

Lecture 6-I

Softmax classification:
Multinomial classification

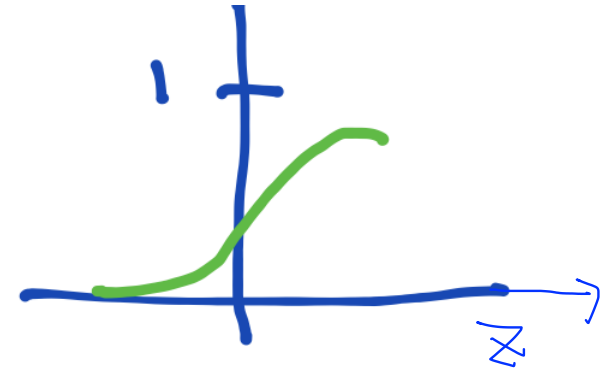
Sung Kim <hunkim+mr@gmail.com>

Linear 하기 때문에 0이나 1로 고르는 것에 적합하지 않았어

Logistic regression

$$H_L(x) = \underline{Wx} - \left\{ \begin{array}{c} 100 \\ 200 \\ 710 \end{array} \right\}$$

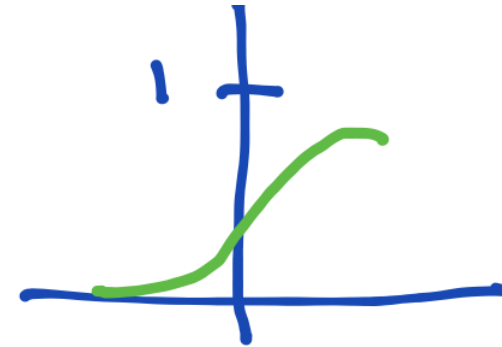
$$z = H_L(x), \quad g(z) = \begin{cases} 0 \\ 1 \end{cases}$$



Logistic regression

$$H_L(x) = Wx$$

$$z = H_L(x), \quad g(z)$$



$$g(z) = \frac{1}{1 + e^{-z}}$$

logistic 또는 sigmoid라고 부름

$$H_R(x) = g(H_L(x))$$

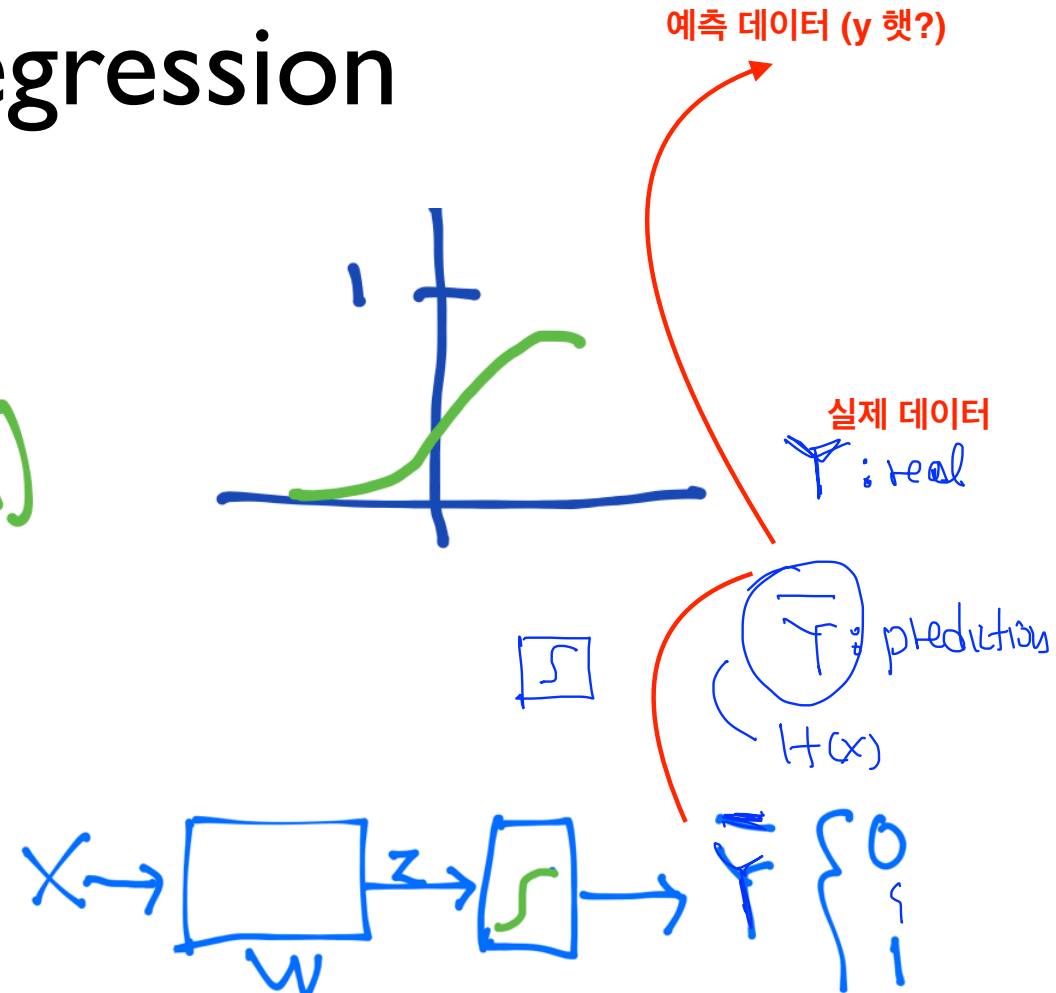
Logistic regression

$$H_L(x) = Wx$$

$$z = H_L(x), \quad g(z)$$

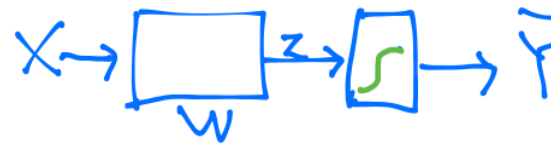
$$g(z) = \frac{1}{1 + e^{-z}}$$

$$H_R(x) = g(H_L(x))$$



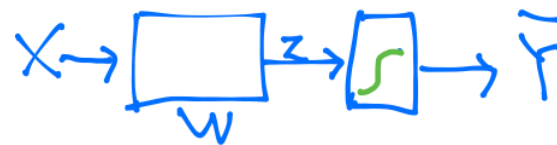
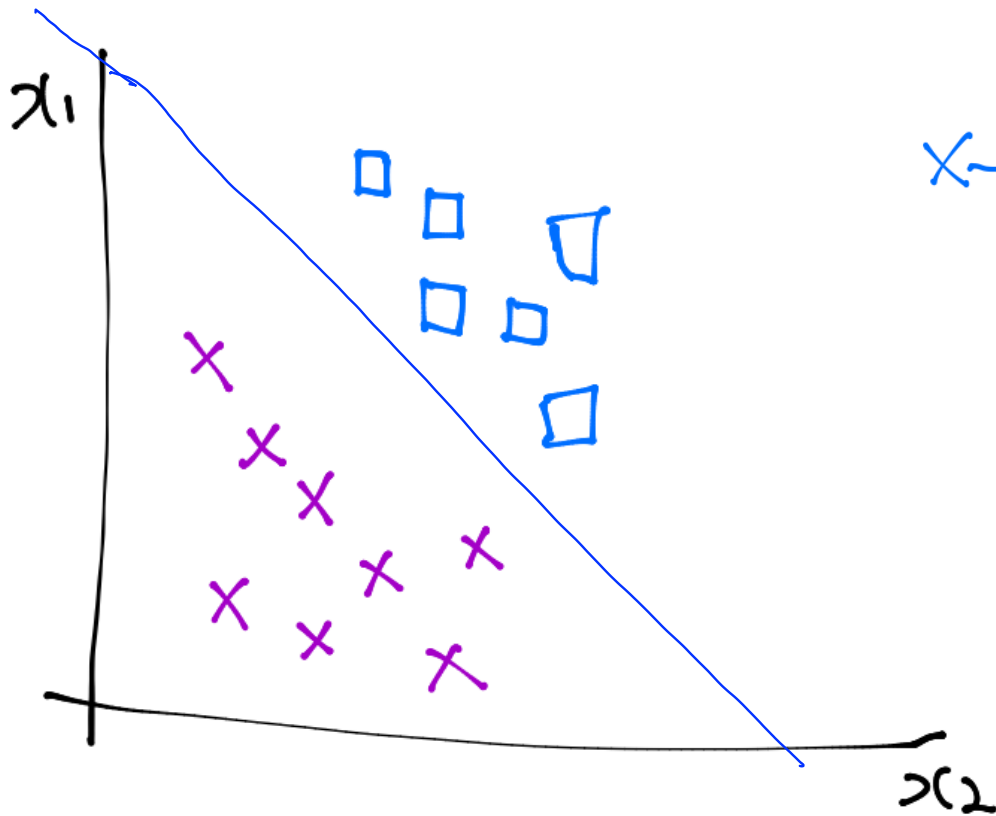
Logistic regression

$$g(z) = \frac{1}{1 + e^{-z}} \quad H_R(x) = g(H_L(x))$$



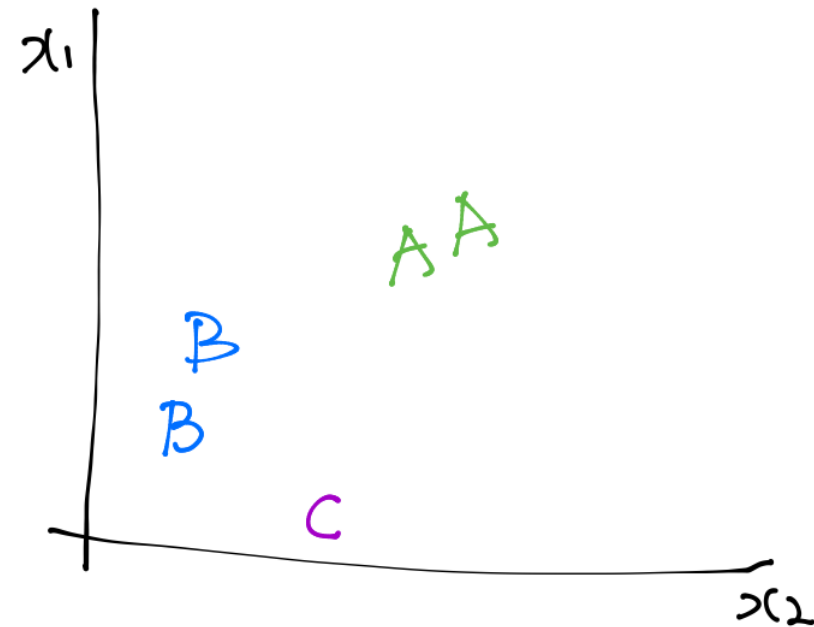
Logistic regression

$$g(z) = \frac{1}{1 + e^{-z}} \quad H_R(x) = g(H_L(x))$$

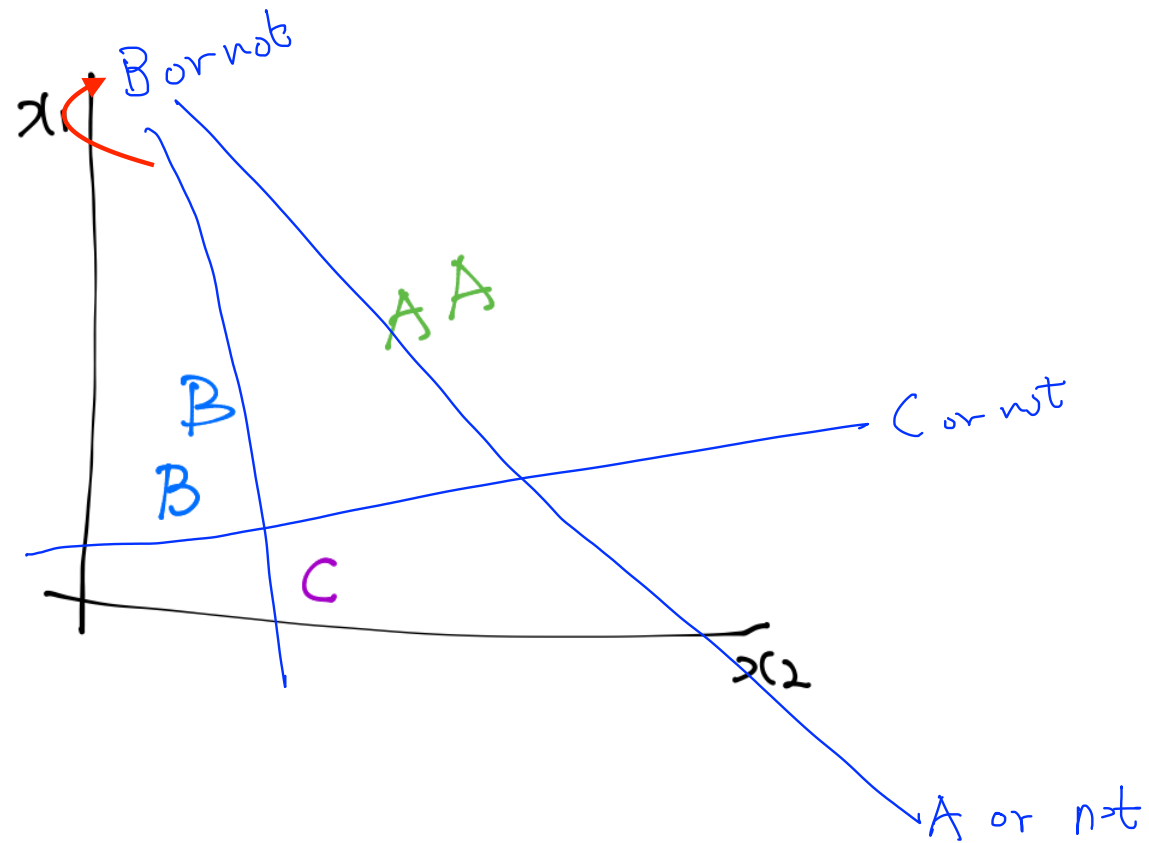


Multinomial classification

x1 (hours)	x2 (attendance)	y (grade)
10	5	A
9	5	A
3	2	B
2	4	B
11	1	C

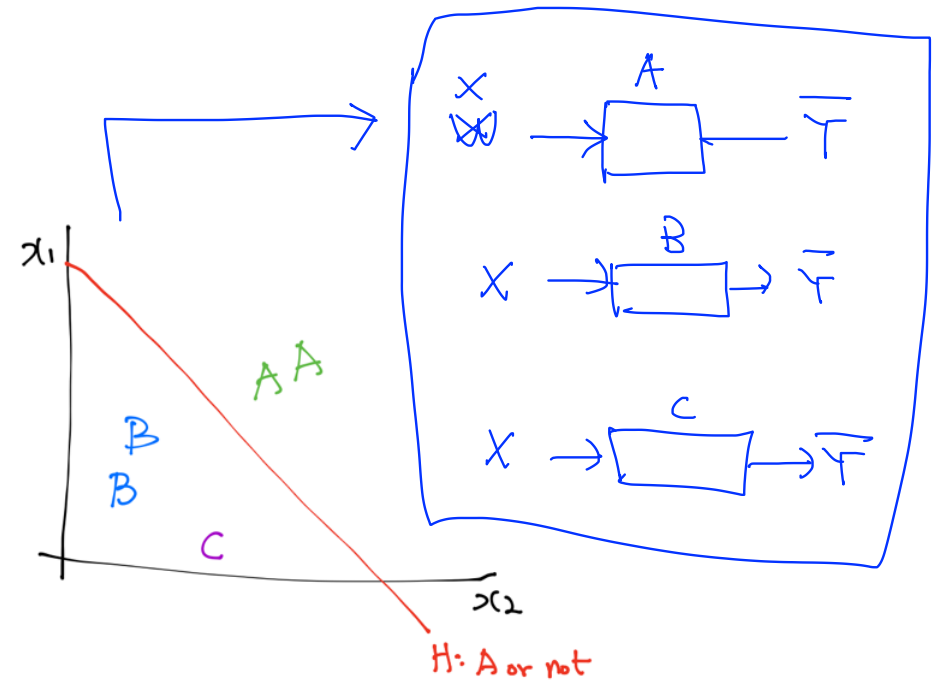
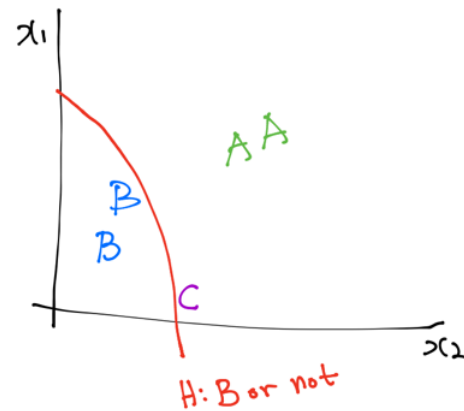
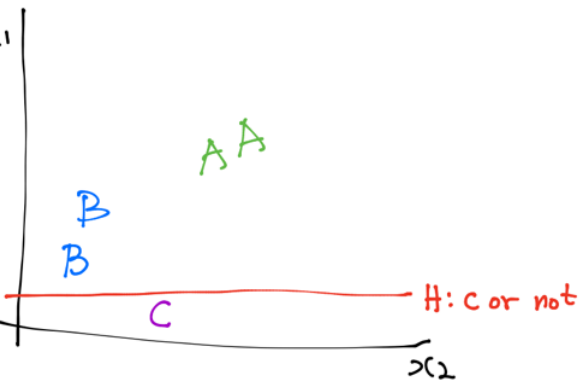


Multinomial classification

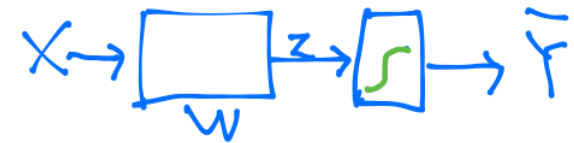
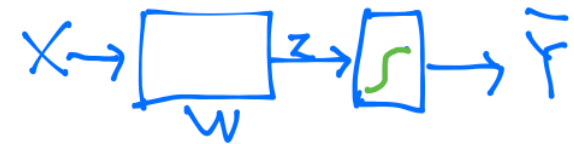
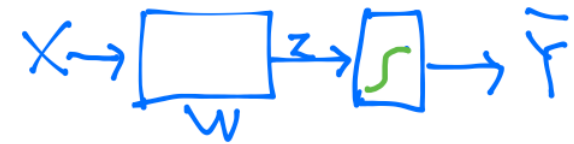


Multinomial classification

x가 주어졌을 때 각각 A, B, C
인지 아닌지 판단하는 독립된
classifier



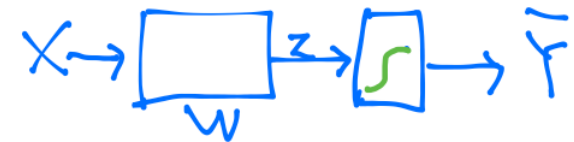
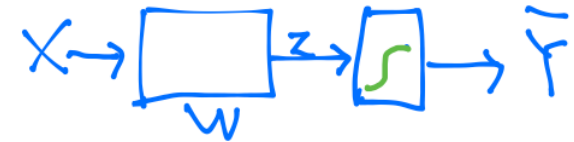
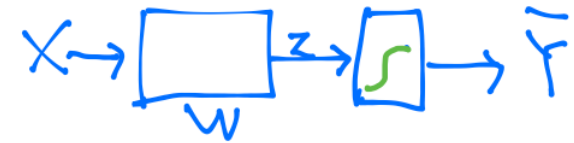
Multinomial classification



실제로 구현할 때는 행렬로

Multinomial classification

$$\begin{matrix} [w_1 & w_2 & w_3] \\ \quad \quad \quad \omega \end{matrix} \begin{matrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ \quad \quad \quad X \end{matrix} = \begin{matrix} \underbrace{[w_1 x_1 + w_2 x_2 + w_3 x_3]}_{\quad \quad \quad H(X)} \end{matrix}$$

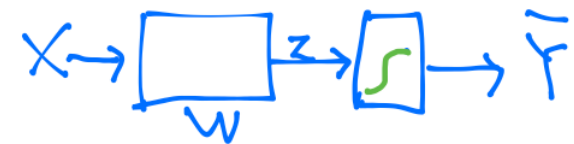
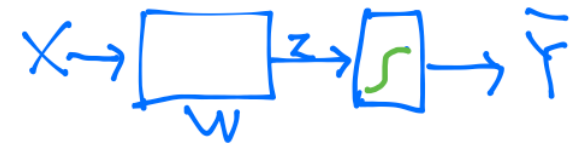
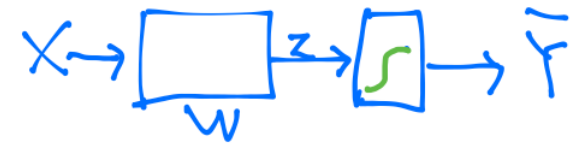


Multinomial classification

$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = [w_1 x_1 + w_2 x_2 + w_3 x_3]$$

$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = [w_1 x_1 + w_2 x_2 + w_3 x_3]$$

$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = [w_1 x_1 + w_2 x_2 + w_3 x_3]$$



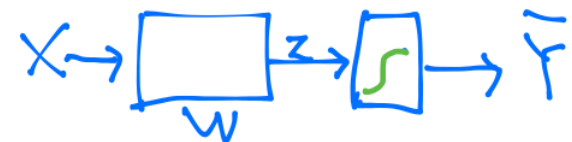
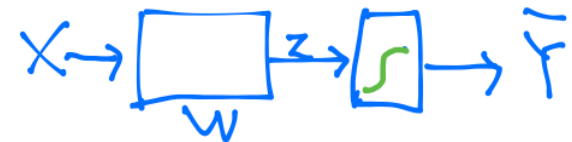
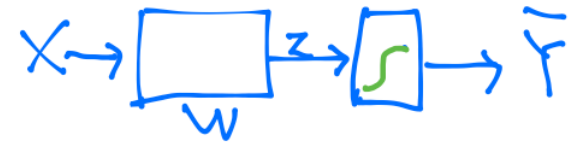
Multinomial classification

각각 구현하면 복잡해서 하나로 합친 후 구현

$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = [w_1 x_1 + w_2 x_2 + w_3 x_3]$$



$$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \hat{y}$$



Matrix multiplication

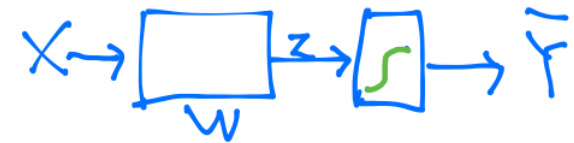
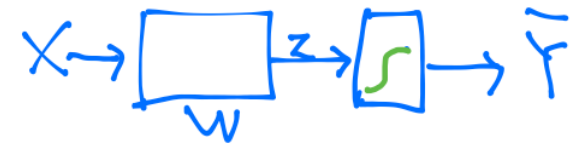
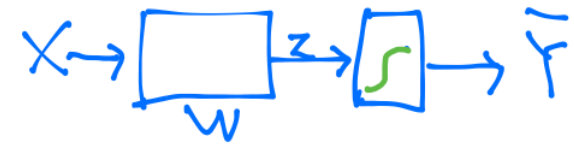
The diagram illustrates the dot product of two matrices. The first matrix is $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ and the second matrix is $\begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}$. A yellow curved arrow labeled "Dot Product" connects the first row of the first matrix to the first column of the second matrix. The result is shown as $\begin{bmatrix} 58 \end{bmatrix}$. The numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 58 are highlighted in yellow.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix} = \begin{bmatrix} 58 \end{bmatrix}$$

<https://www.mathsisfun.com/algebra/matrix-multiplying.html>

Multinomial classification

$$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \\ w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \\ w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix}$$



Multinomial classification

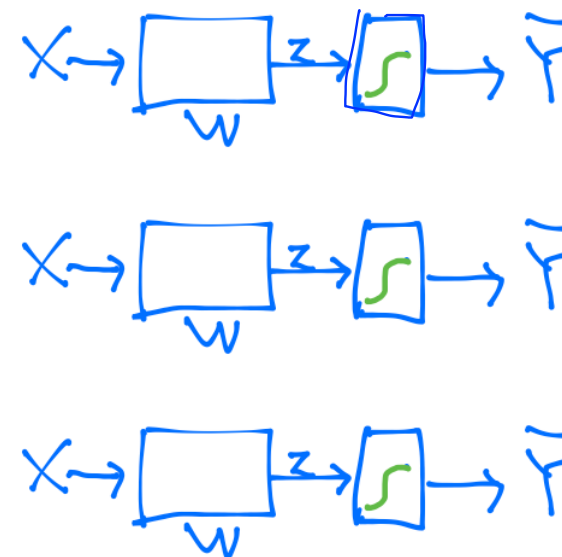
$$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \\ w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \\ w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix}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

$H_A(x)$
 $H_B(x)$
 $H_C(x)$

$X \rightarrow \boxed{W} \xrightarrow{z} \boxed{\sigma} \rightarrow \hat{Y}$
 $X \rightarrow \boxed{W} \xrightarrow{z} \boxed{\sigma} \rightarrow \hat{Y}$
 $X \rightarrow \boxed{W} \xrightarrow{z} \boxed{\sigma} \rightarrow \hat{Y}$

Where is sigmoid?

각각 sigmoid를 적용하면 돼

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


The diagram illustrates the process of applying the sigmoid function to the outputs of a linear layer. It shows three parallel paths, each representing a different output node (A, B, and C). Each path starts with an input vector X entering a box labeled W , which represents the linear transformation. The output of this box is a scalar value z . This value z then enters a second box containing a green sigmoid function symbol σ . The final output of each path is \hat{Y} , which is the result of applying the sigmoid function to z . A red arrow points from the text '각각 sigmoid를 적용하면 돼' (Apply sigmoid to each) to the sigmoid boxes in the diagram.