



Multinomial Classification

머신러닝/딥러닝

임 경 태

CONTENTS

—

①

~~Logistic Regression~~

②

Multinomial Classification

02. Multinomial Classification (개요)

❑ Binary(Pass/fail) 분류가 아닌 N개의 class 분류 문제

- 우리반 학생들의 공부시간 대비 수능 등급이 아래와 같다. 나는 7,3 했는데 무슨 등급일까?

이름	수학 공부시간	영어 공부시간	등급
야쓰오	2	3	C
오공	4	8	C
블리츠	6	2	B
페이커	8	4	A

02. Multinomial Classification (multi-class classification) VS Logistic Regression

❑ Logistic Regression

(Binary Classification) 문제?

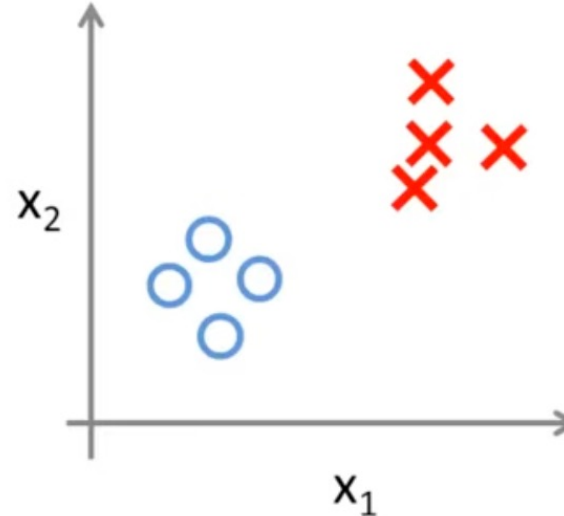
- ❑ 스팸메일 검출: Yes, No
- ❑ 영화 추천: 추천, 비추천
- ❑ 수능 합/불: 합격, 불합격
- ❑ Where $y=\{0, 1\}$

❑ Multinomial Classification

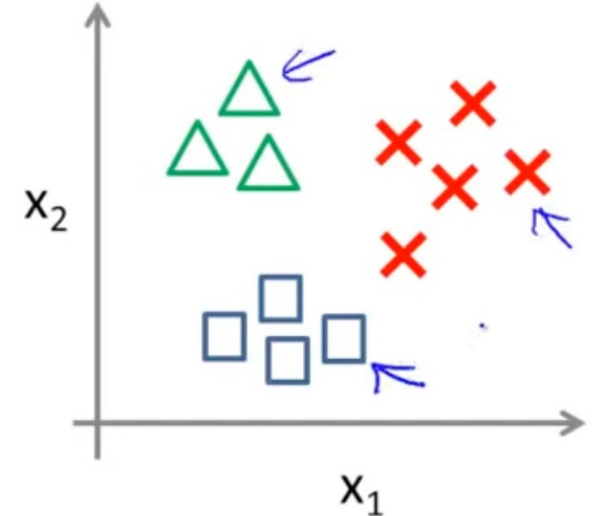
(Multi-class Classification) 문제?

- ❑ 수능 등급: 1등급, 2등급, 3등급
- ❑ 객체 인식: 고양이, 강아지, 사자
- ❑ 감정 인식: 기쁨, 슬픔, 냉소, 분노
- ❑ Where $y=\{0, 1, 2, .. n\}$

Binary classification:



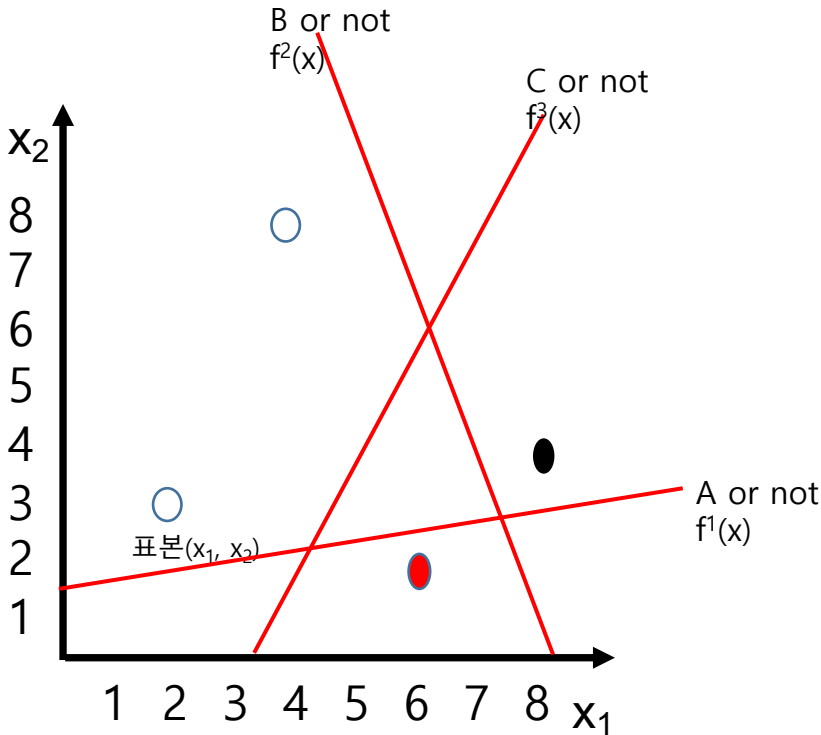
Multi-class classification:



04. Multinomial Classification (개념)

- 수능 등급을 어떻게 나눌 수 있을까?
 - Linear 선을 여러개 그어보자!

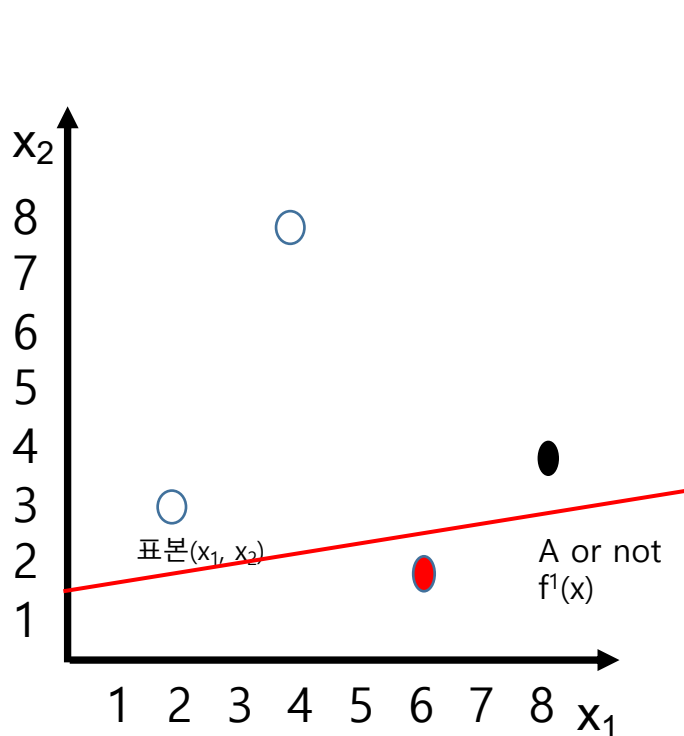
이름	수학 공부시간	영어 공부시간	등급
야쓰오	2	3	C
오공	4	8	C
블리츠	6	2	B
페이커	8	4	A



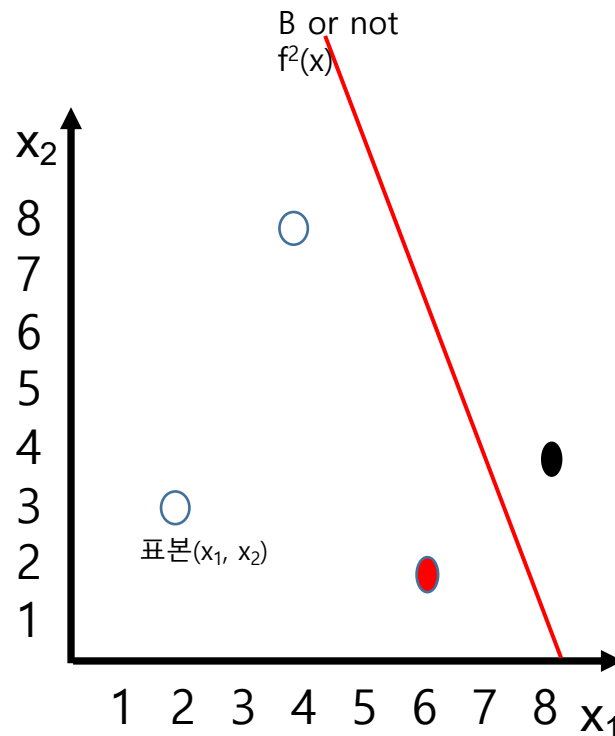
04. Multinomial Classification (개념)

❑ A,B,C 등급은 숫자가 아닌데 어떻게 분류하지?

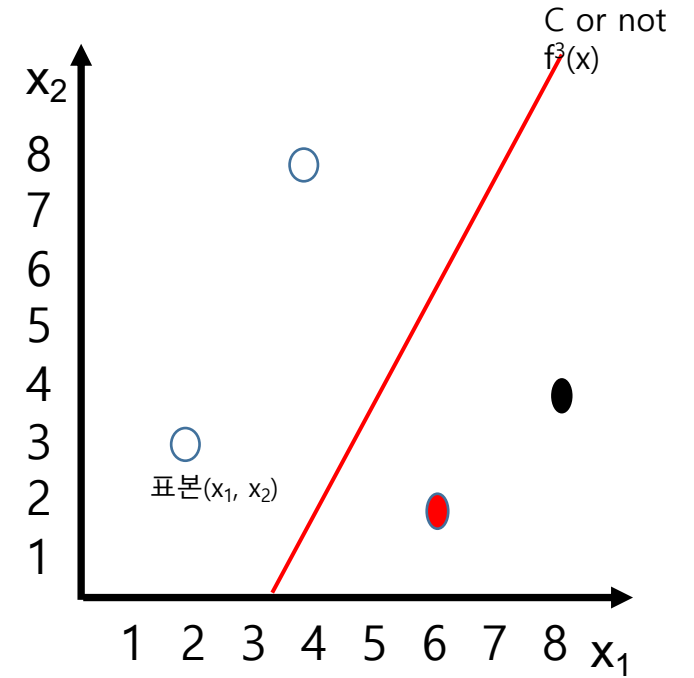
- ❑ 숫자로 만들지 뭐! 종속 변수 y 결과의 범위를 **index** [0 ~ n]로 만들자 where n = Number of classes
- ❑ Input $X_i=\{x_1, x_2\}_i$ 를 f^1, f^2, f^3 에 각각 넣은 후 그중 가장 적합한 결과를 반환하는 분류기(f_i)를 선택



$$X \rightarrow W_a X = Z_a \rightarrow S(Z_a)$$



$$X \rightarrow W_b X = Z_b \rightarrow S(Z_b)$$



$$X \rightarrow W_c X = Z_c \rightarrow S(Z_c)$$

04. Multinomial Classification (개념)

□ A,B,C 등급은 숫자가 아닌데 어떻게 분류하지?

- 숫자로 만들지 뭐! 종속 변수 y 결과의 범위를 **index** [0 ~ n]로 만들자 where n= Number of classes
- Input $X_i=\{x_1,x_2\}_i$ 를 f^1,f^2,f^3 에 각각 넣은 후 그중 가장 적합한 결과를 반환하는 분류기(f_i)를 선택

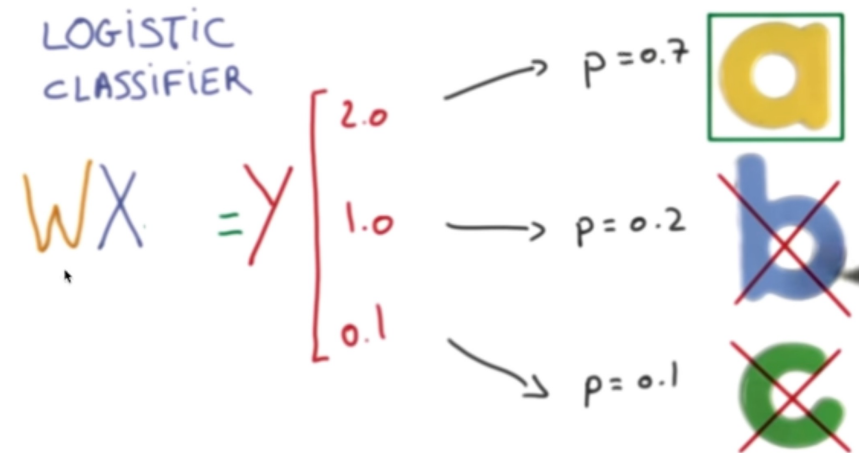
$$\begin{aligned} X &\rightarrow W_a X = Z_a \rightarrow S(Z_a) & (x_1 \ x_2 \ x_3) \cdot \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} &= (x_1 w_1 + x_2 w_2 + x_3 w_3) \\ X &\rightarrow W_b X = Z_b \rightarrow S(Z_b) & (x_1 \ x_2 \ x_3) \cdot \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} &= (x_1 w_1 + x_2 w_2 + x_3 w_3) \\ X &\rightarrow W_c X = Z_c \rightarrow S(Z_c) & (x_1 \ x_2 \ x_3) \cdot \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} &= (x_1 w_1 + x_2 w_2 + x_3 w_3) \end{aligned}$$

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

04. Multinomial Classification (Softmax)

- A,B,C 등급은 숫자가 아닌데 어떻게 분류하지?
 - 숫자로 만들지 뭐! 종속 변수 y 결과의 범위를 **index** $[0 \sim n]$ 로 만들자 where $n = \text{Number of classes}$
 - Input $X_i = \{x_1, x_2\}_i$ 를 f^1, f^2, f^3 에 각각 넣은 후 그중 가장 적합한 결과를 반환하는 분류기(f_i)를 선택

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

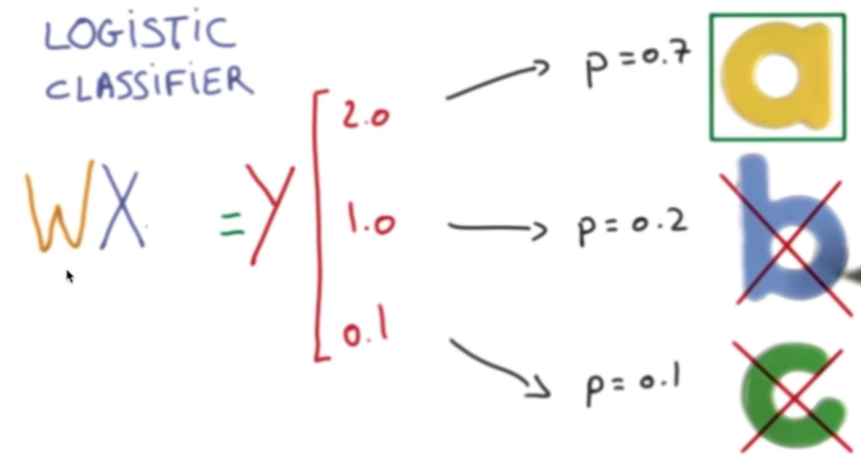


04. Multinomial Classification (Softmax)

- ❑ **Softmax:** Classifier에서 나온 값을 0~1사이 값을 가지는 확률 분포 값으로 변경해보자!

$$\text{softmax}(\mathbf{x})_i = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

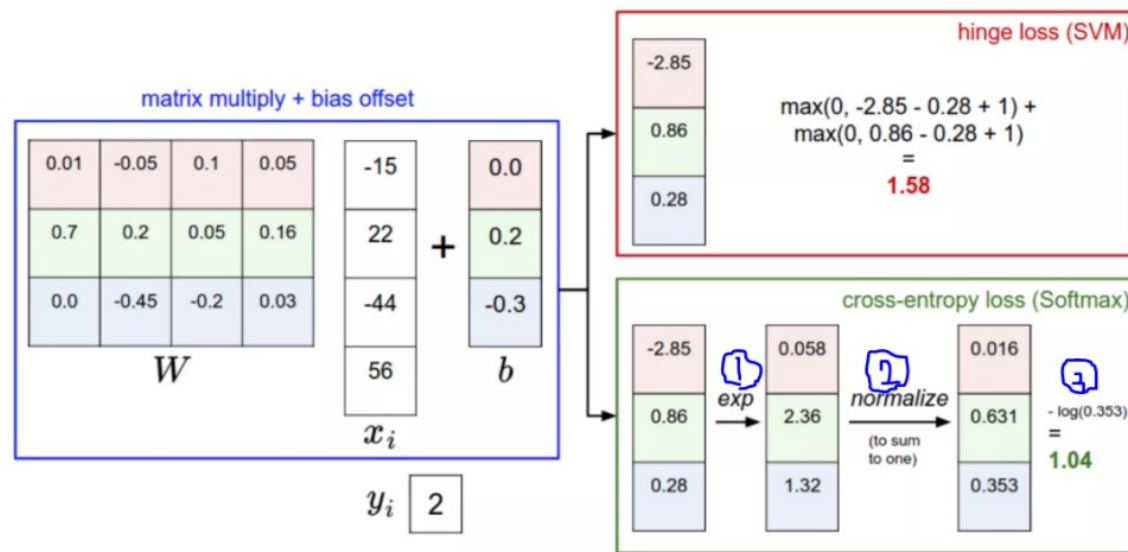
$$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \\ w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \\ w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix}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



04. Multinomial Classification (Softmax)

❑ **Softmax**: Classifier에서 나온 값을 0~1사이 값을 가지는 확률 분포 값으로 변경해보자!

$$\text{softmax}(x)_i = \frac{e^{x_i}}{\sum_j e^{x_j}}$$



Softmax의 의미: logits의 각 값들을 1) 양수로 만들고(파란색 1번 exp) 2)거기에 전체합으로 나누어 normalized(파란색2번) 한다

05. Multinomial Classification (Cross Entropy)

❑ **Cross Entropy Loss:** 우리가 예측한 확률분포값과 실제 정답을 비교 해서 손실값 산정

cross-entropy가 무엇이나 바로 두 확률분포의 거리(차이) 를 계산하는 방법이다.

$$CE(y, \hat{y}) = - \sum_{i=1}^{N_c} y_i \log(\hat{y}_i)$$

$$\hat{y} = [0.016, 0.631, 0.353]$$

$$y = [0, 0, 1]$$

일 경우 수식에 의해 $-\log(0.353) = 1.04$

$$\hat{y} = [0, 0, 1]$$

$$y = [0, 0, 1]$$

일 경우 수식에 의해 $-\log(1) = 0$

$$\hat{y} = [0, 1, 0]$$

$$y = [0, 0, 1]$$

일 경우 수식에 의해 $-\log(0) = \text{무한대}$

$$\hat{y} = [0, 0.1, 0.9]$$

$$y = [0, 0, 1]$$

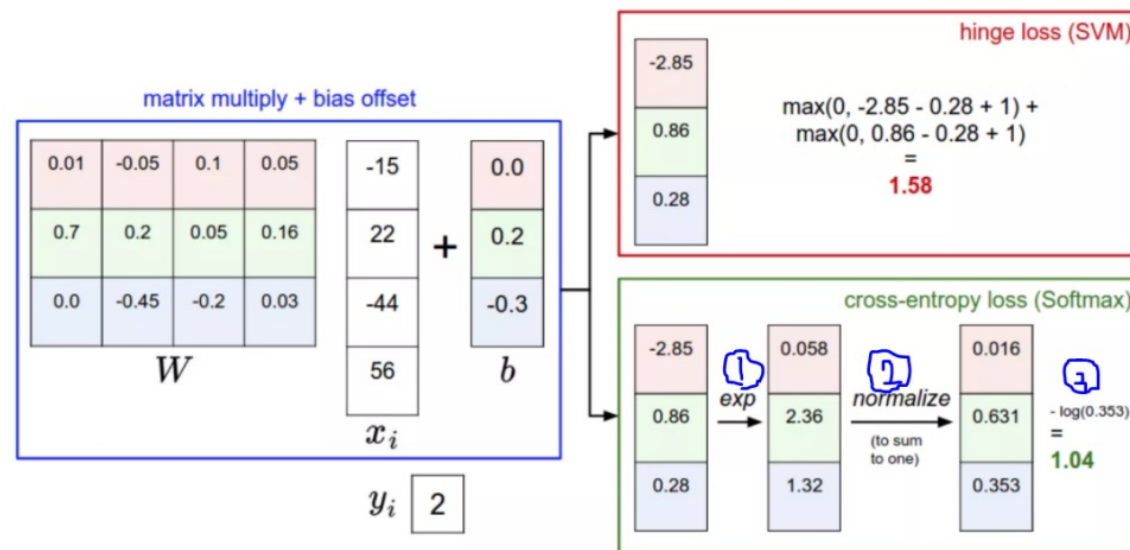
일 경우 수식에 의해 $-\log(0.9) = 0.045$

05. Multinomial Classification (Cross Entropy)

❑ **Cross Entropy Loss:** 우리가 예측한 확률분포값과 실제 정답을 비교 해서 손실값 산정

cross-entropy가 무엇이나 바로 두 확률분포의 거리(차이) 를 계산하는 방법이다.

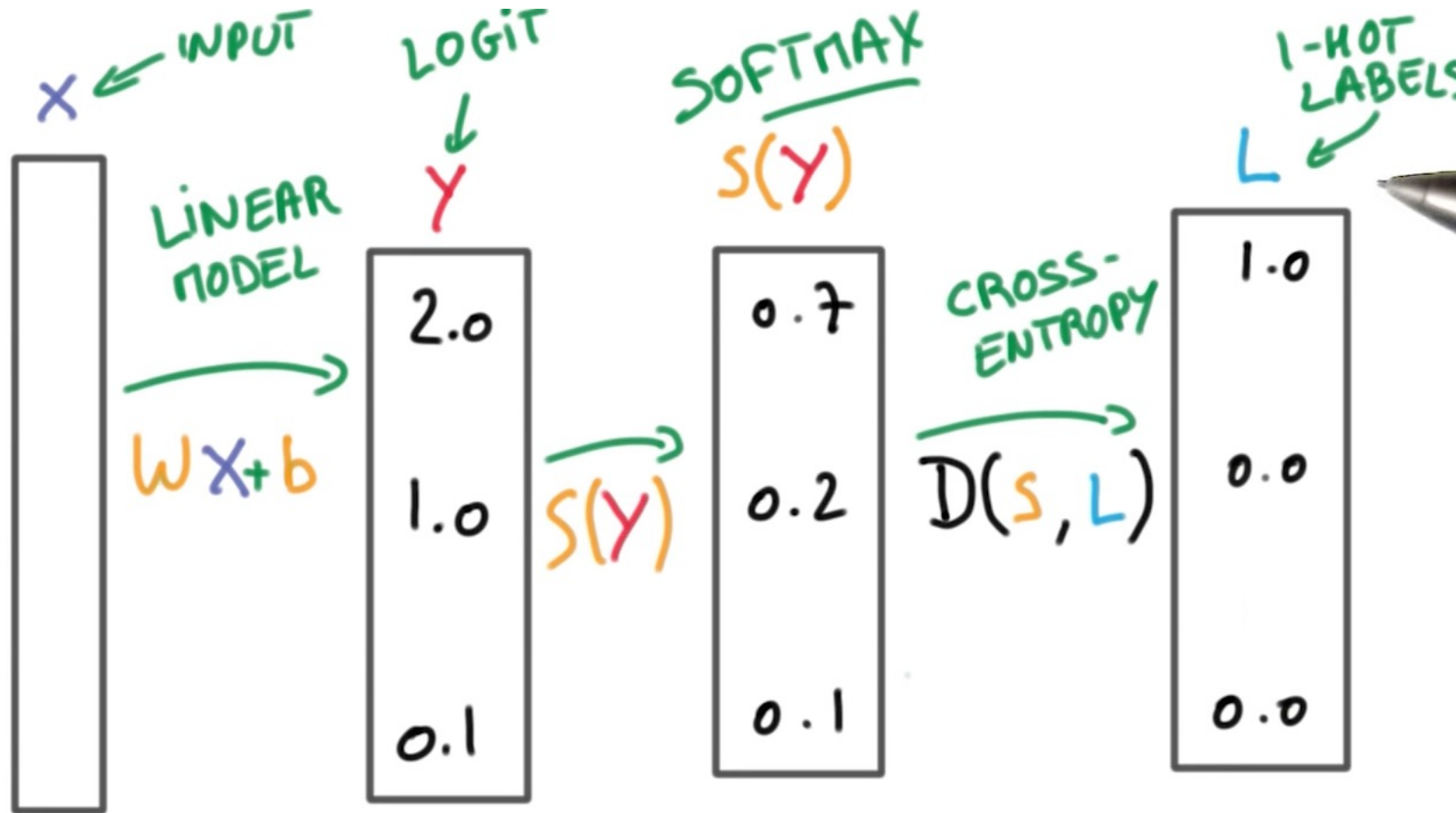
$$CE(y, \hat{y}) = - \sum_{i=1}^{N_c} y_i \log(\hat{y}_i)$$



3)을 보면 일단 아래 첫번째 줄에서 $y_i = [0, 0, 1]$ 이므로 $y_i = 1$ 일때를 제외하고 다 0이 된다.

05. Multinomial Classification (정리)

- ❑ **Cross Entropy Loss:** 우리가 예측한 확률분포값과 실제 정답을 비교 해서 손실값 산정





감사합니다.