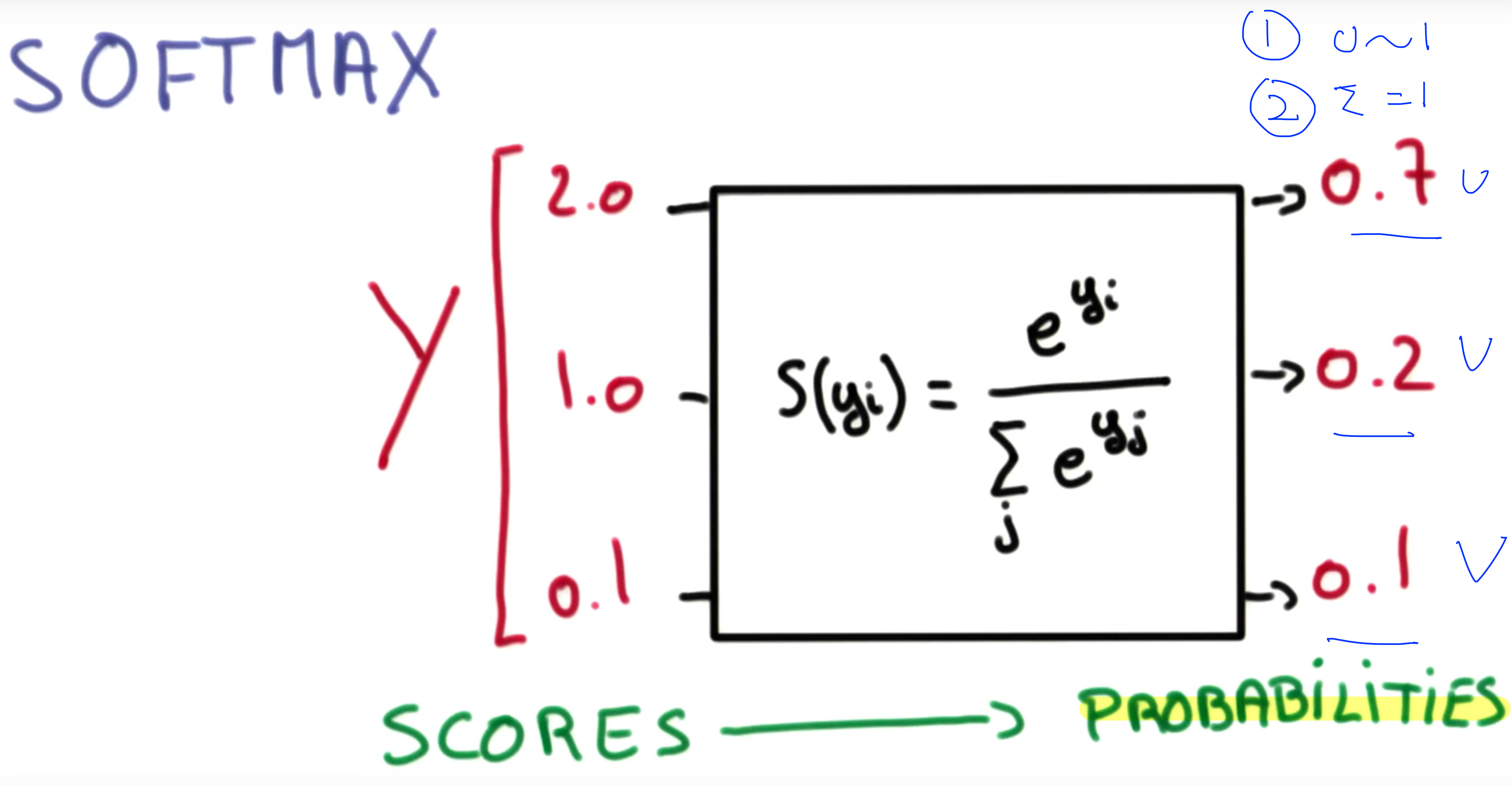
확률로 표현되는 느낌

Sigmoid?

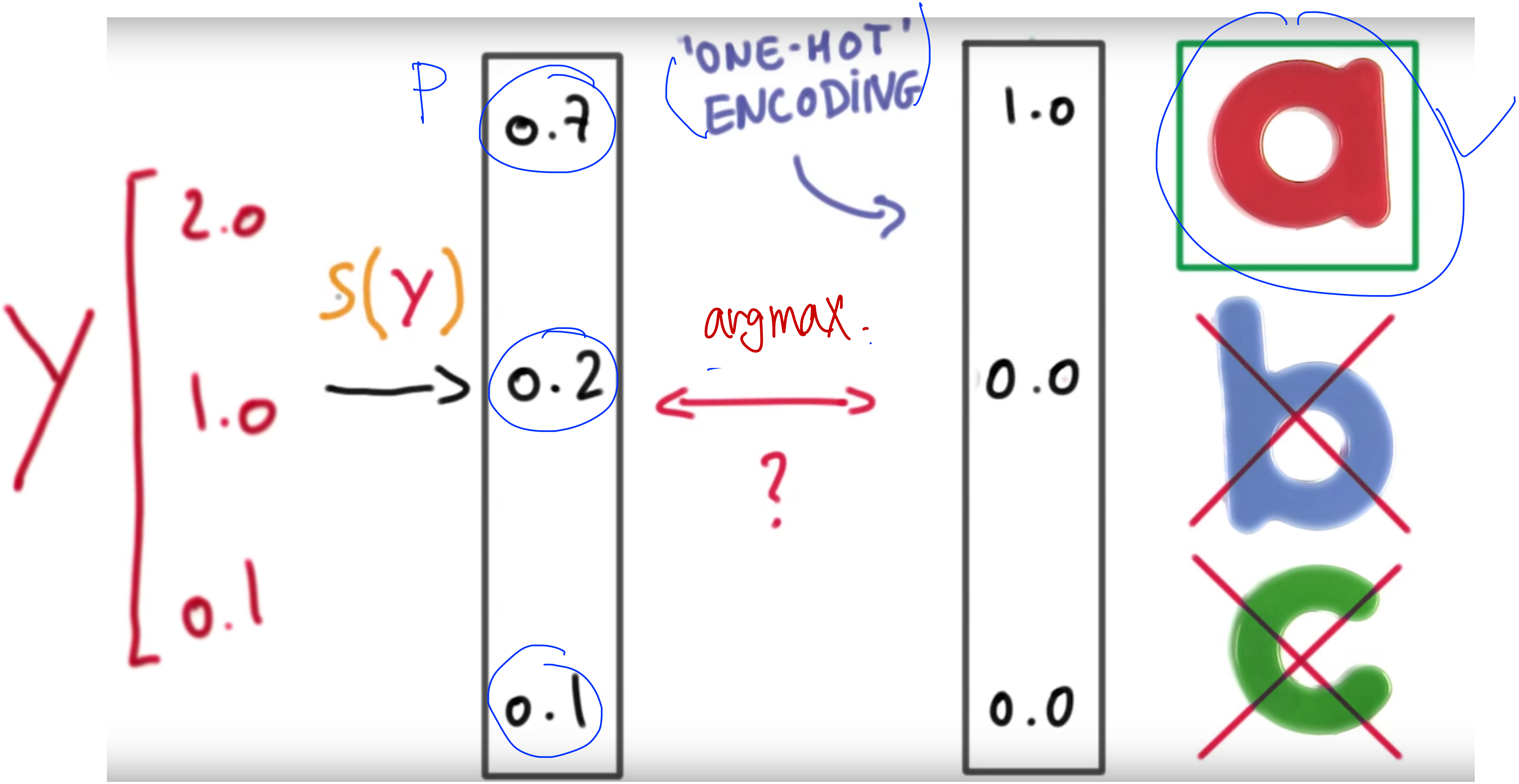
0 ~ 1



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

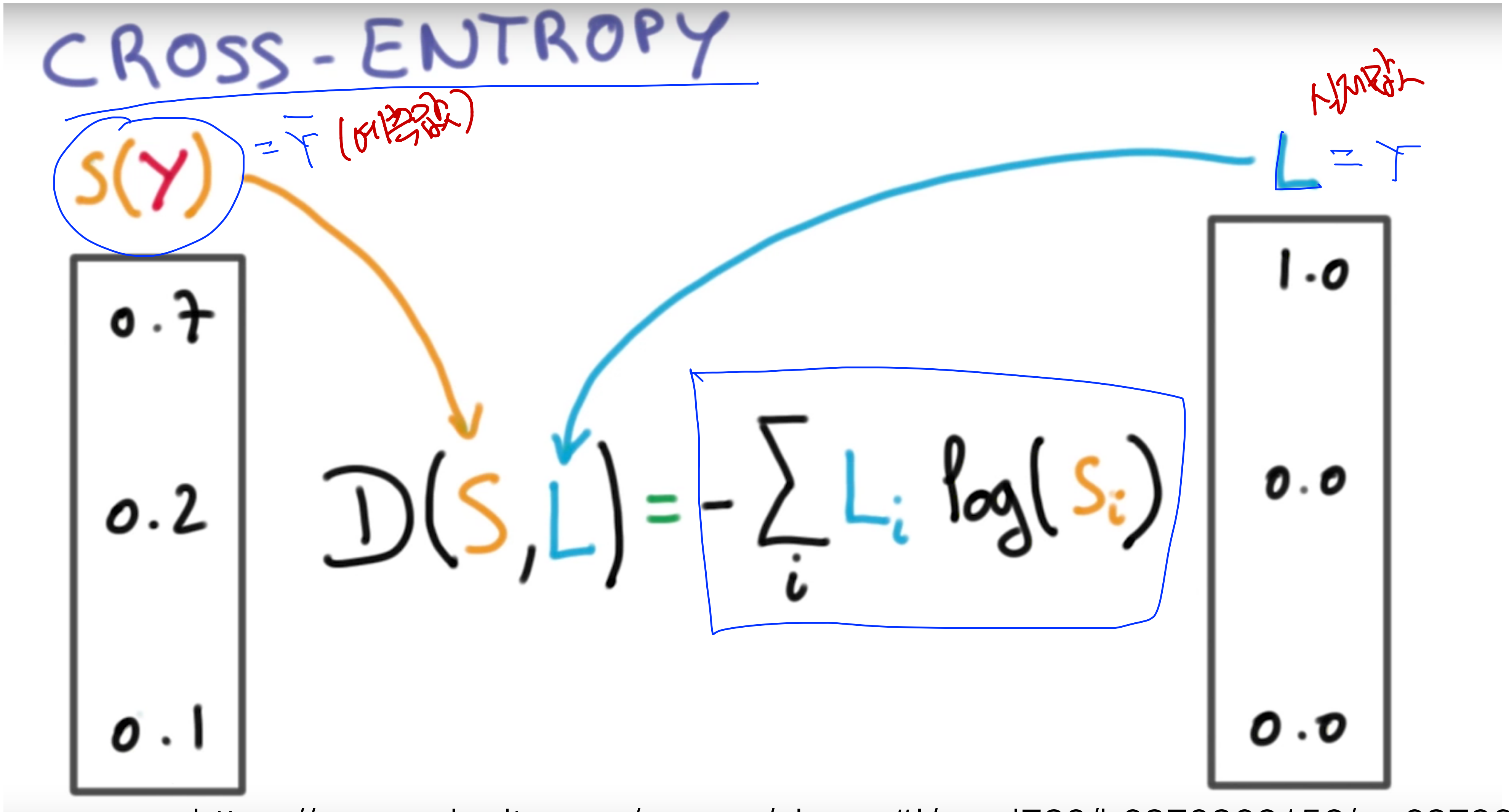


https://www.udacity.com/course/viewer#!/c-ud730/l-6370362152/m-6379811817



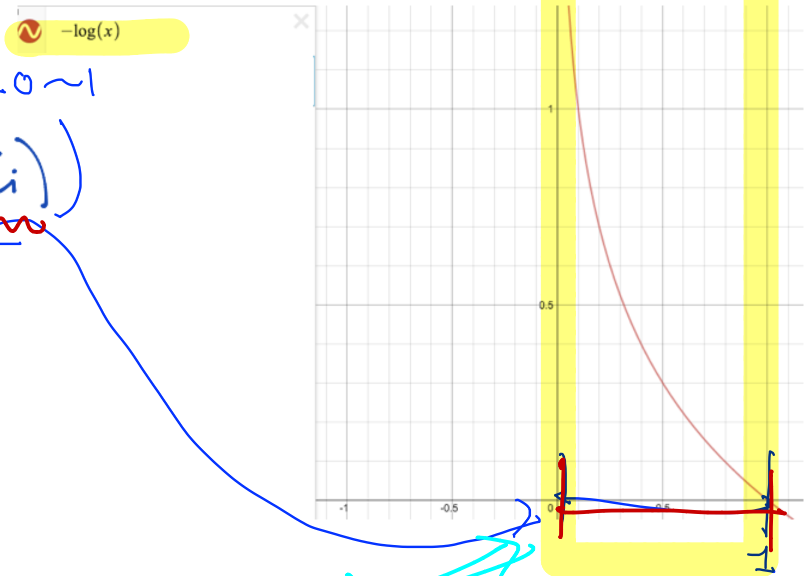
https://www.udacity.com/course/viewer#!/c-ud730/l-6370362152/m-6379811817

Cost function



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

Cross-entropy cost function



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

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

$$\hat{Y} = \begin{array}{c} A \\ B \end{array}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = B \text{ (ok)}$$

경우. cost↓

$$\overset{L}{\begin{bmatrix} 0 \\ 1 \end{bmatrix}} \circledast -\log \overset{\text{예측}}{\begin{bmatrix} 0 \\ 1 \end{bmatrix}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \circledast \begin{bmatrix} \infty \; (0) \\ 0 \; (1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \xRightarrow{\text{sum}} 0$$

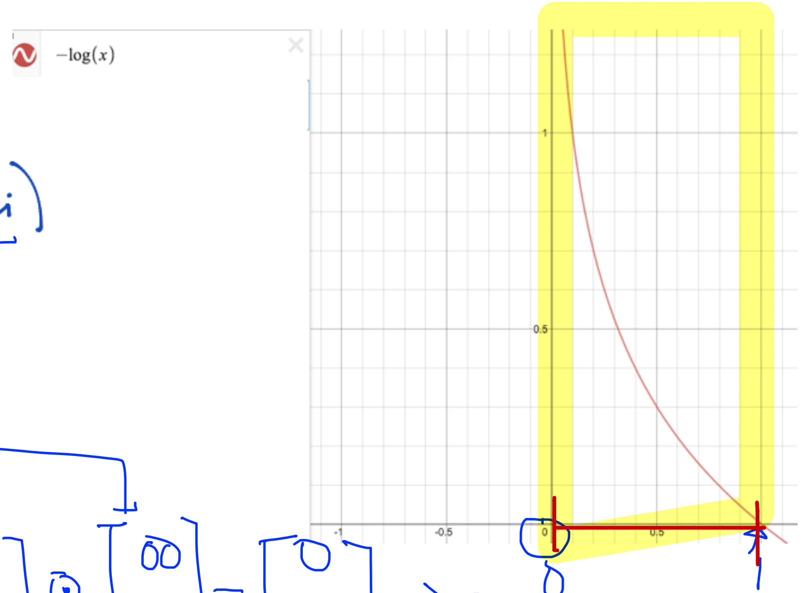
$$\hat{Y} = \begin{array}{c} A \\ B \end{array}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = A(x)$$

cost↑

실제(label)과 어긋나면

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \circledast -\log \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \circledast \begin{bmatrix} 0 \\ \infty \end{bmatrix} = \begin{bmatrix} 0 \\ \infty \end{bmatrix} \xRightarrow{\text{sum}} \infty$$

Cross-entropy cost function



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

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

$$\underline{Y} = L = \begin{array}{c c} A & \\ B & \end{array}\begin{bmatrix}0\\1\end{bmatrix} = \underline{\underline{B}}$$

$$\bar{Y} = \begin{array}{c} A \\ B \end{array}\begin{bmatrix}0\\1\end{bmatrix} \; B \text{ (ok)}, \quad \begin{bmatrix}0\\1\end{bmatrix} \odot -\log\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix} \odot \begin{bmatrix}\infty\\0\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} \Rightarrow 0$$

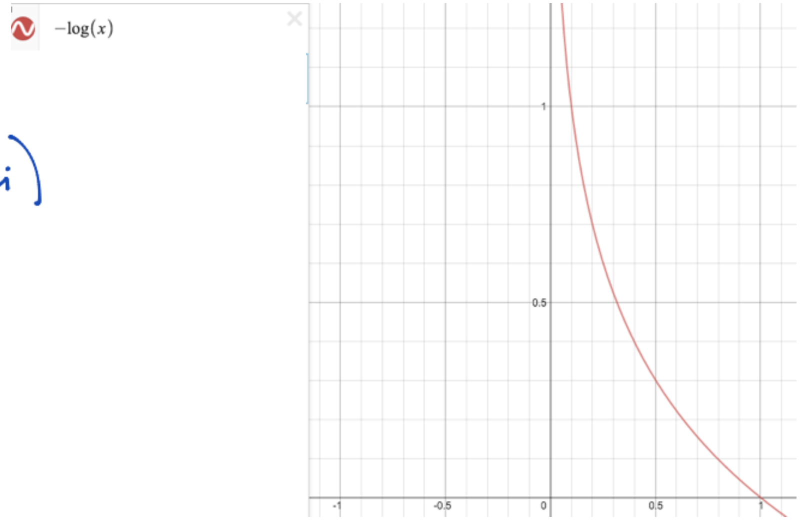
$-\log$ $\quad -\log$

$$\bar{Y} = \begin{array}{c} A \\ B \end{array}\begin{bmatrix}1\\0\end{bmatrix} = A \text{ (X)}, \quad \begin{bmatrix}0\\1\end{bmatrix} \odot -\log\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix} \odot \begin{bmatrix}0\\\infty\end{bmatrix} = \begin{bmatrix}0\\\infty\end{bmatrix} \Rightarrow \infty \uparrow$$

Cross-entropy cost function

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

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



$$L = \begin{array}{c|c} A & \\ B & \end{array}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = A$$

element mul

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

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

Logistic cost VS cross entropy

결과.

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

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

// ?

독리

$S_i = \underline{\underline{Wx_i + b}}$.

$H(x)$

곱하는 0으로 변경.

element mul

$L\begin{bmatrix}0\\1\end{bmatrix} \Rightarrow L_1 \times \log(H_1(x_1)) + L_2 \times \log(H_2(x_2))$

Cost function

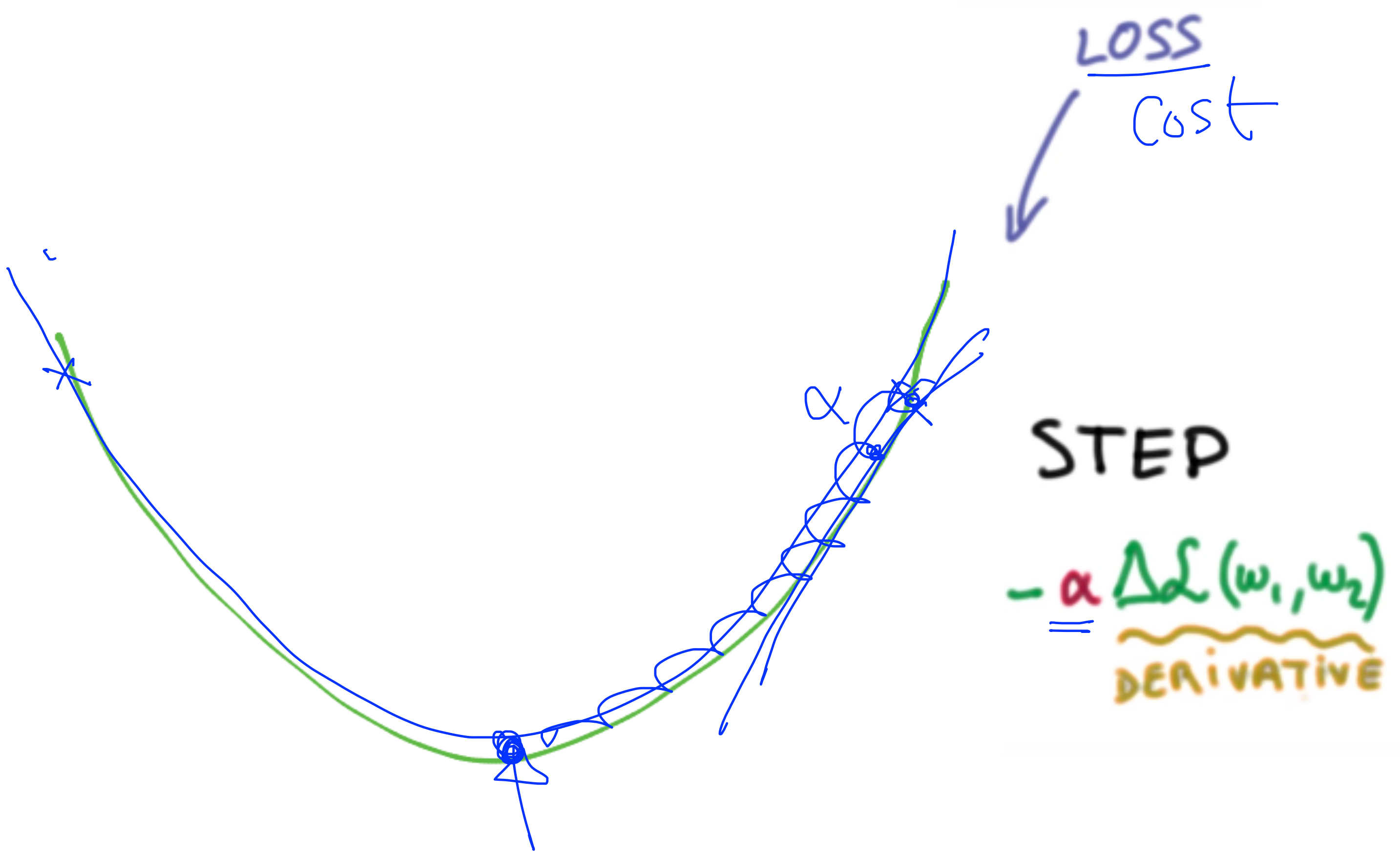
LOSS

$$\mathcal{L} = \frac{1}{N} \sum_i D(S(wx_i + b), L_i)$$

TRAINING SET

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