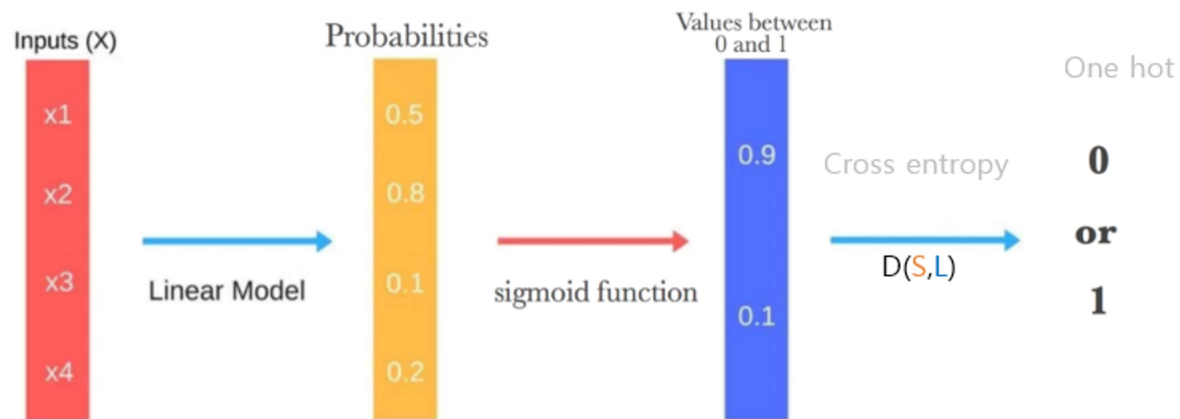


4. Softmax Classifier

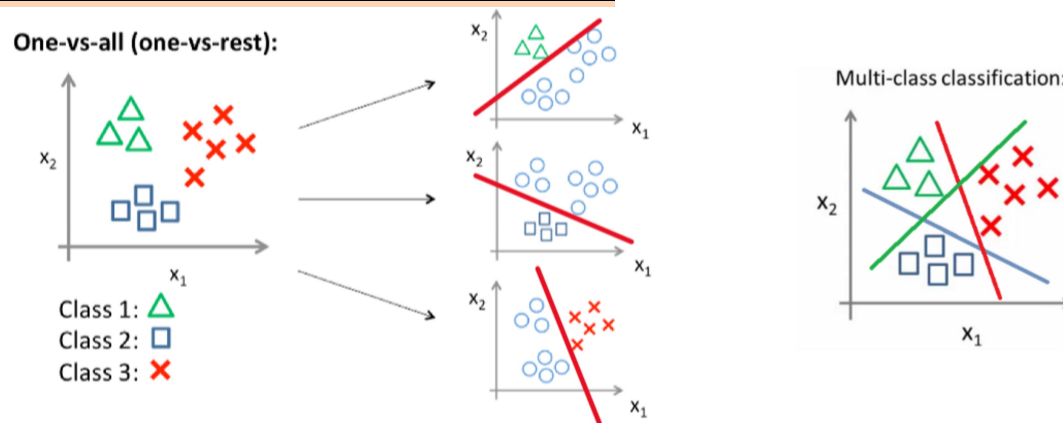
2019년 3월 29일 금요일 오후 4:43

1. Definition

- N개의 결과값을 추측할 때




- Multinomial classification : 데이터 -> 분류



- Linear model : 데이터 -> 값

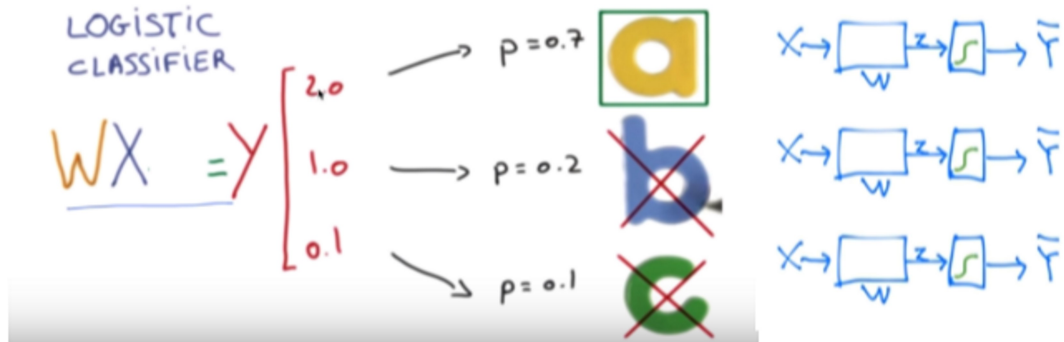
$$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \\ w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \\ w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix}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



- Sigmoid : 값 -> 확률

- $0 < p < 1$: 확률
- P값의 합은 1

Sigmoid?



- Cross-entropy : 확률, 값 -> cost

$$D(\hat{\mathbf{y}}, \mathbf{y}) = - \sum_j y_j \ln \hat{y}_j$$

- **One hot**

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

2. **Hypothesis**

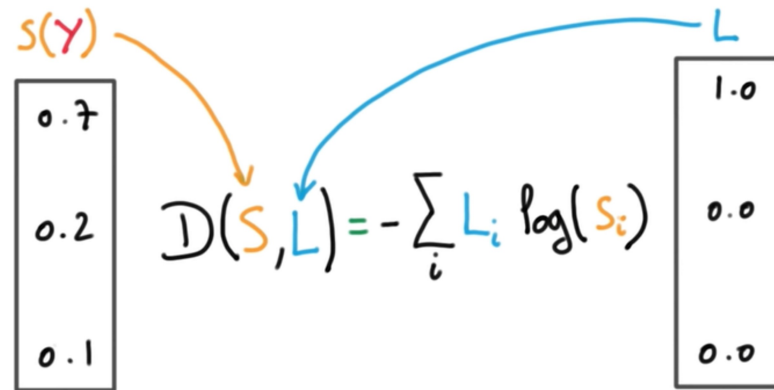
```
logits = tf.matmul(X, W) + b
hypothesis = tf.nn.softmax(logits)
```

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

3. **Cost function**

Cost function

CROSS-ENTROPY



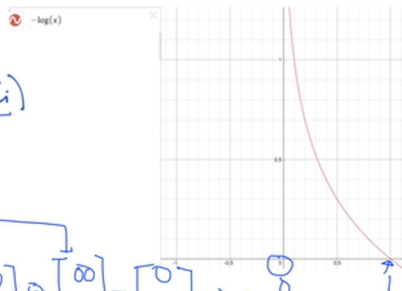
Cross-entropy cost function

$$-\sum_i L_i \log(S_i) \quad -\sum_i L_i \log(\tilde{y}_i) = \sum_i L_i \times -\log(\tilde{y}_i)$$

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

$$\underline{Y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{ (OK)}, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{matrix} -\log \\ -\log \end{matrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{bmatrix} \infty \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow 0$$

$$\underline{Y} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ (X)}, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{matrix} -\log \\ -\log \end{matrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{bmatrix} 0 \\ \infty \end{bmatrix} = \begin{bmatrix} 0 \\ \infty \end{bmatrix} \Rightarrow \infty \uparrow$$



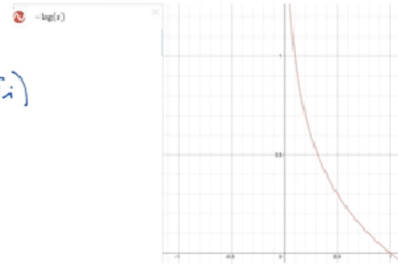
Cross-entropy cost function

$$-\sum_i L_i \log(s_i) \quad -\sum_i L_i \log(\tilde{y}_i) = \sum_i L_i \times -\log(\tilde{y}_i)$$

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

$$\tilde{Y} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{\tilde{A}} (0) \text{ ; } \begin{bmatrix} 1 \\ 0 \end{bmatrix} \odot \begin{bmatrix} 0 \\ \infty \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow 0$$

$$\tilde{Y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}_{\tilde{B}} (\infty) \text{ ; } \begin{bmatrix} 1 \\ 0 \end{bmatrix} \odot \begin{bmatrix} \infty \\ 0 \end{bmatrix} = \begin{bmatrix} \infty \\ 0 \end{bmatrix} \Rightarrow \infty \uparrow$$



Logistic cost VS cross entropy

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

$$\mathcal{D}(S, L) = -\sum_i L_i \log(s_i)$$

// ?

4. Gradient descent algorithm