

4. Softmax Classifier

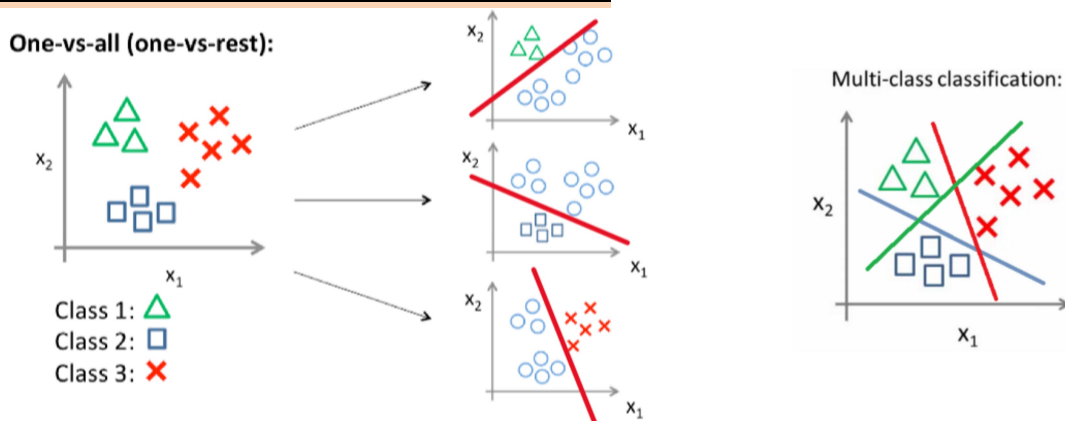
2019년 3월 29일 금요일 오후 4:43

1. Definition

- N개의 결과값을 추측할 때



- **Multinomial classification : 데이터 -> 분류**



- **Linear model : 데이터 -> 값**

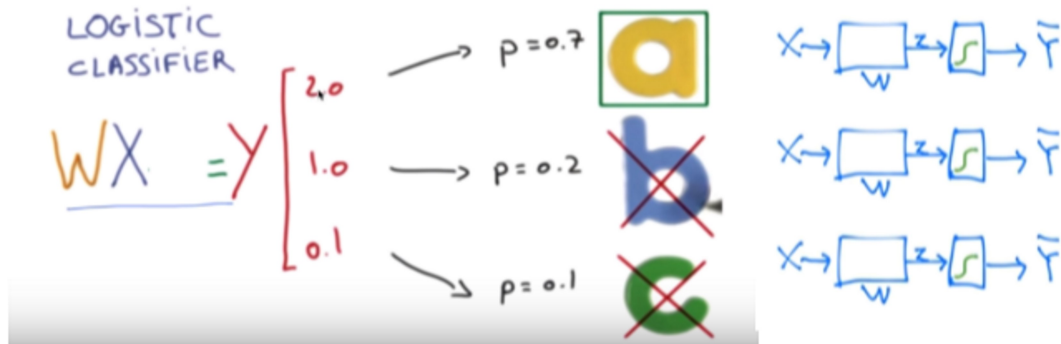
$$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \\ w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \\ w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix}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

Red arrows point from the values 2.0, 1.0, and 0.1 to three boxes containing the letters 'a', 'b', and 'c' respectively. The box with 'b' and 'c' has a red 'X' over it, indicating they are not the predicted class.



- **Sigmoid : 값 -> 확률**

- o $0 < p < 1$: 확률
- o P값의 합은 1

Sigmoid?




- Cross-entropy : 확률, 값 \rightarrow cost

\hat{y}  y 

$\begin{bmatrix} 0.1 \\ 0.5 \\ 0.4 \end{bmatrix}$
 $D(\hat{y}, y) = - \sum_j y_j \ln \hat{y}_j$
 $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

- One hot

Color		Red	Yellow	Green
Red				
Red		1	0	0
Yellow		1	0	0
Green		0	1	0
Yellow		0	0	1

2. Hypothesis

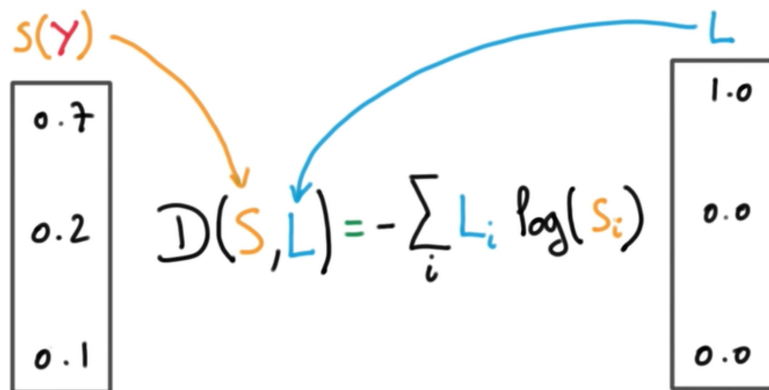
```
logits = tf.matmul(X, W) + b
hypothesis = tf.nn.softmax(logits)
```

```
hypothesis = tf.nn.softmax(tf.matmul(X, W))
```

3. Cost function

Cost function

CROSS-ENTROPY



Cross-entropy cost function

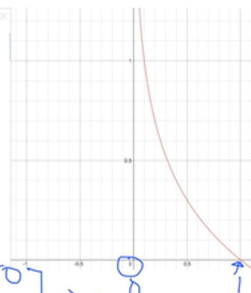
$$-\sum_i L_i \log(S_i)$$

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

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

$$\bar{Y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \xrightarrow{B(OK)}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{matrix} \xrightarrow{-\log} \\ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{matrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{bmatrix} \infty \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow 0$$

$$\bar{Y} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \xrightarrow{A(X)}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{matrix} \xrightarrow{-\log} \\ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{matrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{bmatrix} 0 \\ \infty \end{bmatrix} = \begin{bmatrix} 0 \\ \infty \end{bmatrix} \Rightarrow \infty \uparrow$$



Cross-entropy cost function

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

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

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

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

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



Logistic cost VS cross entropy

$$C(H(x), y) = y \log(H(x)) - (1 - y) \log(1 - H(x))$$
$$D(S, L) = - \sum_i L_i \log(s_i)$$

4. Gradient descent algorithm