

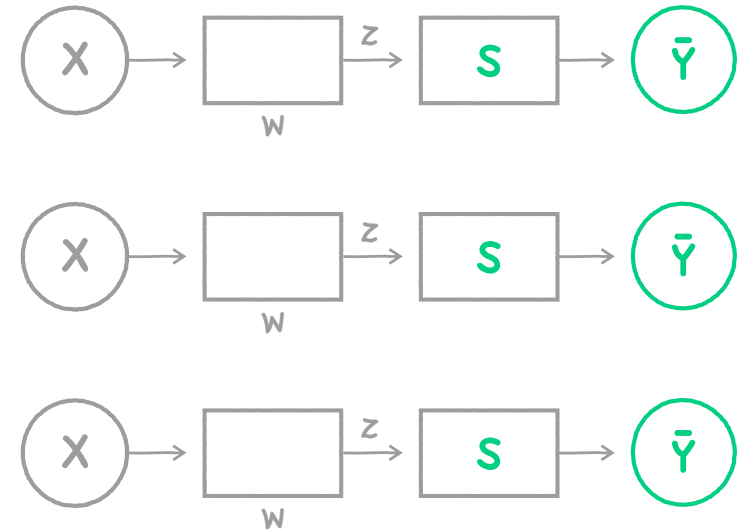
LECTURE 6-2

# SOFTMAX CLASSIFICATION: SOFTMAX & COST FUNCTION

Sung Kim <hunkim+ml@gmail.com>  
<http://hunkim.github.io/ml>




# Where Is Sigmoid?

$$\begin{bmatrix} W_{A1} & W_{A2} & W_{A3} \\ W_{B1} & W_{B2} & W_{B3} \\ W_{C1} & W_{C2} & W_{C3} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} W_{A1}X_1 + W_{A2}X_2 + W_{A3}X_3 \\ W_{B1}X_1 + W_{B2}X_2 + W_{B3}X_3 \\ W_{C1}X_1 + W_{C2}X_2 + W_{C3}X_3 \end{bmatrix} = \begin{bmatrix} \bar{Y}_A \\ \bar{Y}_B \\ \bar{Y}_C \end{bmatrix}$$

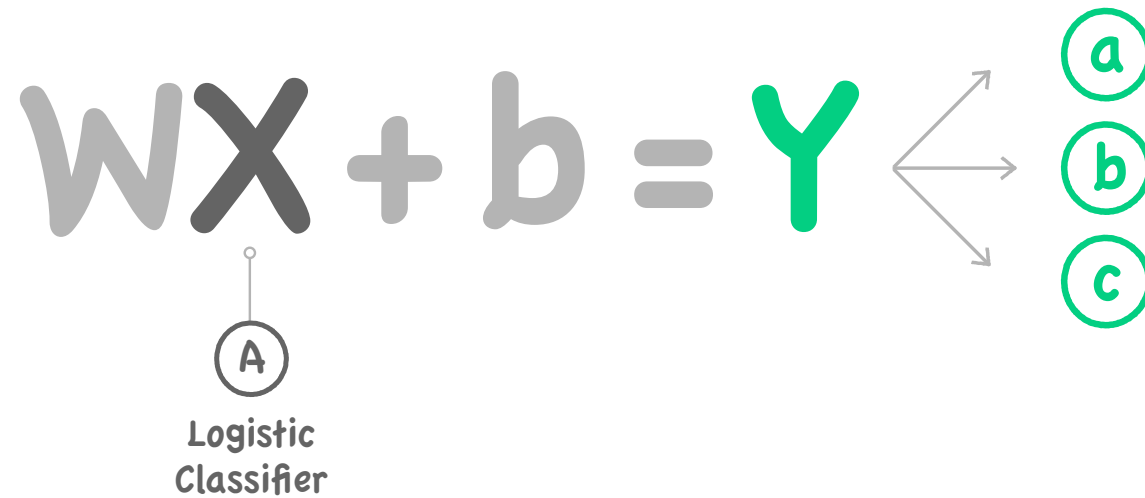


# Where Is Sigmoid?

$$\begin{bmatrix} W_{A1} & W_{A2} & W_{A3} \\ W_{B1} & W_{B2} & W_{B3} \\ W_{C1} & W_{C2} & W_{C3} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} W_{A1}X_1 + W_{A2}X_2 + W_{A3}X_3 \\ W_{B1}X_1 + W_{B2}X_2 + W_{B3}X_3 \\ W_{C1}X_1 + W_{C2}X_2 + W_{C3}X_3 \end{bmatrix} = \begin{bmatrix} \bar{Y}_A \\ \bar{Y}_B \\ \bar{Y}_C \end{bmatrix}$$

$\bar{Y}$	2.0	→ 
	1.0	→ 
	0.1	→ 

# Where Is Sigmoid?

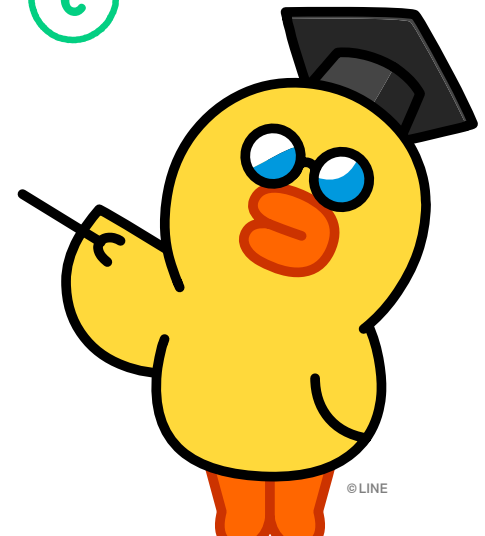


# Where Is Sigmoid?

$$\underbrace{W X + b}_{\text{Trained}} = Y$$

Weights                      Bias




Three arrows point from the result  $Y$  to three green circles labeled  $a$ ,  $b$ , and  $c$ .



# Where Is Sigmoid?

$$WX + b = Y$$

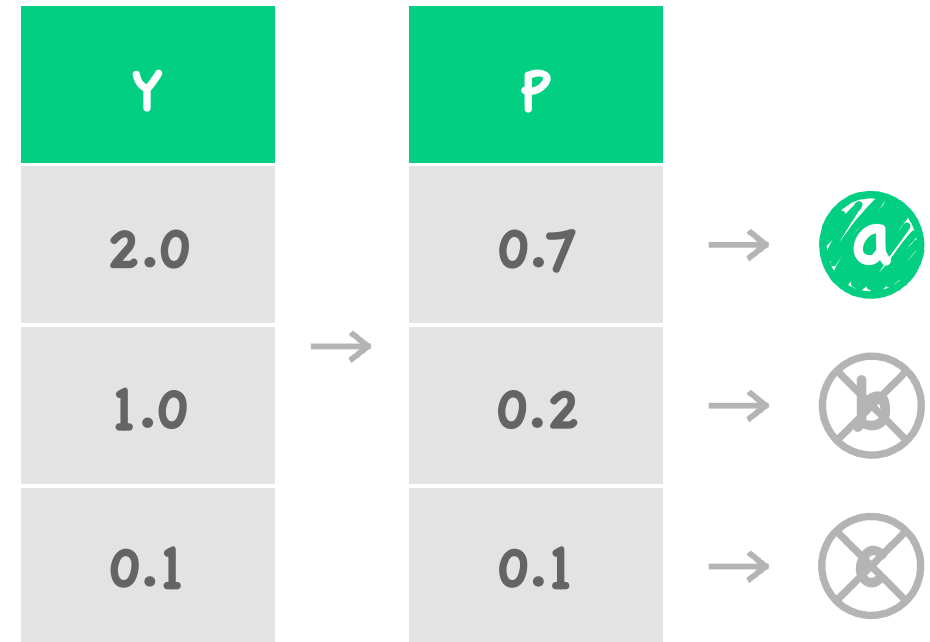
  
Logistic  
Classifier

Y	
2.0	→ 
1.0	→ 
0.1	→ 

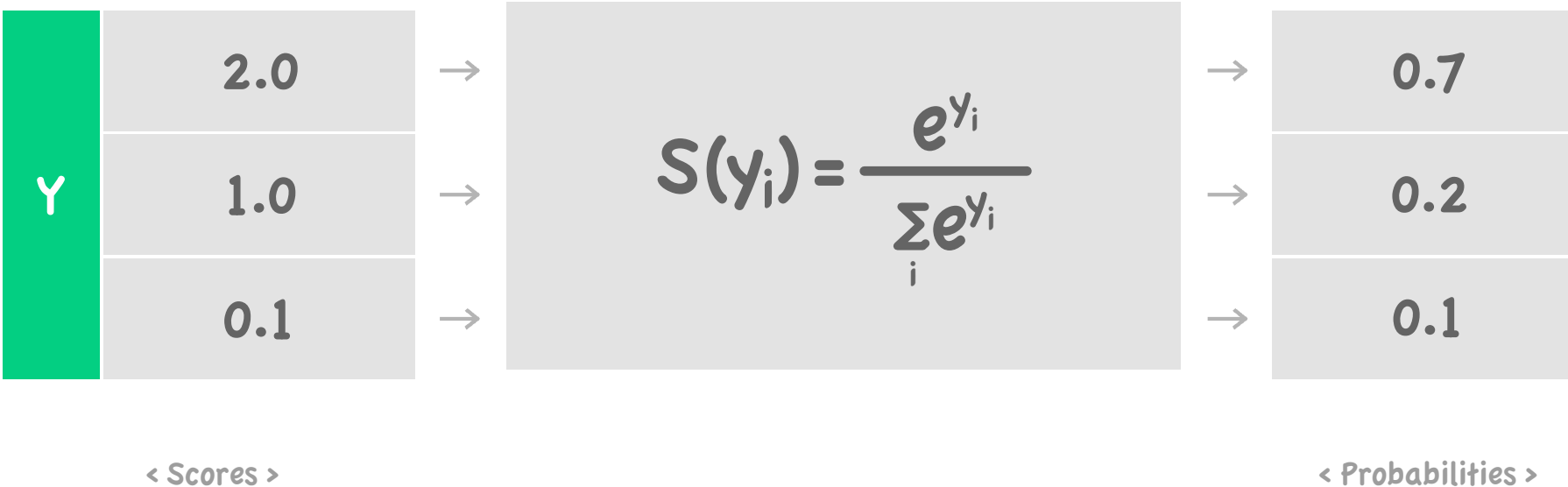
# Sigmoid?

$$WX = Y$$

  
Logistic  
Classifier

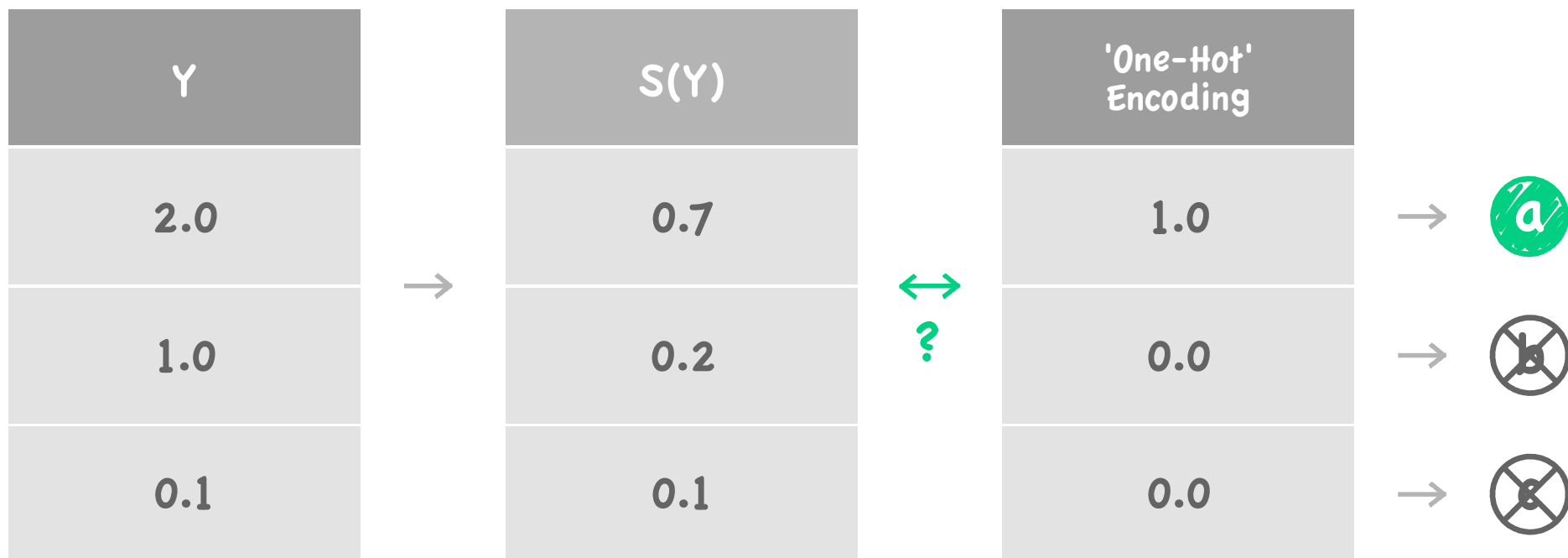


# Softmax





# Softmax



# Cost Function

## Cross - Entropy

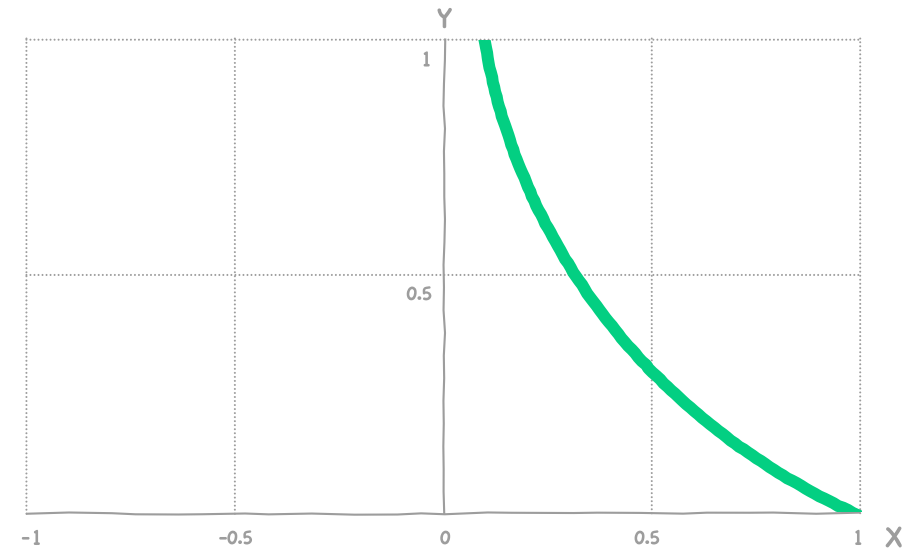
$$D(S, L) = -\sum_i L_i \log(S_i)$$

S(Y)	L
0.7	1.0
0.2	0.0
0.1	0.0

# Cross-Entropy Cost Function

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

$$-\sum_i L_i \log(\bar{Y}_i) = \sum_i L_i x - \log(\bar{Y}_i)$$



(A)  $-\log(x)$

# Cross-Entropy Cost Function

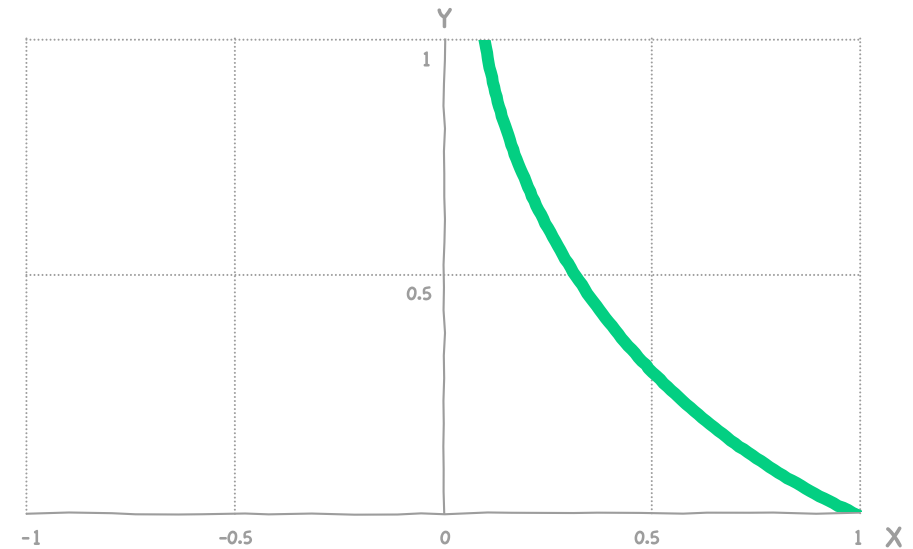
$$-\sum_i L_i \log(S_i)$$

$$-\sum_i L_i \log(\bar{Y}_i) = \sum_i L_i x - \log(\bar{Y}_i)$$

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

$$\bar{Y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\bar{Y} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$



(A)  $-\log(x)$

# Cross-Entropy Cost Function

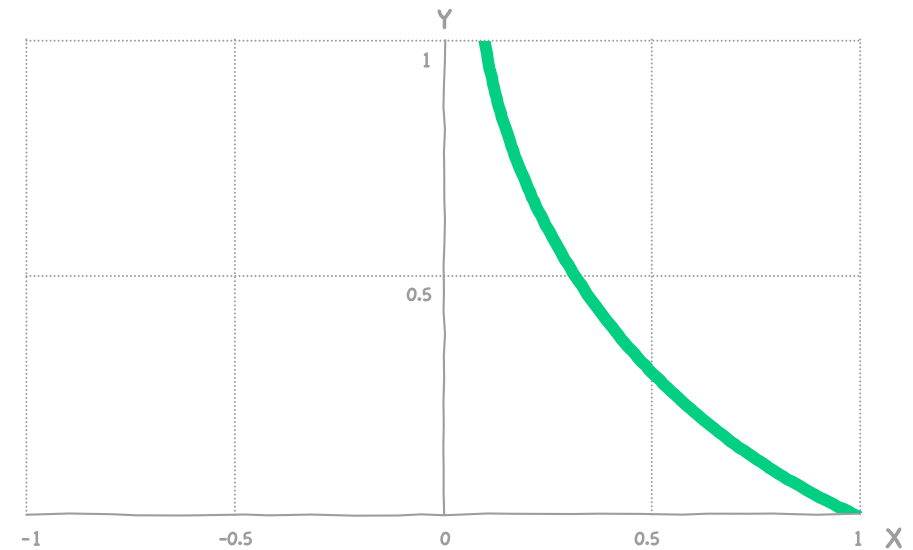
$$-\sum_i L_i \log(S_i)$$

$$-\sum_i L_i \log(\bar{Y}_i) = \sum_i L_i x - \log(\bar{Y}_i)$$

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

$$\bar{Y} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\bar{Y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$



(A)  $-\log(x)$

# Logistic Cost vs Cross Entropy

$$\underline{c(H(x), y) = y \log(H(x)) - (1 - y) \log(1 - H(x))}$$

$$D(\underset{\downarrow}{S}, \underset{\downarrow}{L}) = -\sum_i L_i \log(S_i)$$

# Cost Function

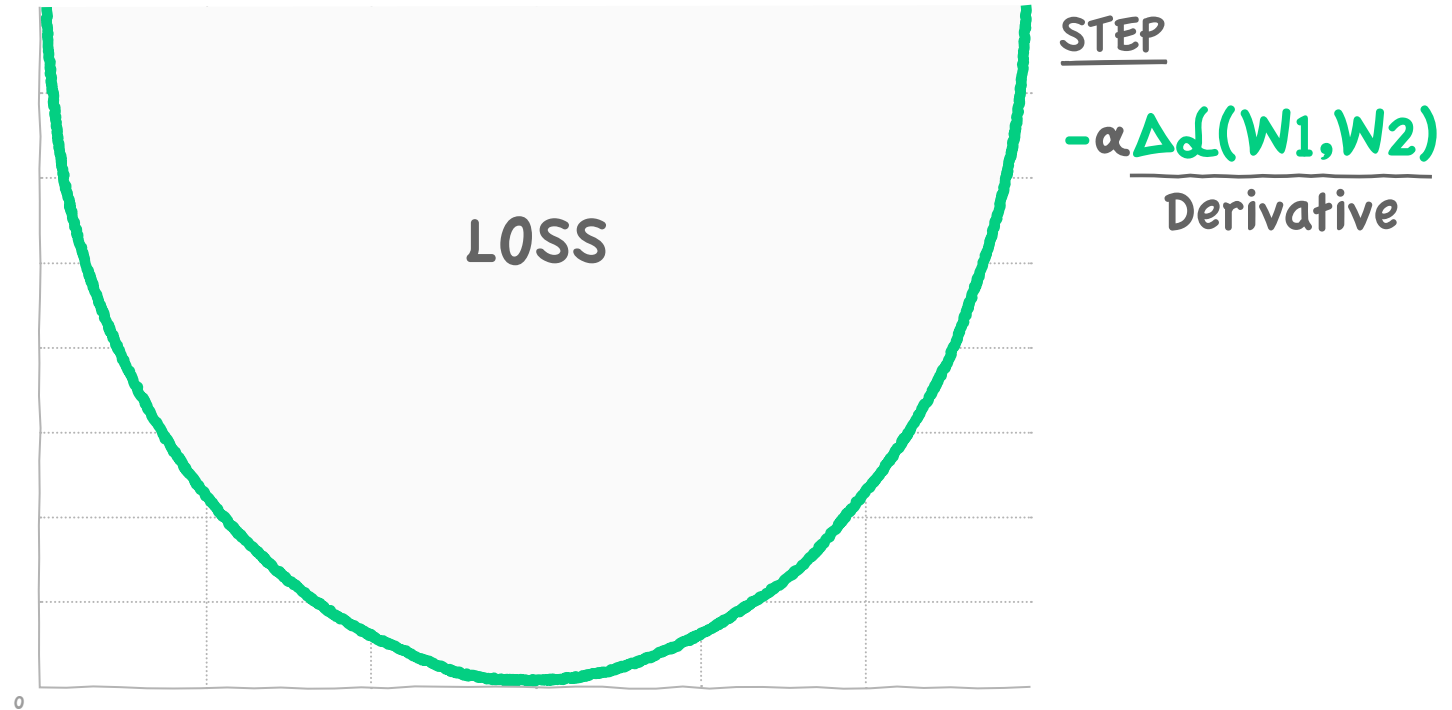
$$\mathcal{L} = \frac{1}{N} \sum_i \mathcal{D}(S(w x_i + b), L_i)$$

The diagram shows the cost function formula with several annotations. A vertical line with a small circle at the top points from the word "Loss" to the symbol  $\mathcal{L}$ . Another vertical line with a small circle at the top points from the word "Training Set" to the summation index  $i$ . A horizontal line with small circles at both ends spans the width of the summation term  $\mathcal{D}(S(w x_i + b), L_i)$ , with the label "Training Set" centered below it.

Loss

Training Set

# Gradient Descent





**NEXT LECTURE**

# **APPLICATION & TIPS**