Lecture 6-1

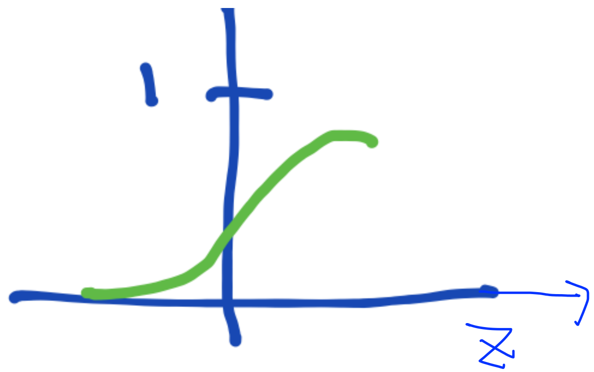
Softmax classification: Multinomial classification

Sung Kim <hunkim+mr@gmail.com>

Logistic regression

$H_L(x) = \underline{Wx} - \left\{ \cancel{\begin{matrix} 100 \\ 200 \\ -10 \end{matrix}} \right\}$

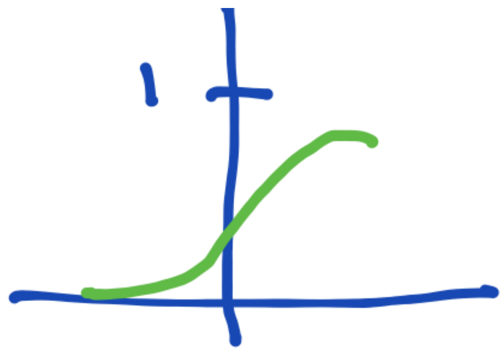
$z = H_L(x), \quad \underbrace{g(z)}$ $= \left\{ \begin{matrix} 0 \\ 1 \end{matrix} \right.$



Logistic regression

$$H_L(x) = Wx$$

$$z = H_L(x), \quad g(z)$$



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

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

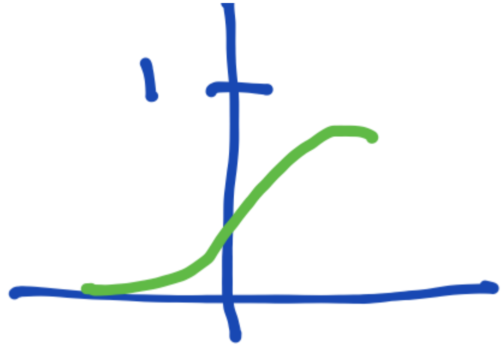
Logistic regression

$$H_L(X) = WX$$

$$z = H_L(X), \quad g(z)$$

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

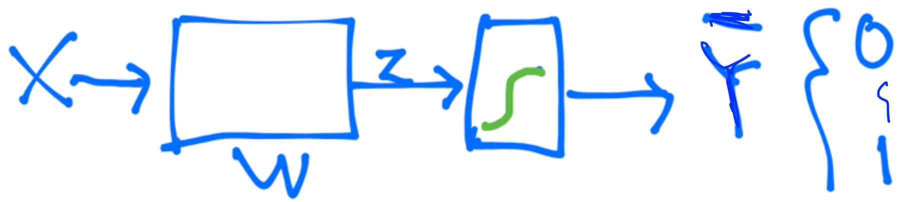
$$H_R(X) = g(H_L(X))$$



Y : real

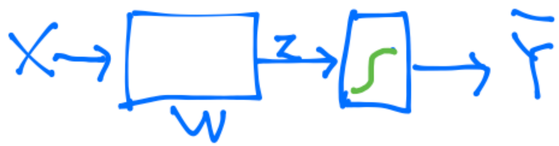
$\bar{Y}$: prediction

$H(x)$



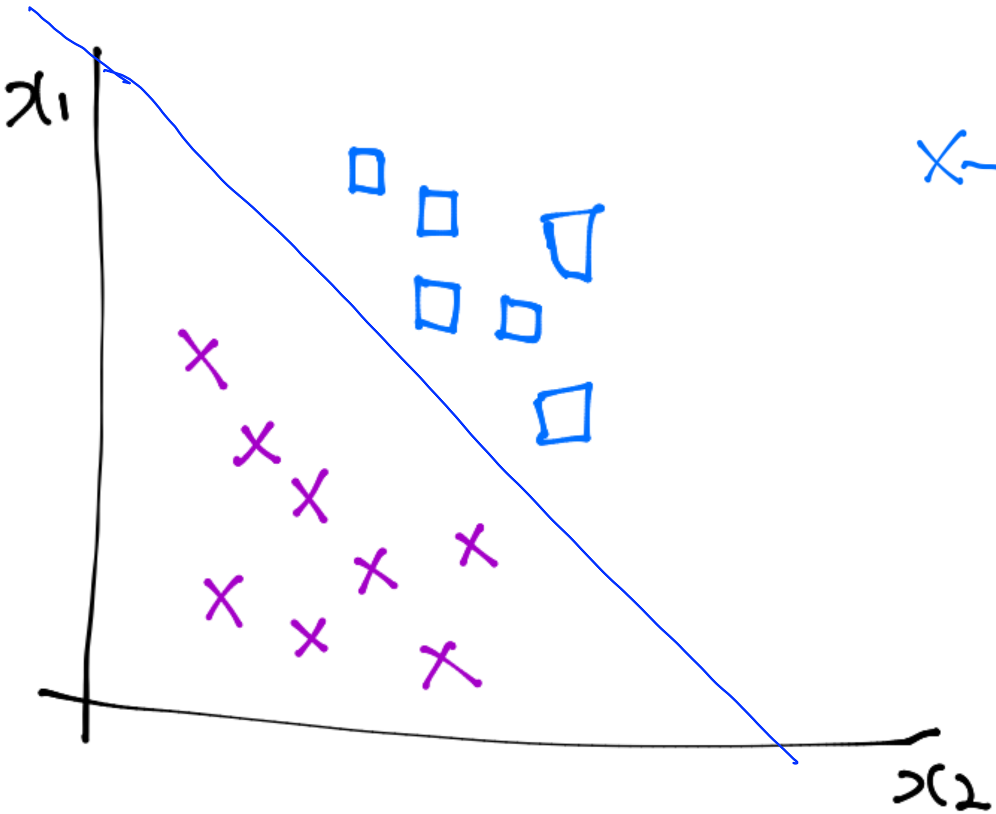
Logistic regression

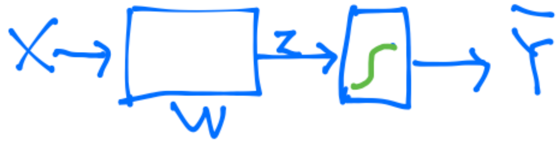
$$g(z) = \frac{1}{1+e^{-z}} \qquad H_R(x) = g(H_L(x))$$



Logistic regression

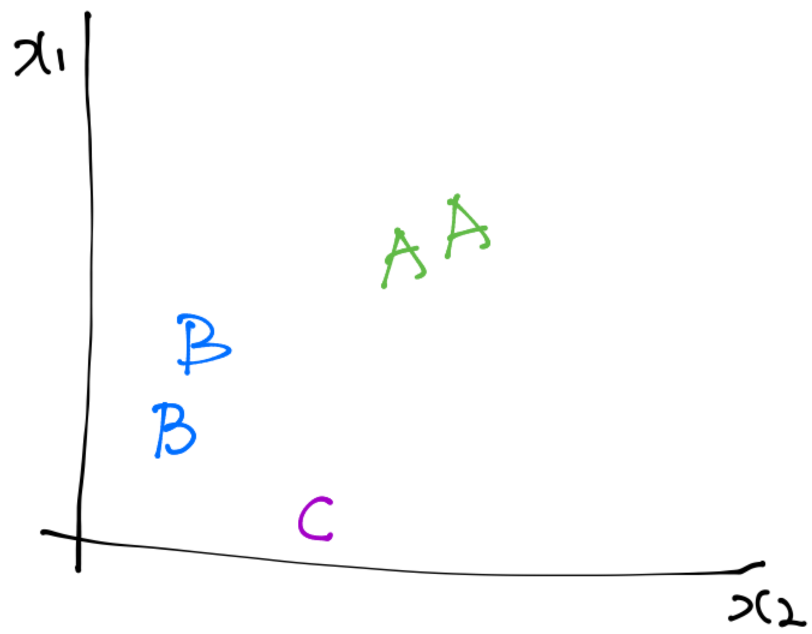
$$g(z) = \frac{1}{1+e^{-z}} \qquad H_R(x) = g(H_L(x))$$



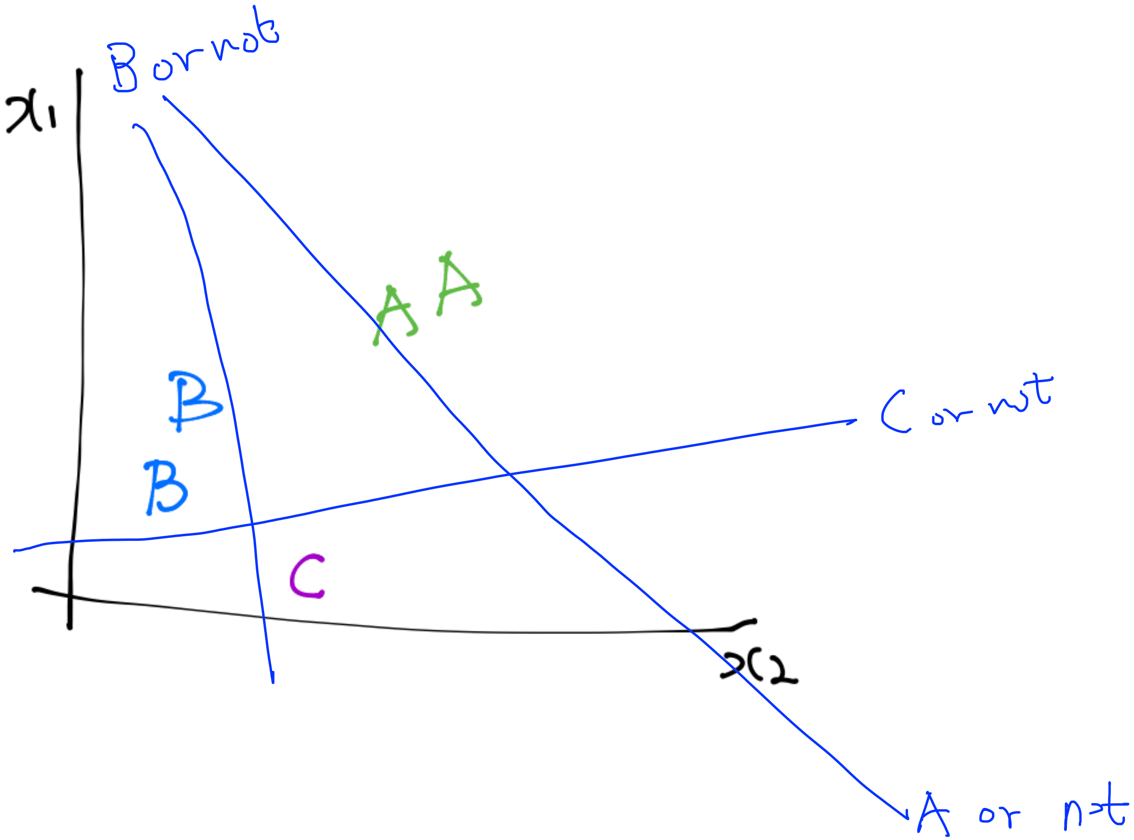


Multinomial classification

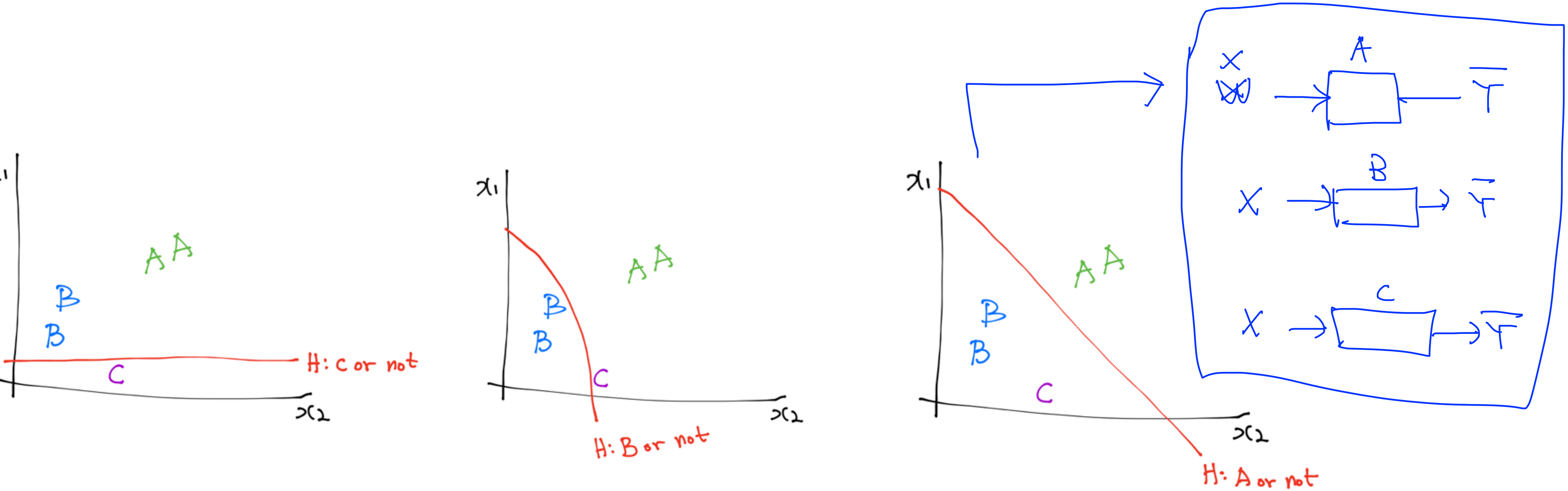
x1 (hours)	x2 (attendance)	y (grade)
10	5	A
9	5	A
3	2	B
2	4	B
11	1	C



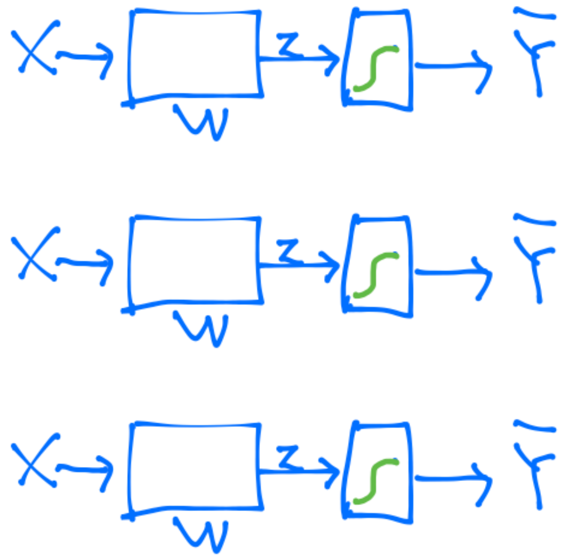
Multinomial classification



Multinomial classification

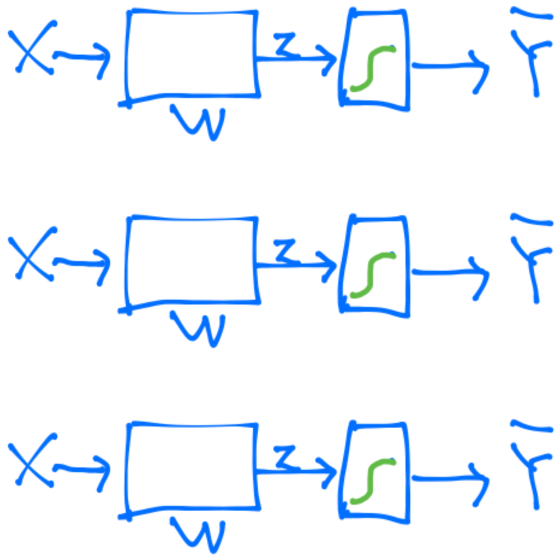


Multinomial classification



Multinomial classification

$$\underbrace{[w_1 \quad w_2 \quad w_3]}_{W}\underbrace{\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}}_{X} = [\underbrace{w_1x_1 + w_2x_2 + w_3x_3}_{H(x)}]$$

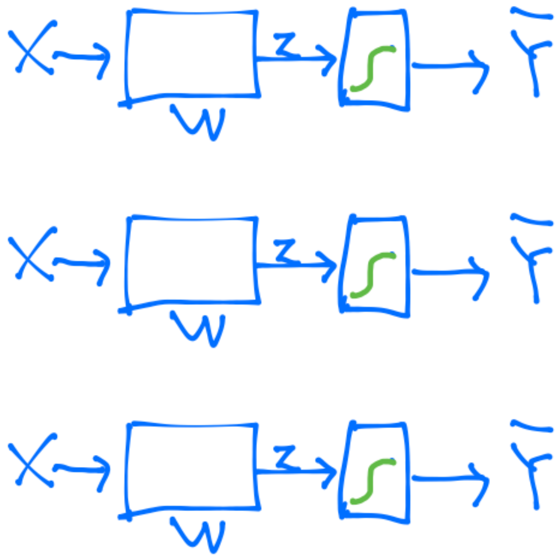


Multinomial classification

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

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

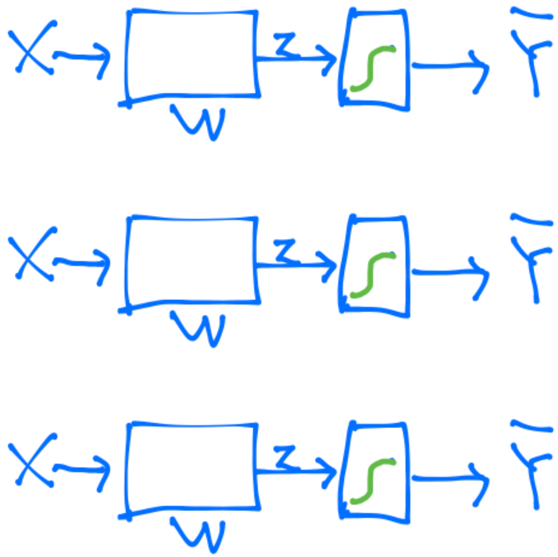
$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_2x_2 + w_3x_3 \end{bmatrix}$$



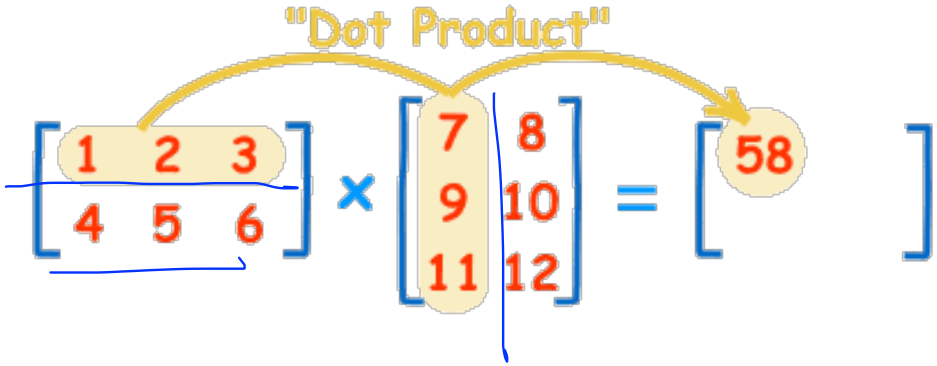
Multinomial classification

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

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



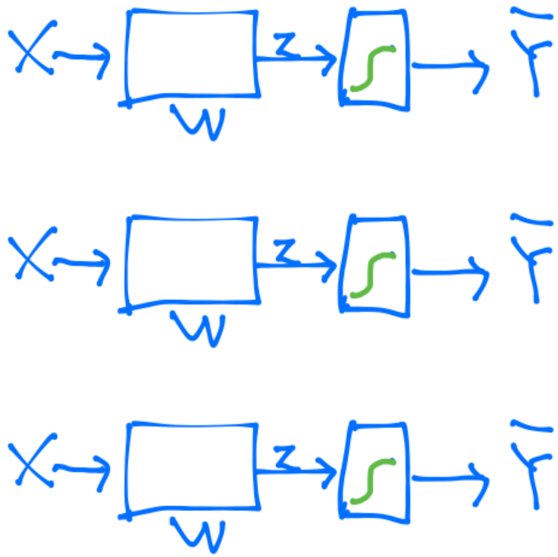
Matrix multiplication



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

Multinomial classification

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



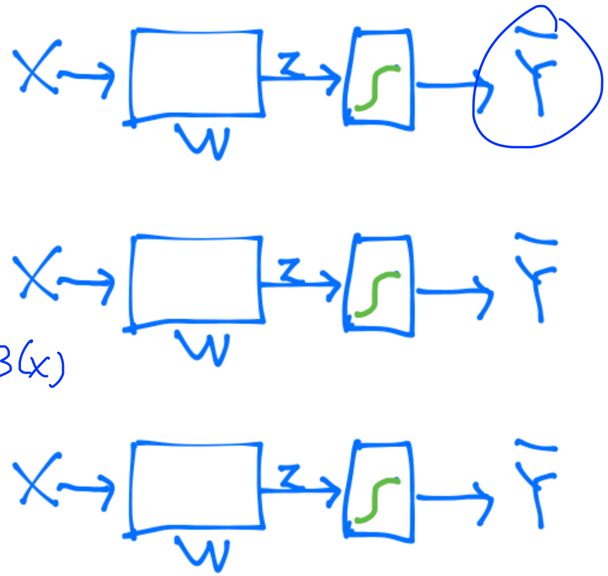
Multinomial classification

$H_A(x)$

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

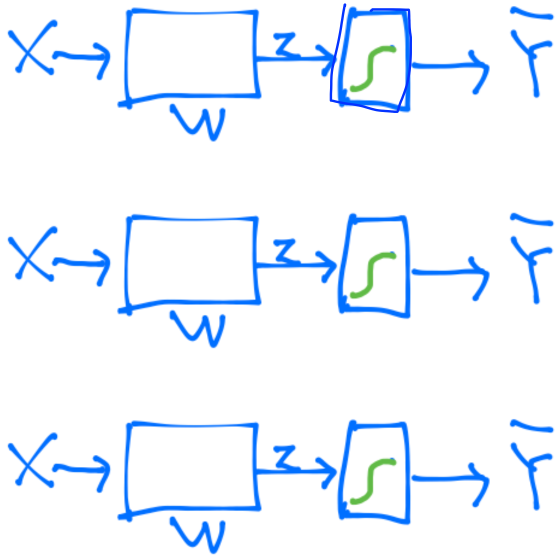
$H_B(x)$

$H_C(x)$



Where is sigmoid?

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



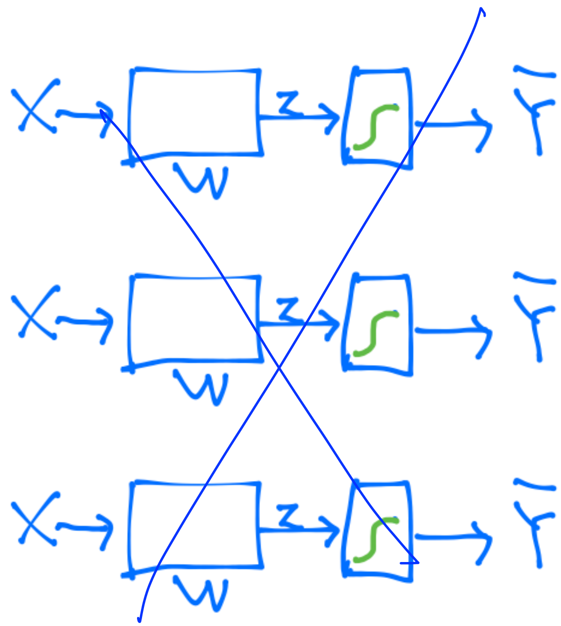
Lecture 6-2

Softmax classification: softmax and cost function

Sung Kim <hunkim+mr@gmail.com>

Where is sigmoid?

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

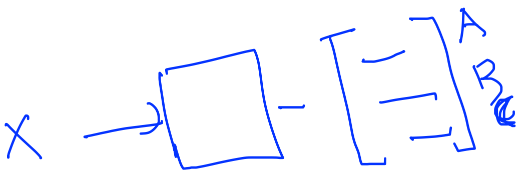


Where is sigmoid?

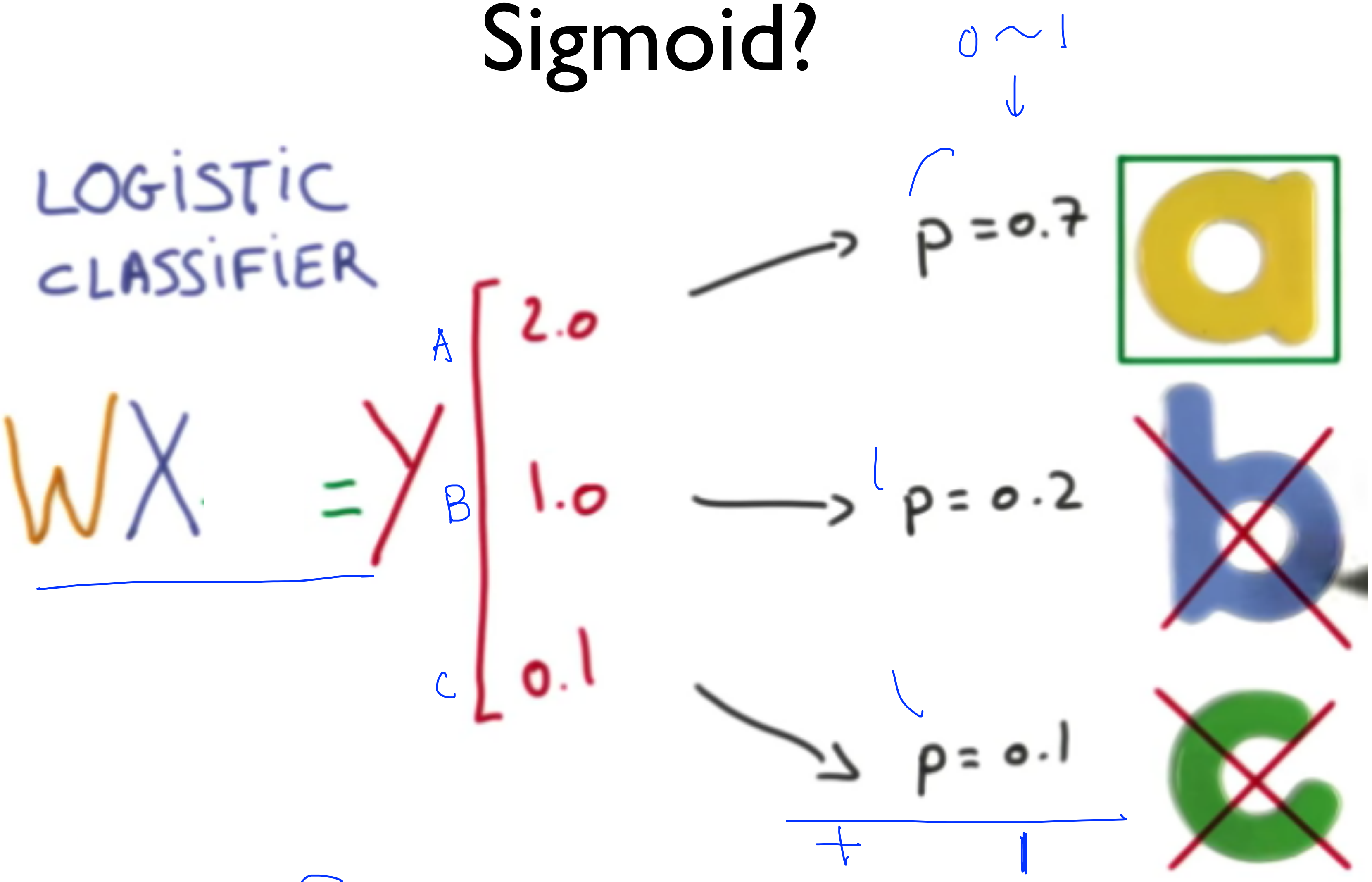
$$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \\ w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \\ w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix}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

$0 \sim 1$

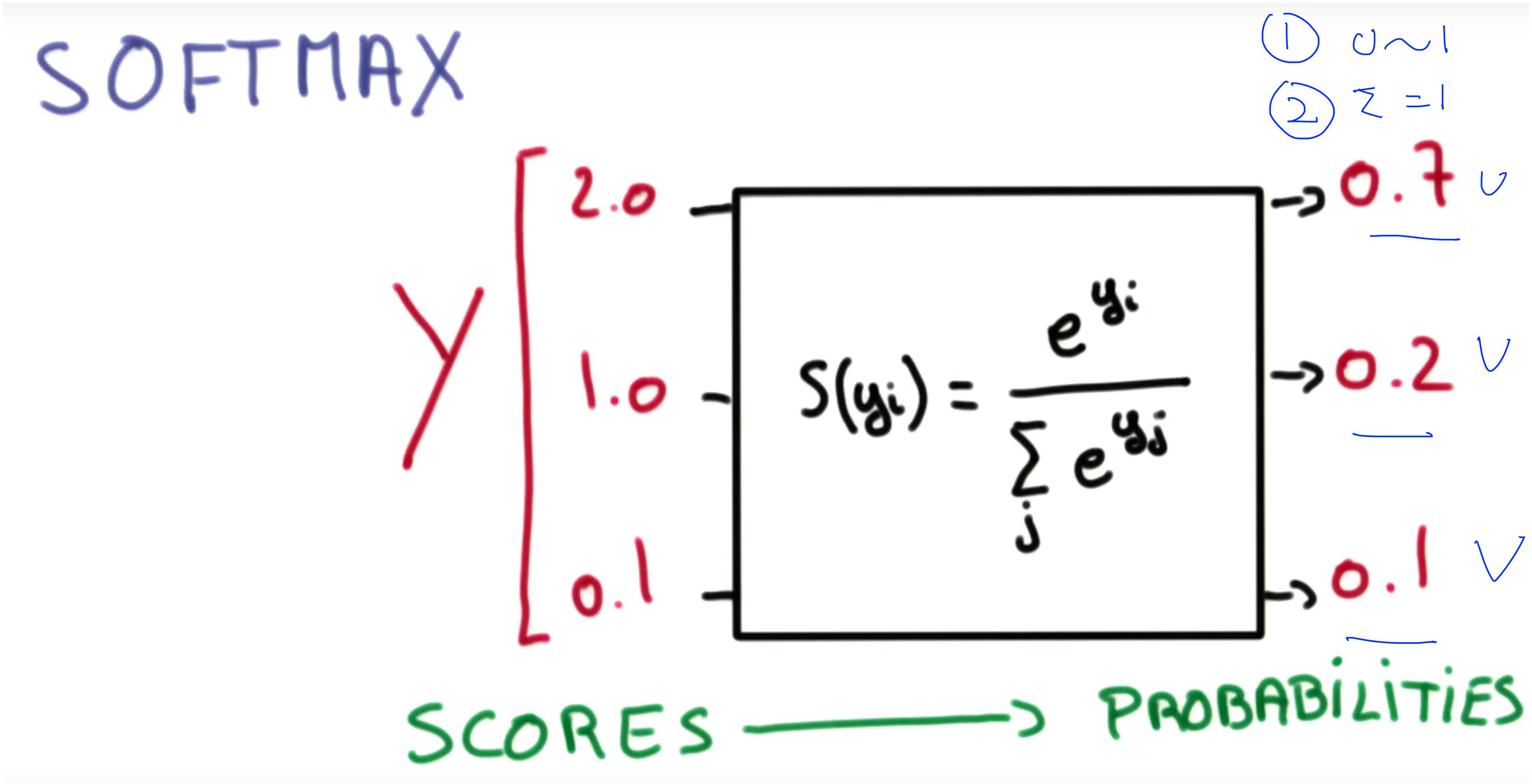
$$\begin{bmatrix} 2.0 \\ 1.0 \\ 0.1 \end{bmatrix}$$



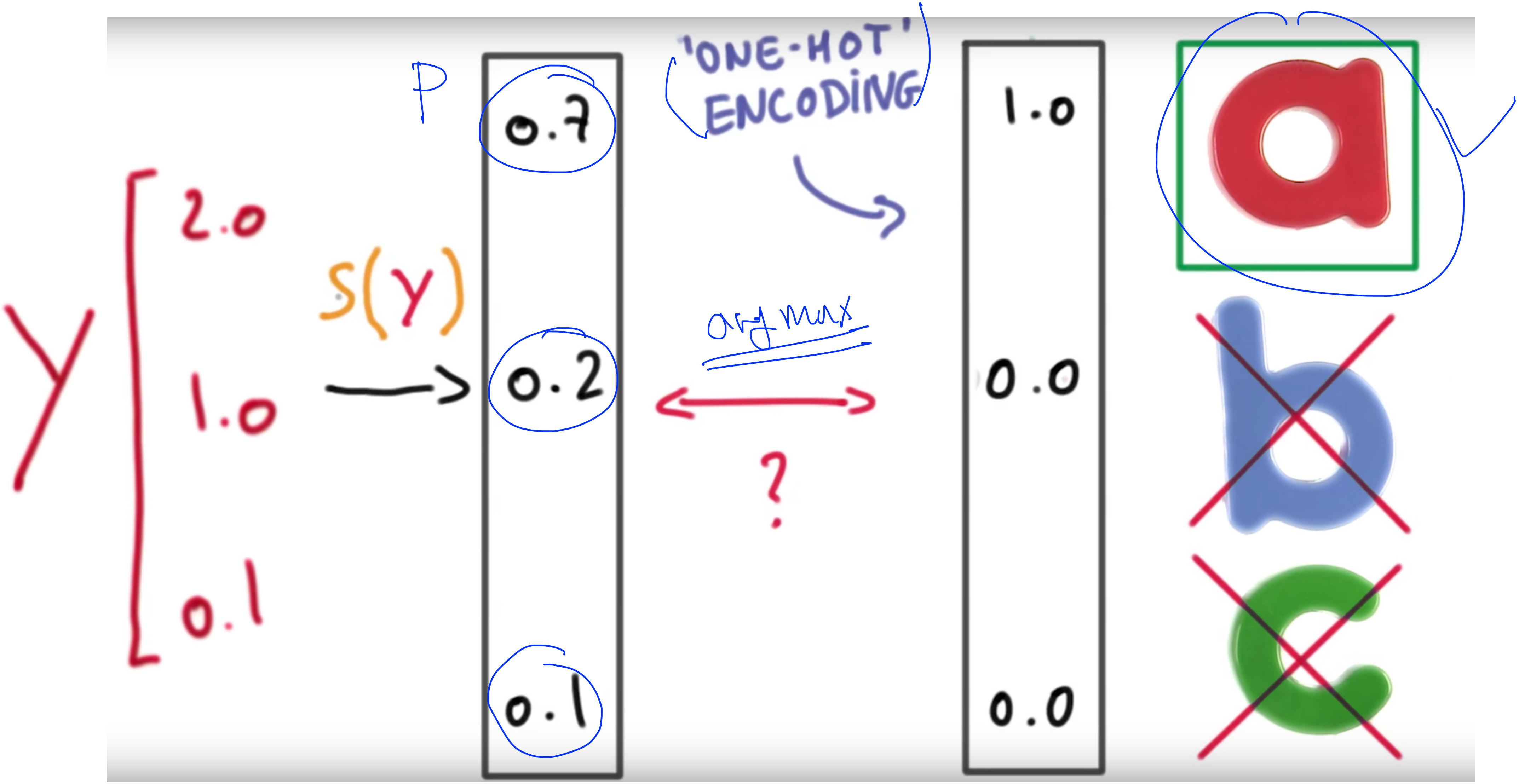
Sigmoid?



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

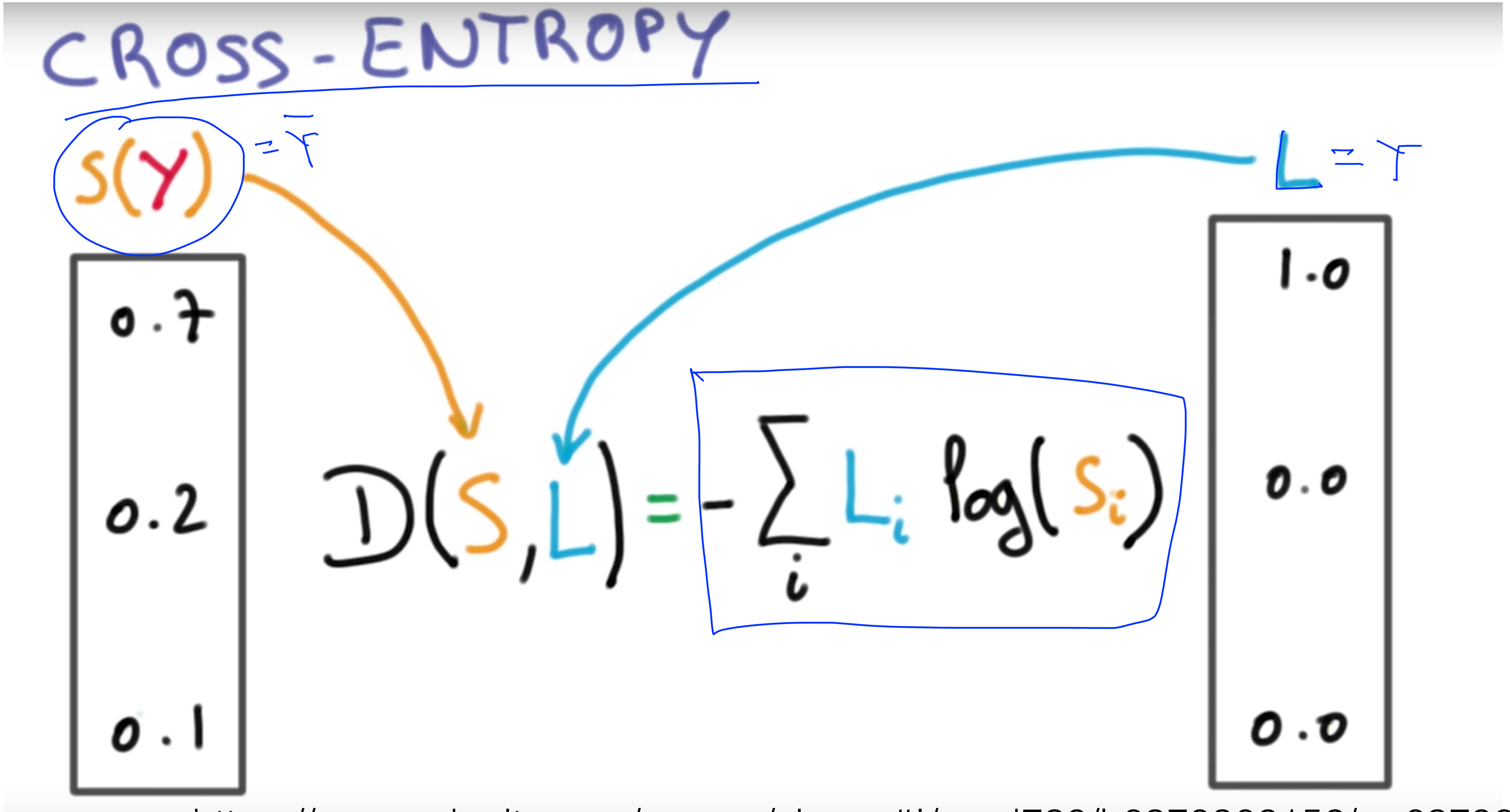


https://www.udacity.com/course/viewer#!/c-ud730/l-6370362152/m-6379811817



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

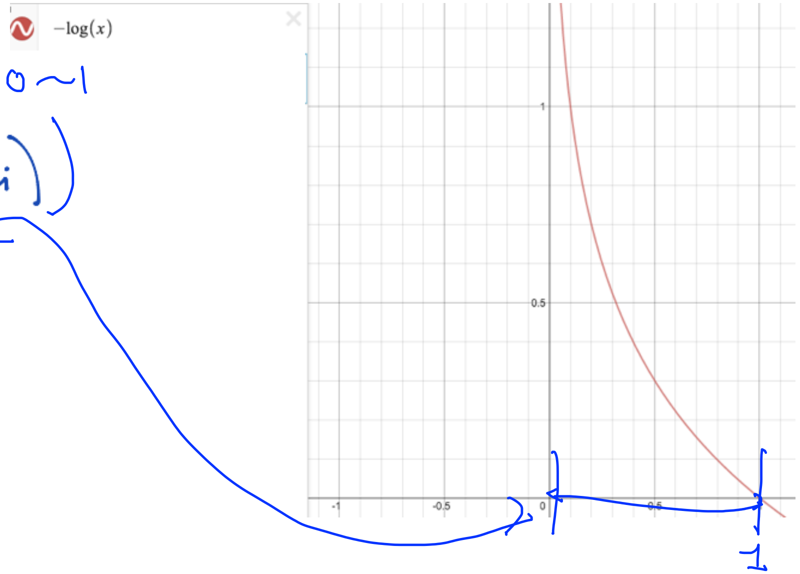
Cost function



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

Cross-entropy cost function

$$-\sum_i L_i \log(S_i) = -\sum_i L_i \log(\bar{y}_i) = \sum_i (L_i) * \left(-\log\left(\overbrace{\bar{y}_i}^{0\sim 1}\right)\right)$$



Cross-entropy cost function

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

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

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

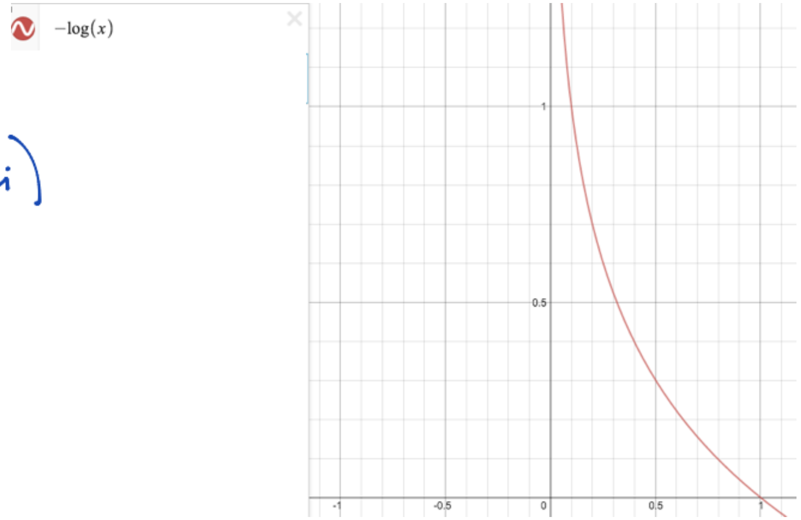
$$\bar{Y} = \begin{array}{c} A \\ B \end{array}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \; B \text{ (ok)}, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot -\log\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{bmatrix} \infty \\ 0 \end{bmatrix} = \begin{bmatrix} 0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

−log

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

Cross-entropy cost function

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



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

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

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

Logistic cost VS cross entropy

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

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

// ?

Cost function

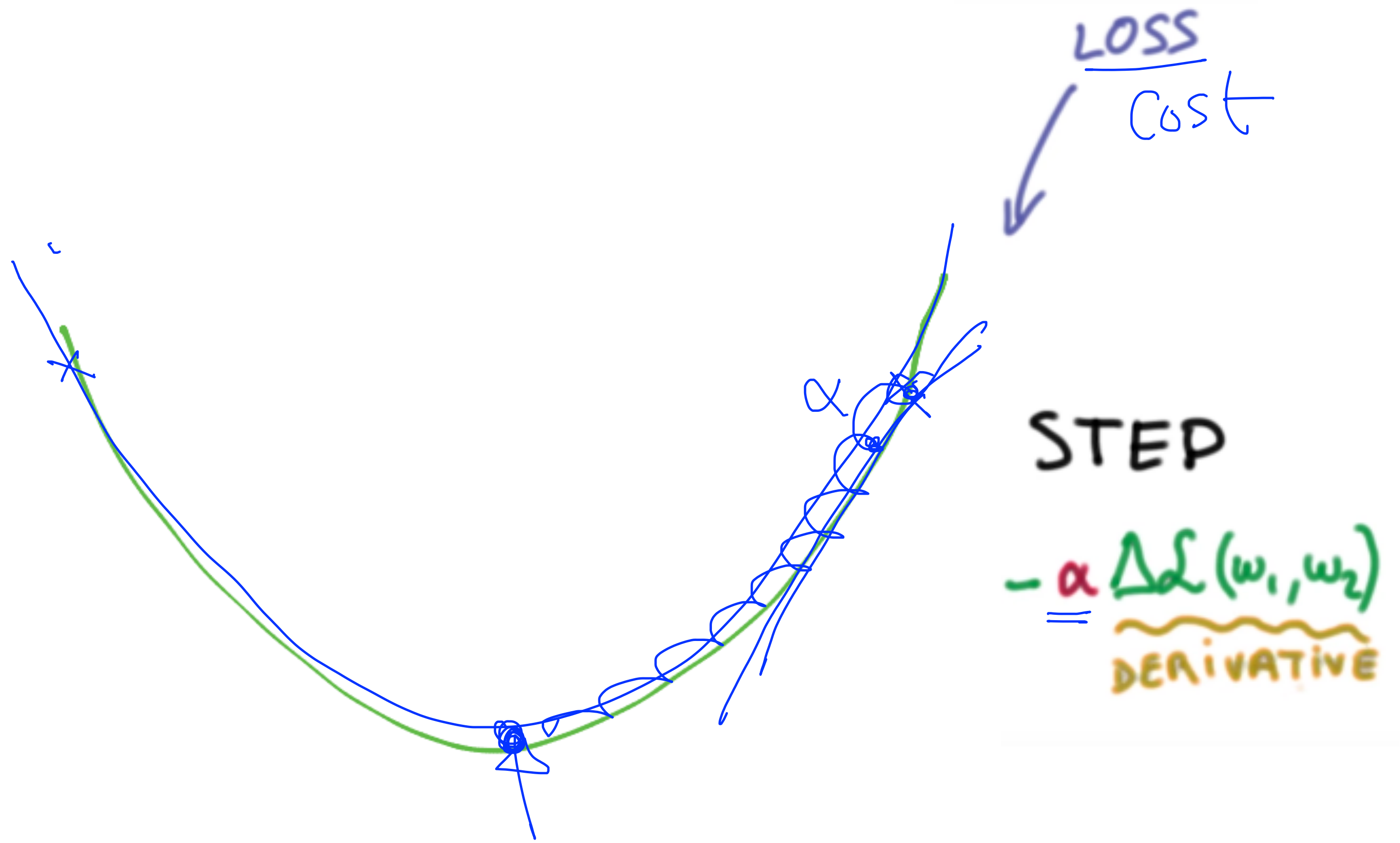
LOSS

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

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