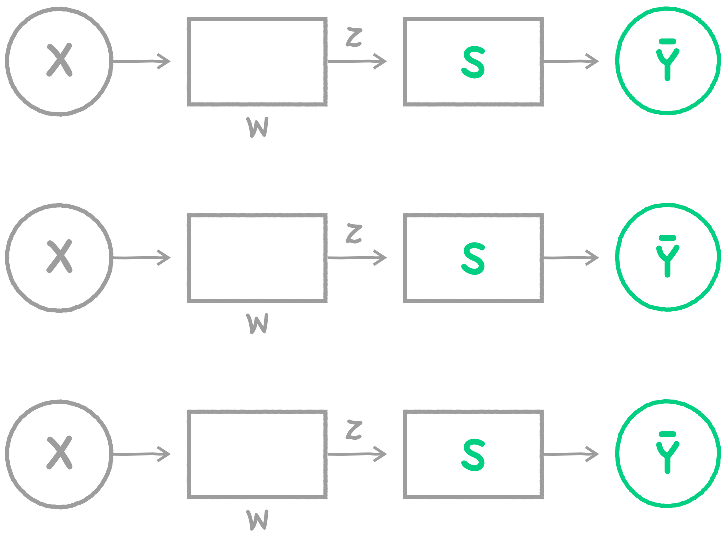
LECTURE 6-2

SOFTMAX CLASSIFICATION : SOFTMAX & COST FUNCTION

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

Where Is Sigmoid?

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

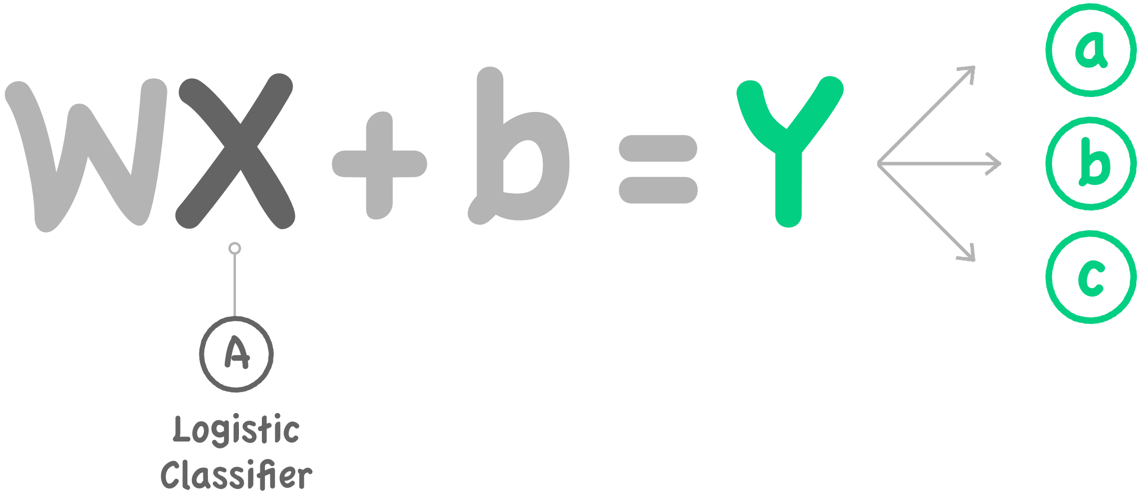


Where Is Sigmoid?

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

$\bar{Y}$			
	2.0	$\rightarrow$	a
	1.0	$\rightarrow$	$\cancel{b}$
	0.1	$\rightarrow$	$\cancel{c}$

Where Is Sigmoid?



Where Is Sigmoid?

$$WX + b = Y$$

a
b
c

Weights

Bias

Trained

Where Is Sigmoid?

$$WX + b = Y$$

A

Logistic Classifier

Y	
2.0	→ a
1.0	→ ~~b~~
0.1	→ ~~c~~

Sigmoid?

$$WX = Y$$

A

Logistic Classifier

Y	P	
2.0	0.7	a
1.0	0.2	~~b~~
0.1	0.1	~~c~~

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

Softmax

	Scores		Probabilities
Y	2.0	→	0.7
	1.0	→	0.2
	0.1	→	0.1

$$S(y_i) = \frac{e^{y_i}}{\sum_i e^{y_i}}$$

< Scores >

< Probabilities >

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

Softmax

Y	S(Y)	'One-Hot' Encoding	
2.0	0.7	1.0	a
1.0	0.2	0.0	~~b~~
0.1	0.1	0.0	~~c~~

?

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

Cost Function

Cross-Entropy

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

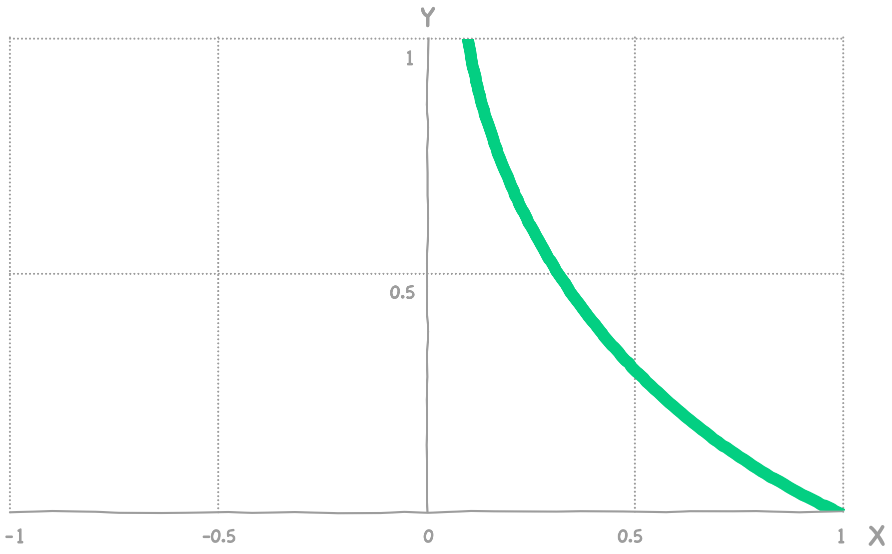
S(Y)	L
0.7	1.0
0.2	0.0
0.1	0.0

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

Cross-Entropy Cost Function

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

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



(A) -log(x)

Cross-Entropy Cost Function

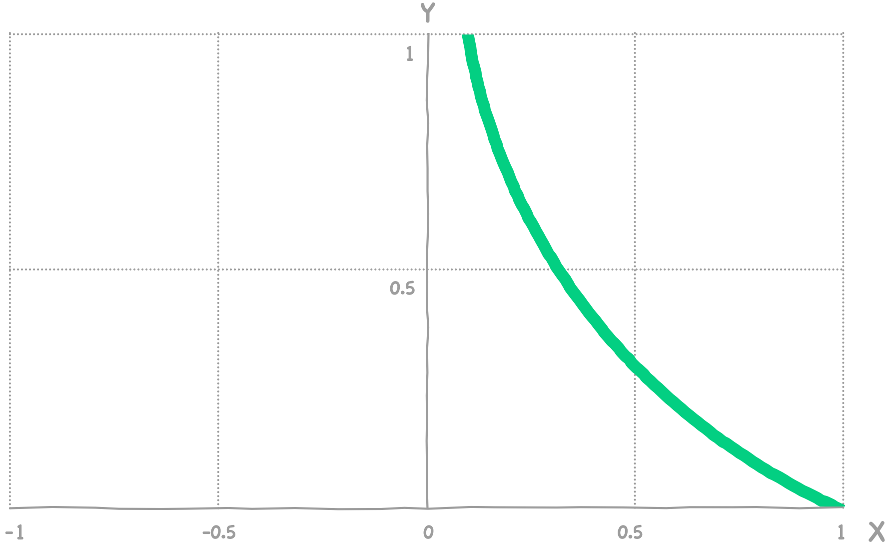
$$-\sum_i L_i \log(S_i)$$

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

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

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

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



(A) -log(x)

Cross-Entropy Cost Function

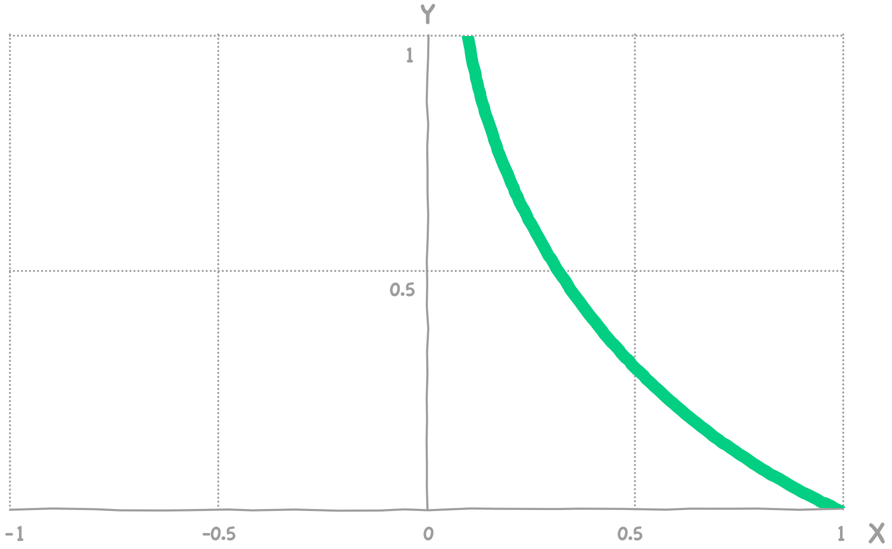
$$-\sum_i L_i \log(S_i)$$

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

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

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

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



(A) -log(x)

Logistic Cost vs Cross Entropy

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

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

Cost Function

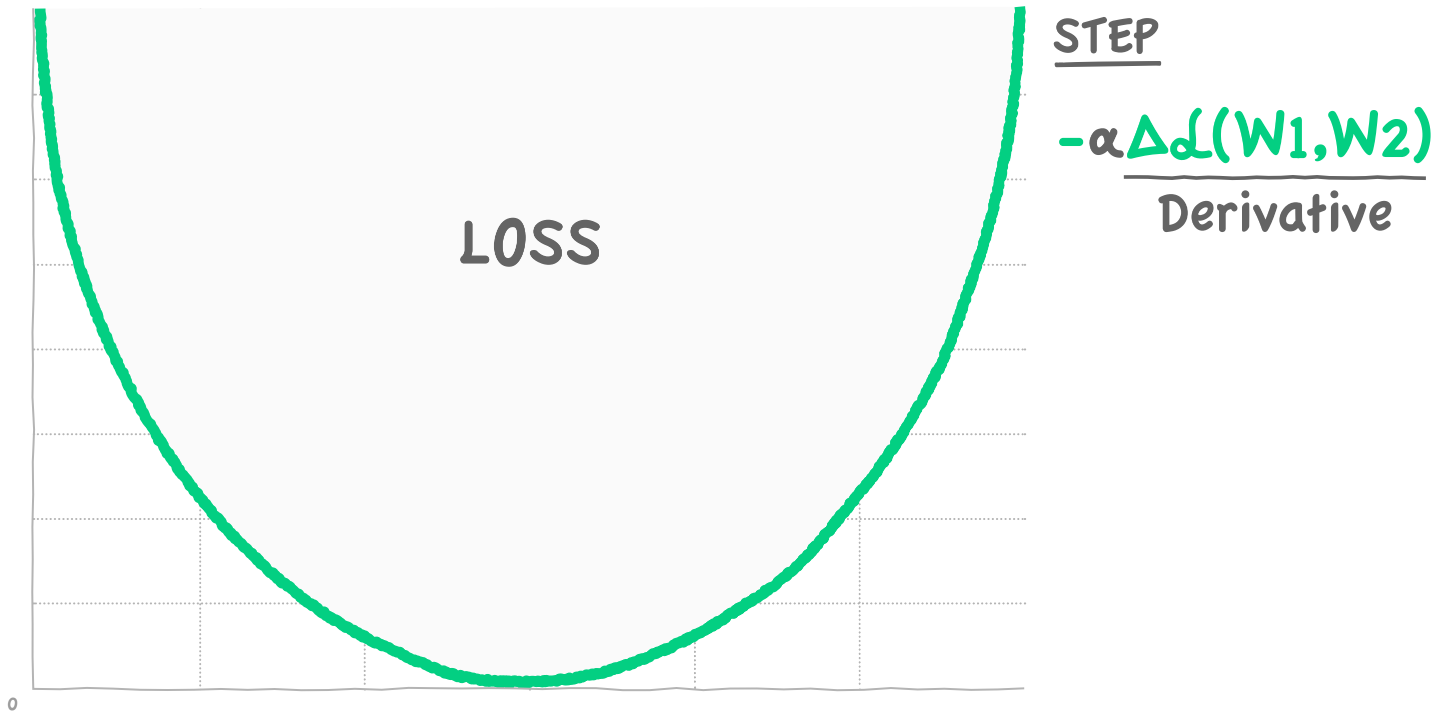
$$\mathcal{L} = \frac{1}{N} \sum_i D(S(wx_i + b), L_i)$$

Loss

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 LECTURE

APPLICATION & TIPS