Lecture 06

Softmax classification: Multinomial classification

Logistic regression

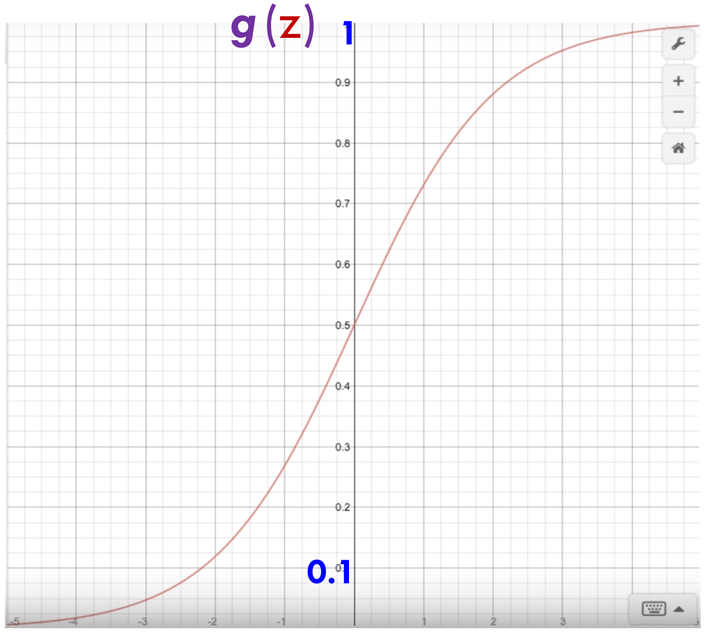
$H_L(x) = Wx$ ➡ 자연수

$z = H_L(x)$, $g(z)$ ➡ 0~1

z

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

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



Logistic regression

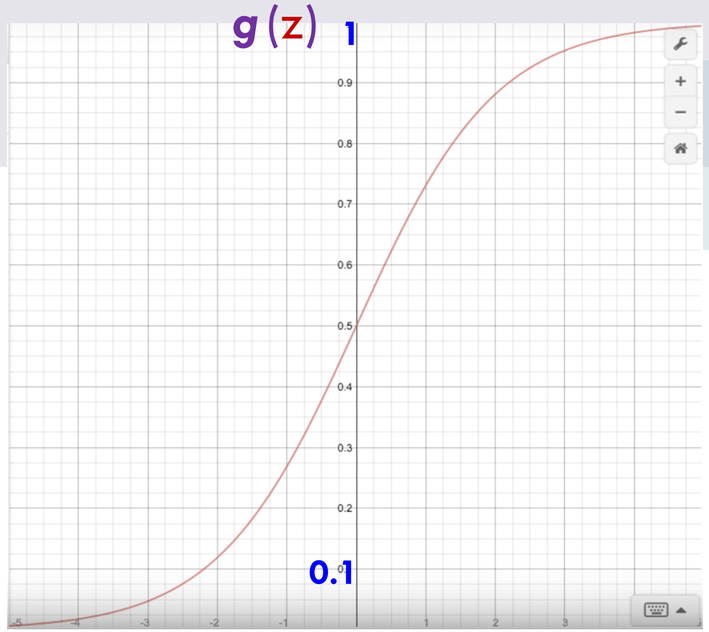
$H_L(x) = Wx$ ➡ 자연수

$z = H_L(x)$, $g(z)$ ➡ 0~1

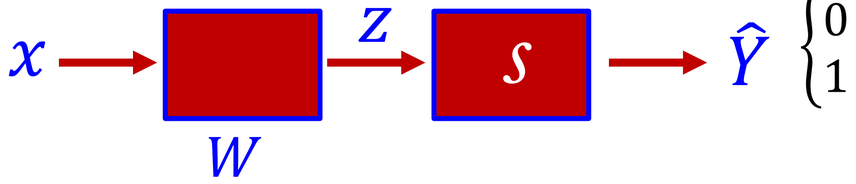
Z

$g(Z)$

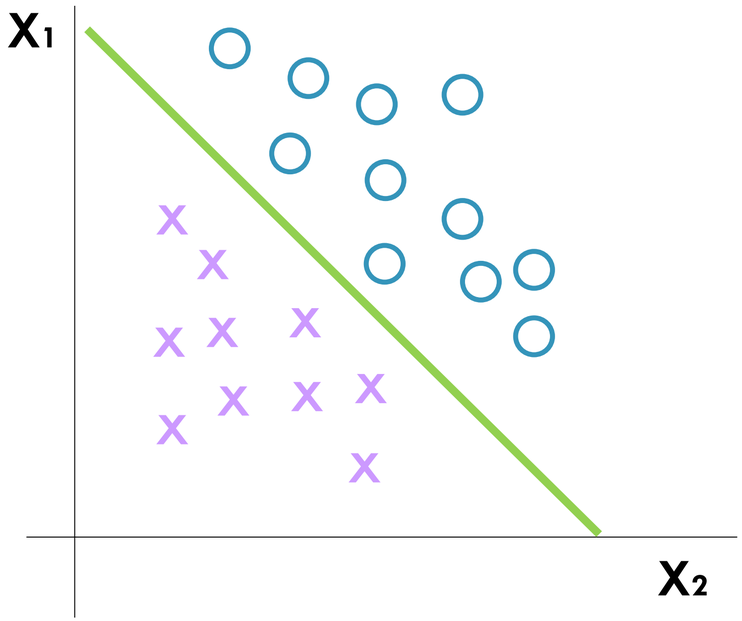
$$g(z) = \frac{1}{(1 + e^{-z})}$$



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

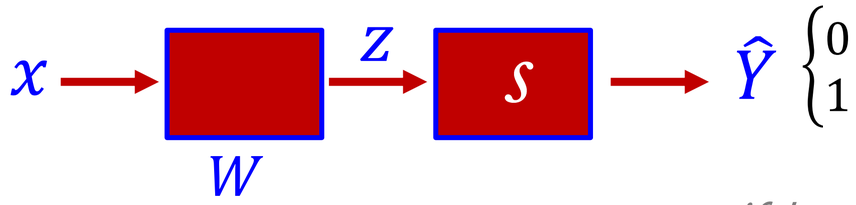


Logistic regression



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

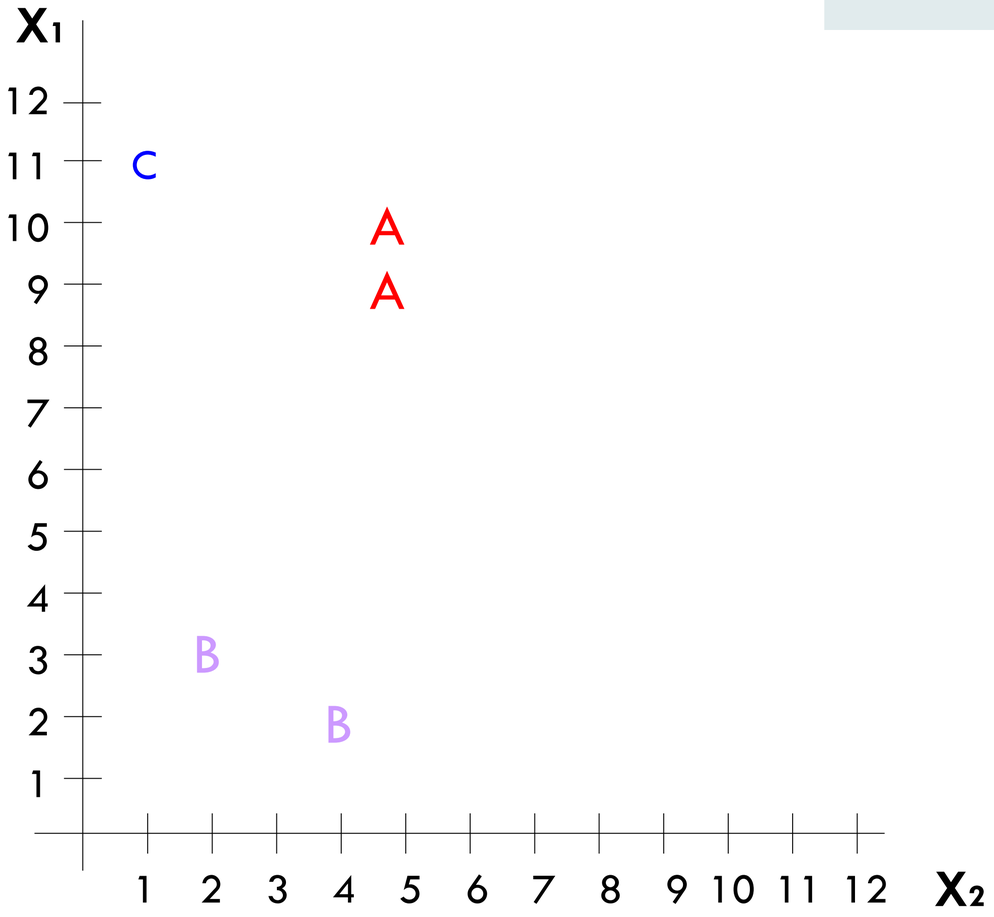
$$H_R(x) = g(H_L(x))$$



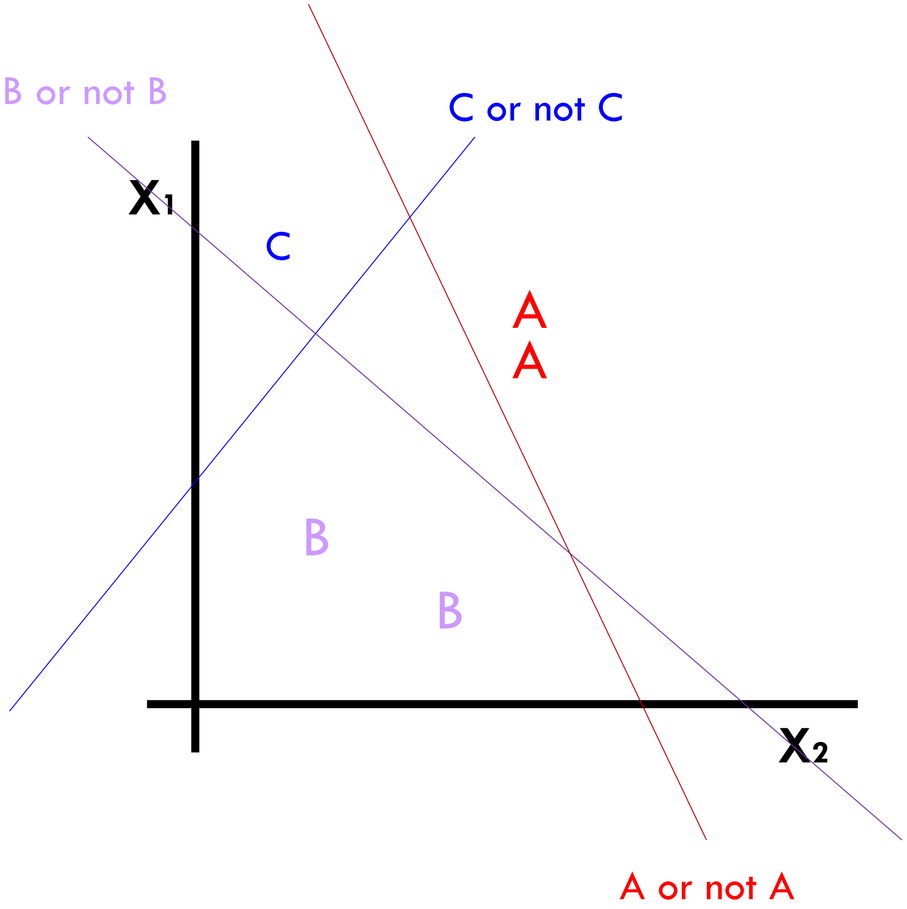
if hypothesis>0.5 :
then 1
else 0

Multinomial classification

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

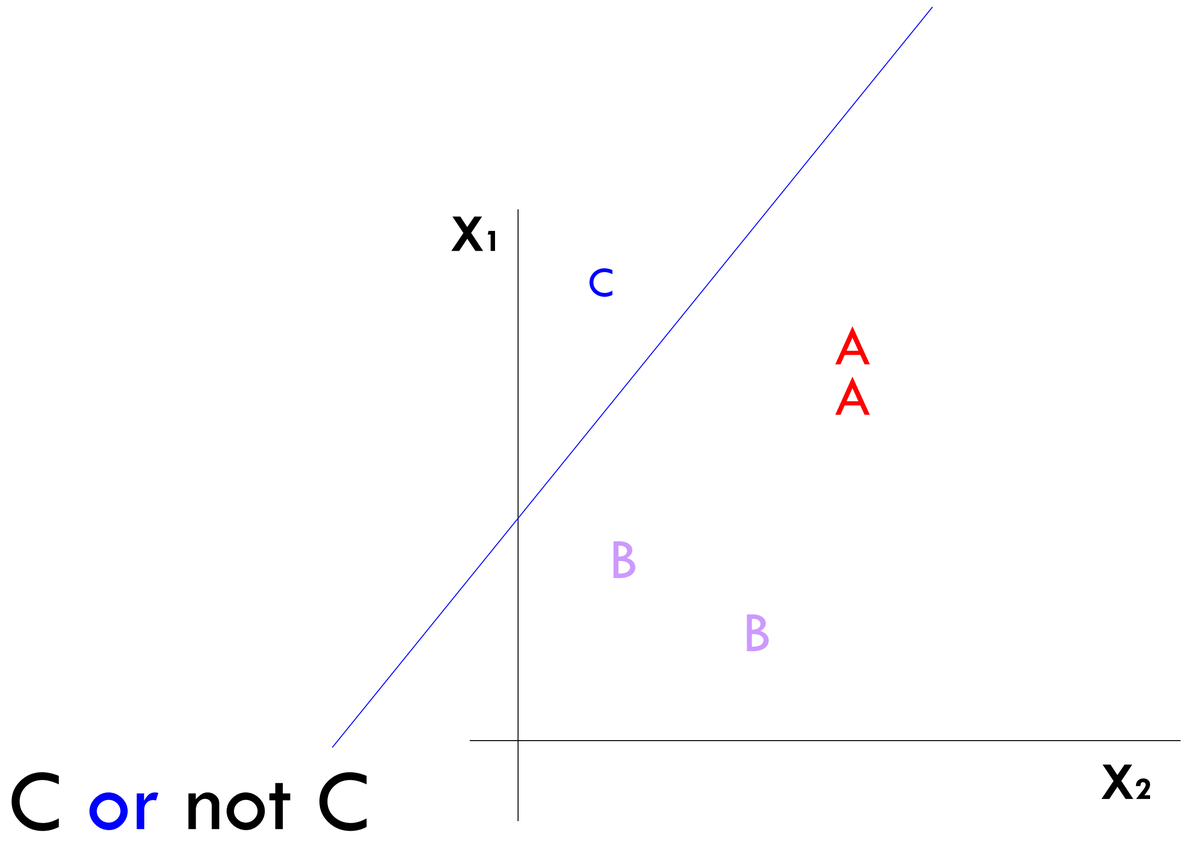


Multinomial classification

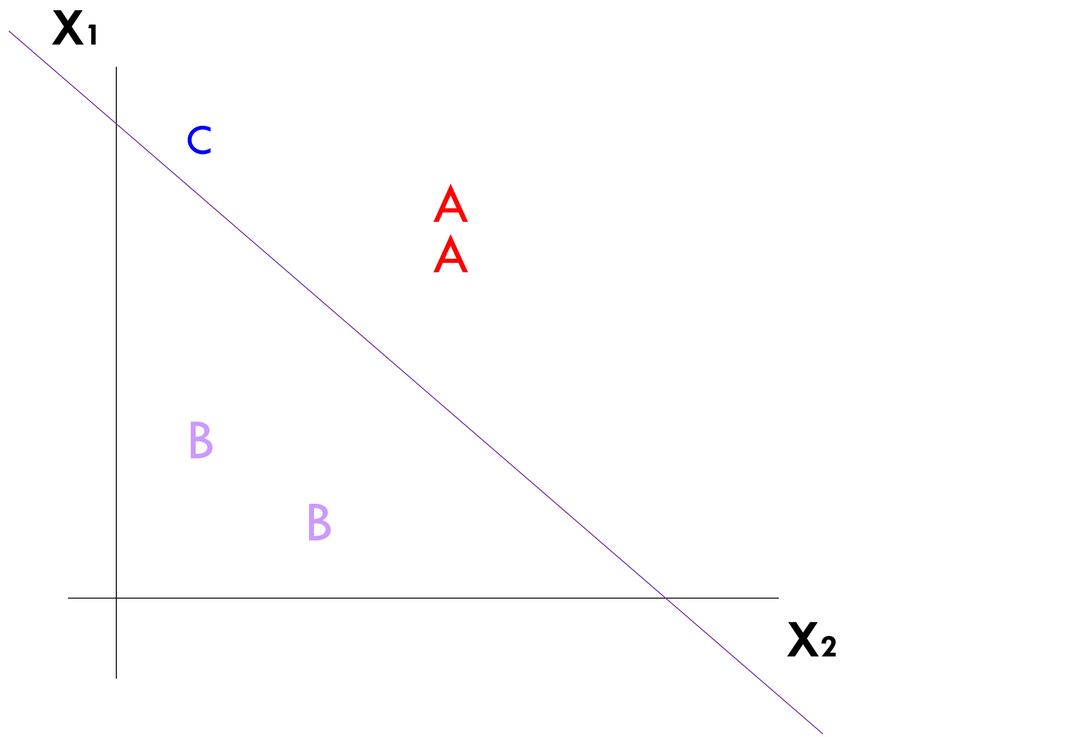


Binary classification

Multinomial classification



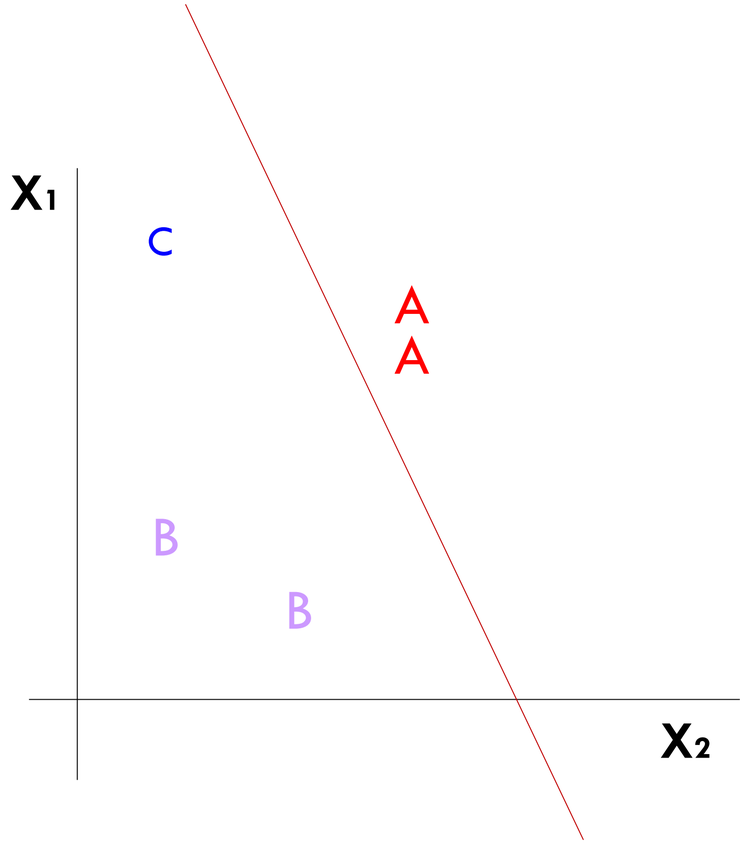
Multinomial classification



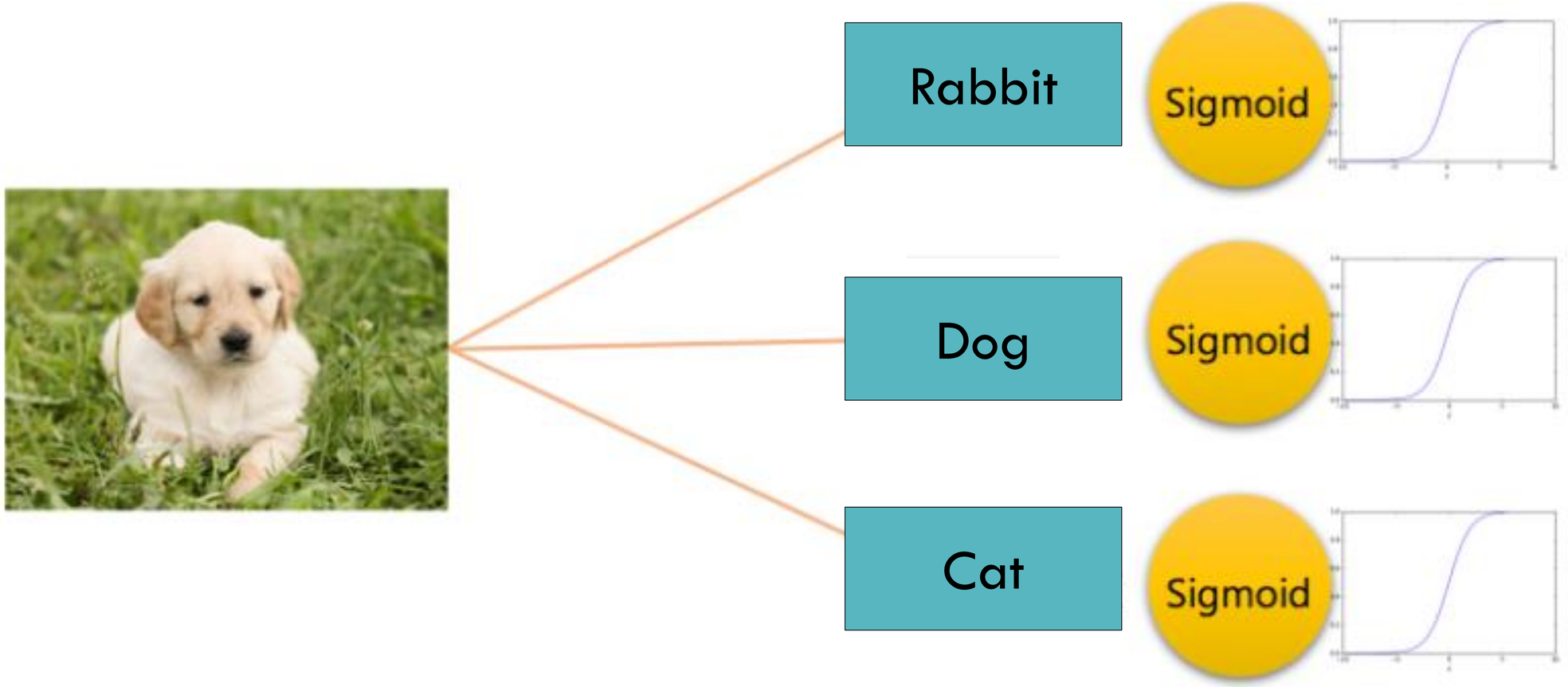
B or not B

Multinomial classification

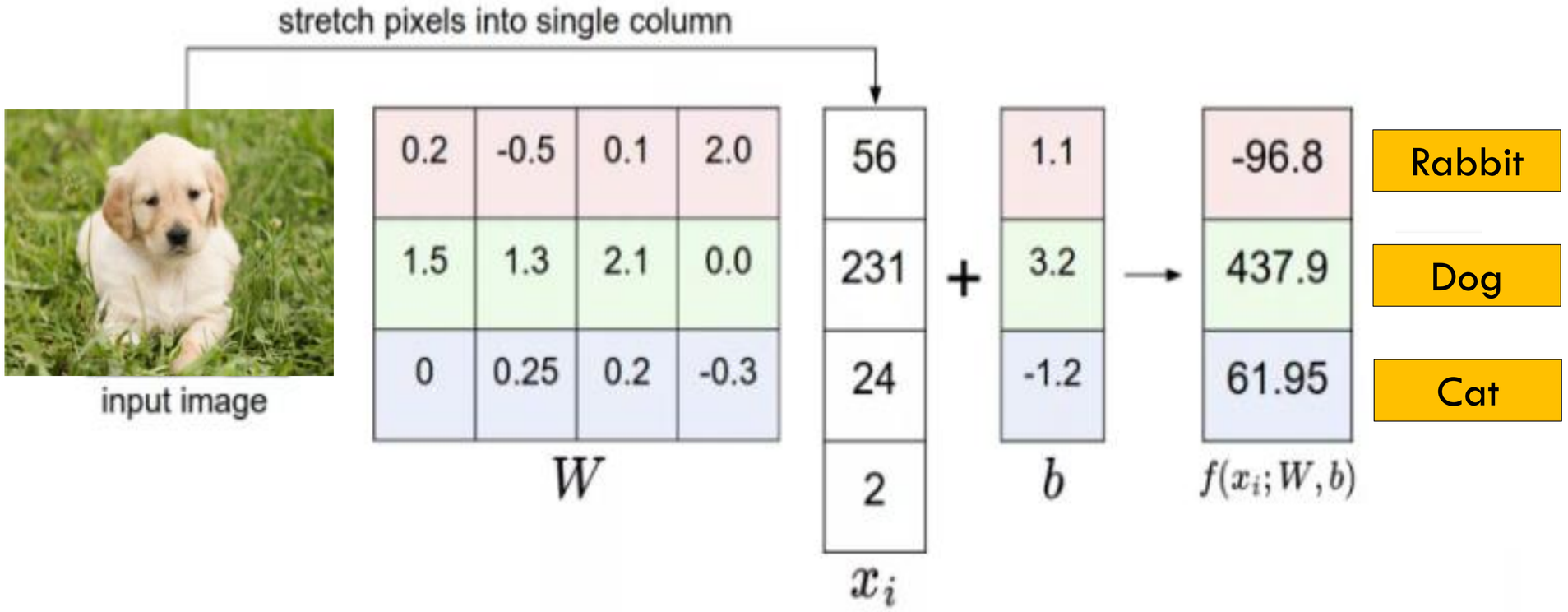
A or not A



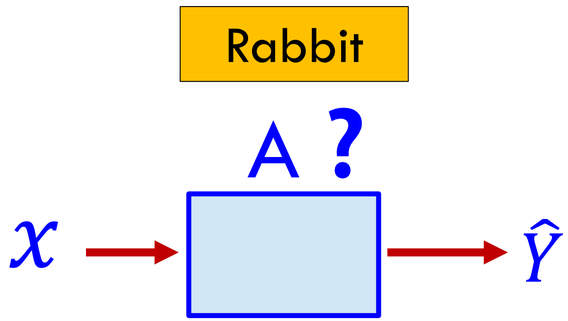
Multinomial classification

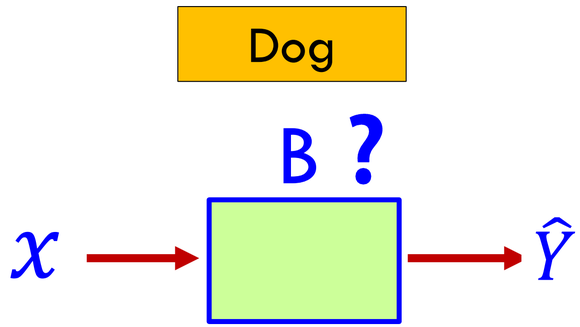


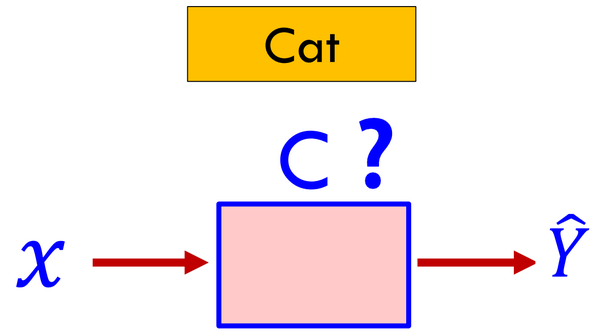
Multinomial classification

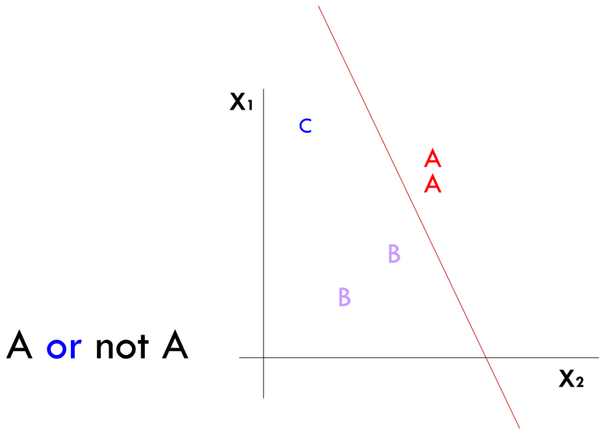


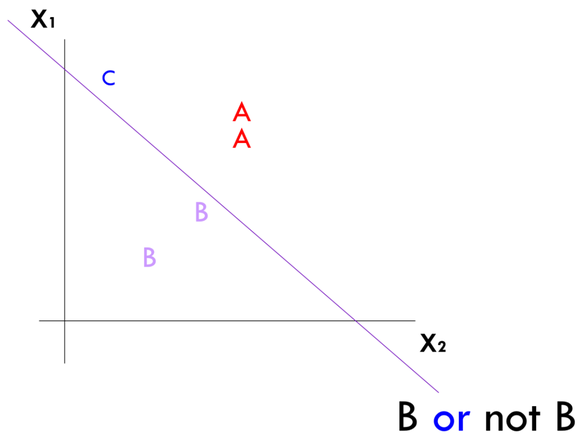
Multinomial classification

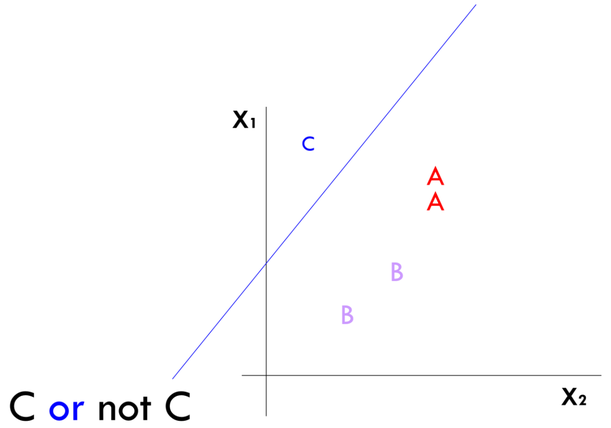








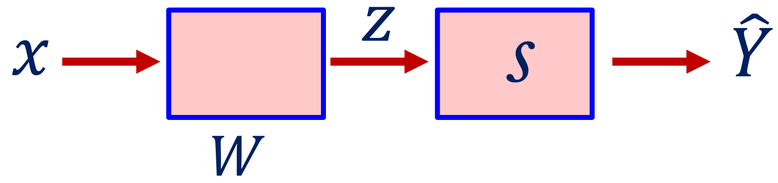




Multinomial classification

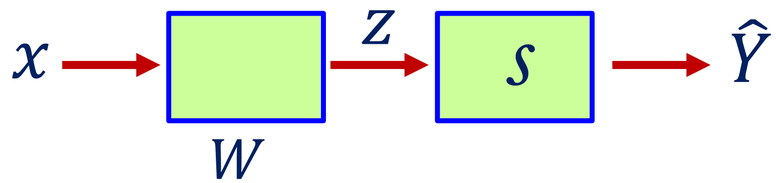
Rabbit

$$[WA1\ WA2\ WA3]\begin{bmatrix}x1\\x2\\x3\end{bmatrix} = [WA1x1 + WA2x2 + WA3x3]$$



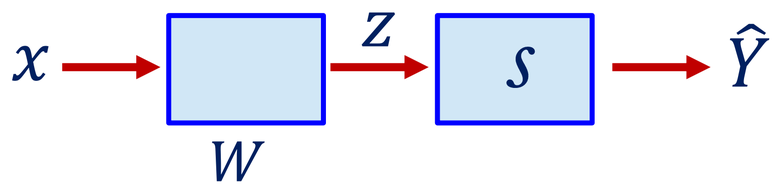
Dog

$$[WB1\ WB2\ WB3]\begin{bmatrix}x1\\x2\\x3\end{bmatrix} = [WB1x1 + WB2x2 + WB3x3]$$



Cat

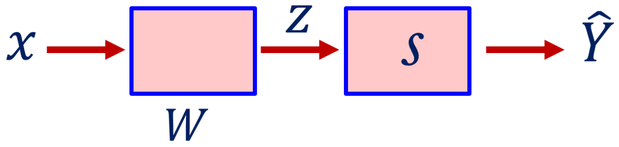
$$[WC1\ WC2\ WC3]\begin{bmatrix}x1\\x2\\x3\end{bmatrix} = [WC1x1 + WC2x2 + WC3x3]$$



Multinomial classification

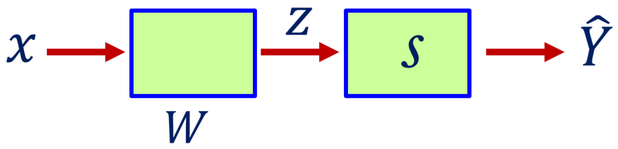
Rabbit

$$[WA1\ WA2\ WA3]\begin{bmatrix}x1\\x2\\x3\end{bmatrix} = [WA1x1 + WA2x2 + WA3x3]$$



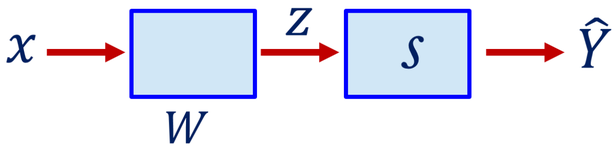
Dog

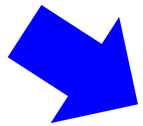
$$[WB1\ WB2\ WB3]\begin{bmatrix}x1\\x2\\x3\end{bmatrix} = [WB1x1 + WB2x2 + WB3x3]$$



Cat

$$[WC1\ WC2\ WC3]\begin{bmatrix}x1\\x2\\x3\end{bmatrix} = [WC1x1 + WC2x2 + WC3x3]$$





Rabbit
Dog
Cat

$$\begin{bmatrix}W_{A1} & W_{A2} & W_{A3}\\ W_{B1} & W_{B2} & W_{B3}\\ W_{C1} & W_{C2} & W_{C3}\end{bmatrix}\begin{bmatrix}x1\\x2\\x3\end{bmatrix} =$$

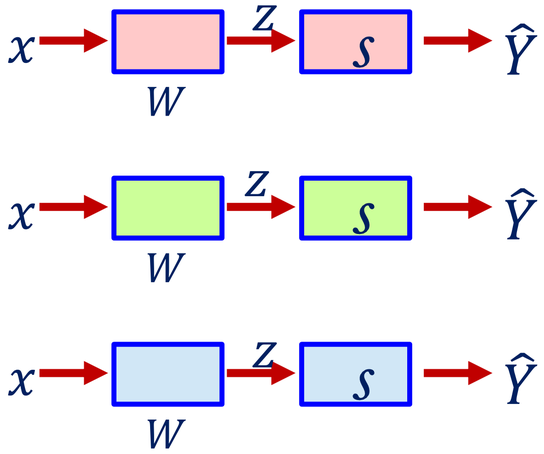
Multinomial classification

Rabbit
Dog
Cat

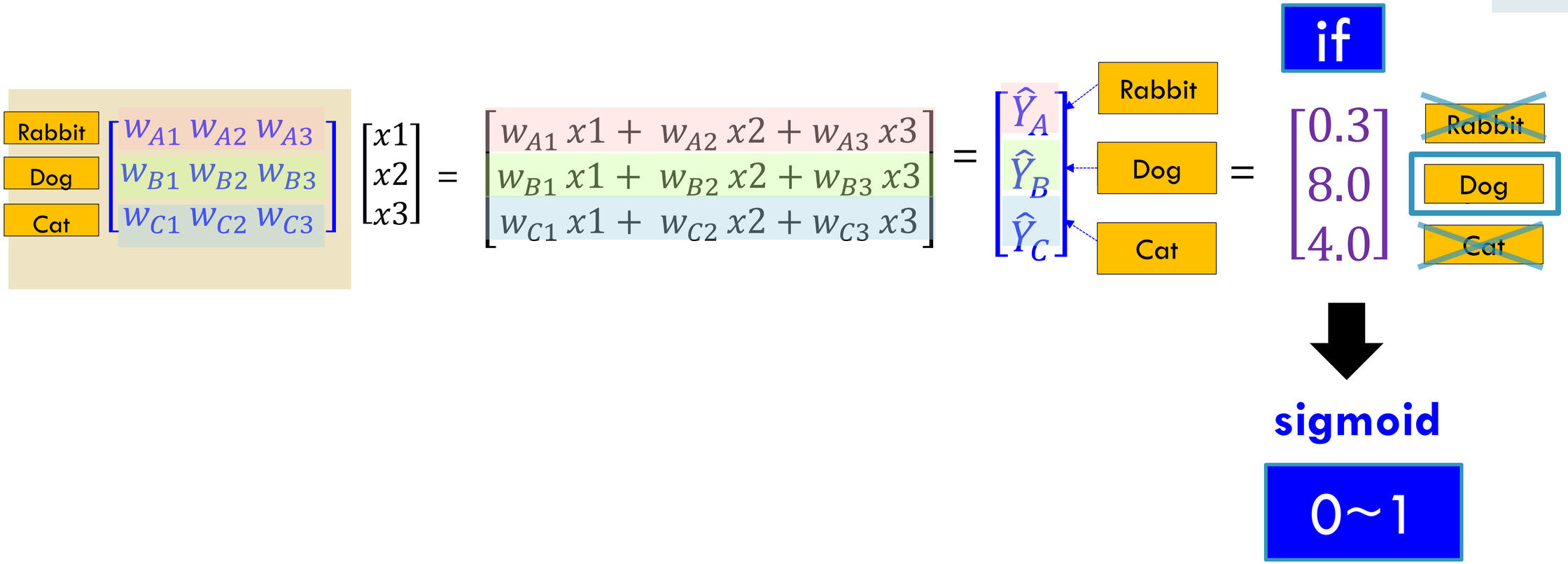
$$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x1 \\ x2 \\ x3 \end{bmatrix} =$$

$$\begin{bmatrix} w_{A1}\,x1 + w_{A2}\,x2 + w_{A3}\,x3 \\ w_{B1}\,x1 + w_{B2}\,x2 + w_{B3}\,x3 \\ w_{C1}\,x1 + w_{C2}\,x2 + w_{C3}\,x3 \end{bmatrix} = \begin{bmatrix} \hat{Y}_A \\ \hat{Y}_B \\ \hat{Y}_C \end{bmatrix}$$

Rabbit ← $H_A(x)$
Dog ← $H_B(x)$
Cat ← $H_C(x)$



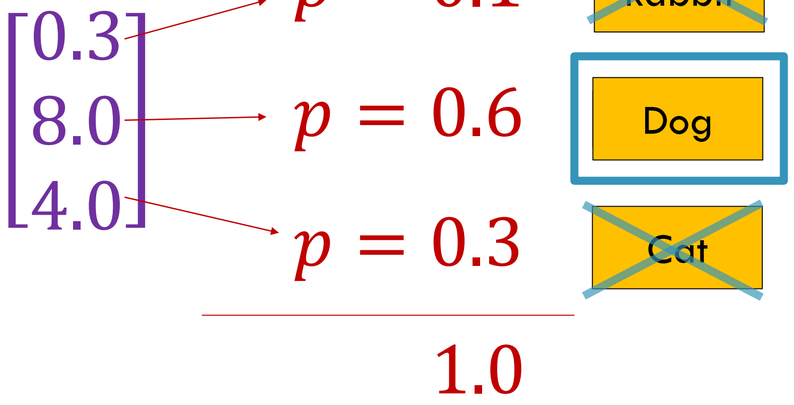
Where is sigmoid?



Sigmoid?

Logistic Classifier

$$WX = \hat{Y}$$

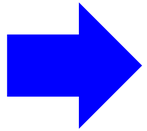


Sigmoid?

SOFTMAX

if $\hat{Y}$

$$y \begin{bmatrix} 0.3 \\ 8.0 \\ 4.0 \end{bmatrix}$$



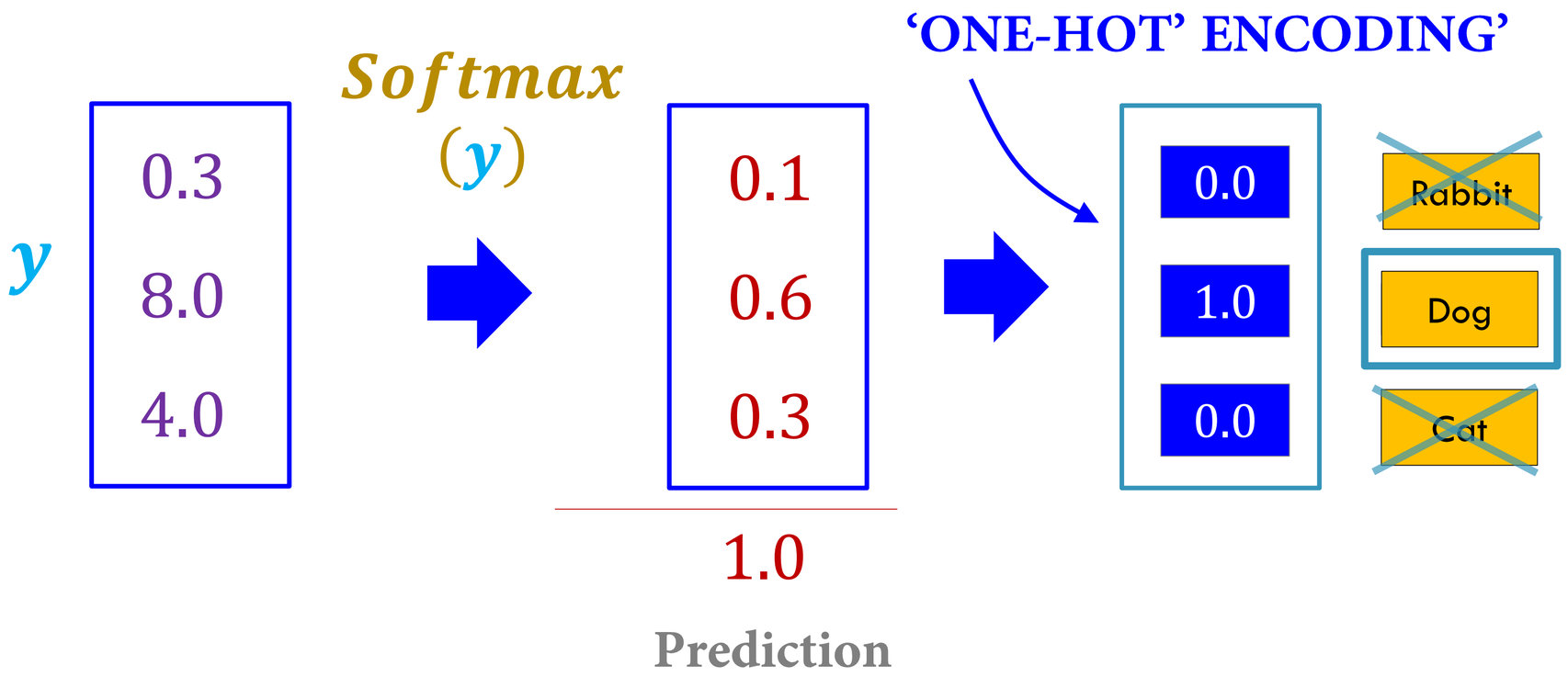
SOFTMAX

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

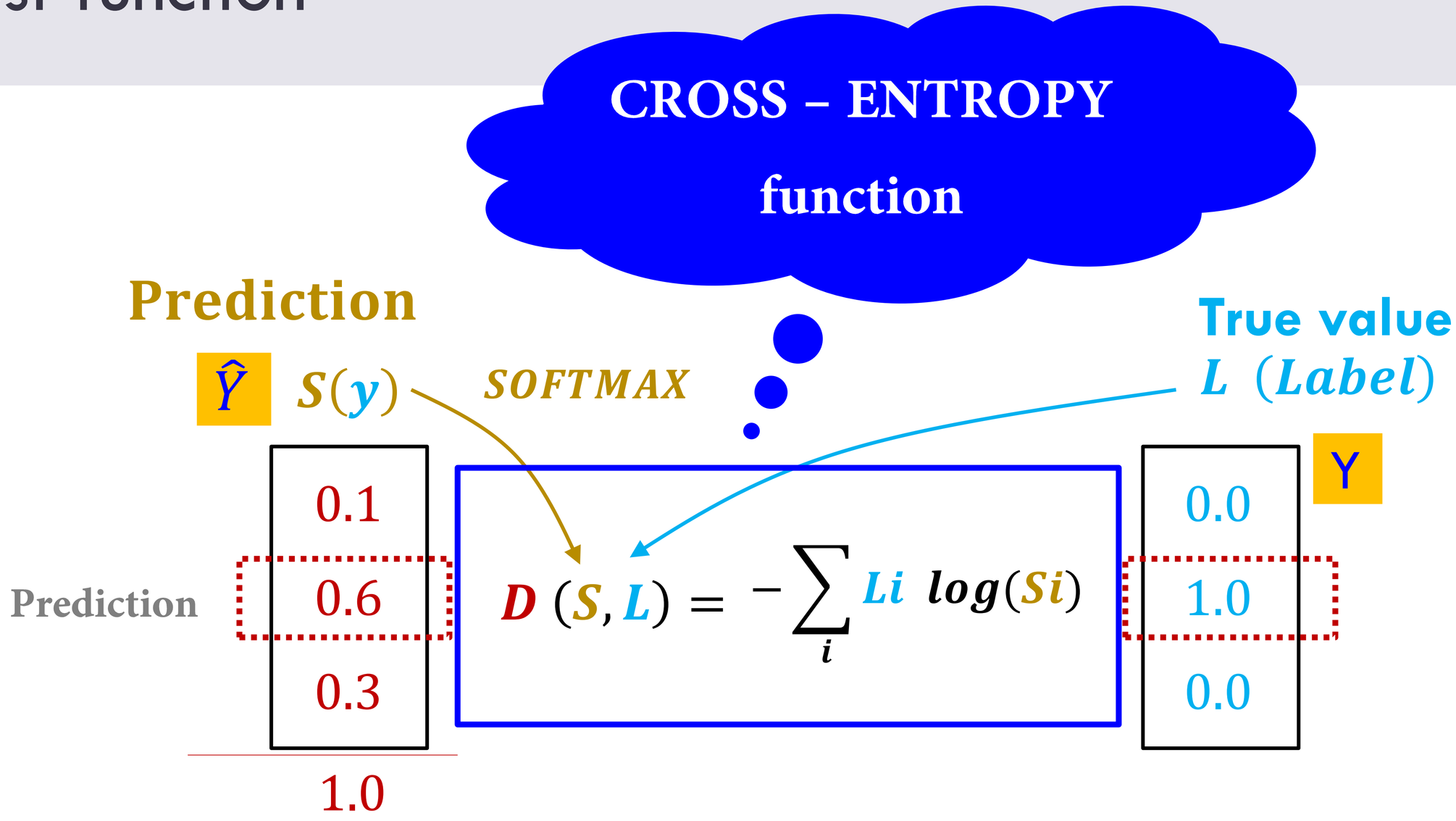
PROBABILITIES

$p = 0.1$	Rabbit	➡	0.0
$p = 0.6$	Dog	➡	1.0
$p = 0.3$	Cat	➡	0.0
1.0			

Cost function



Cost function



Cross-entropy cost function

$$-\sum_i Li\ log(Si) = -\sum_i Li\ log(\hat{Y}_i) = \sum_i Li \cdot \left(-log(\hat{Y}_i)\right)$$

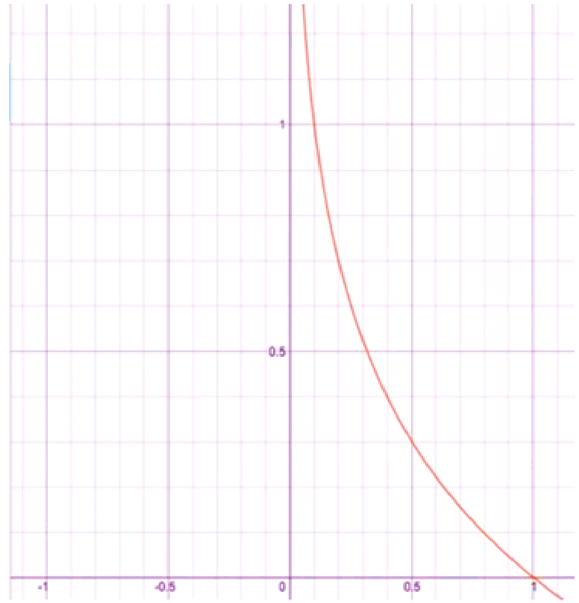
$$L = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = B$$

A

B

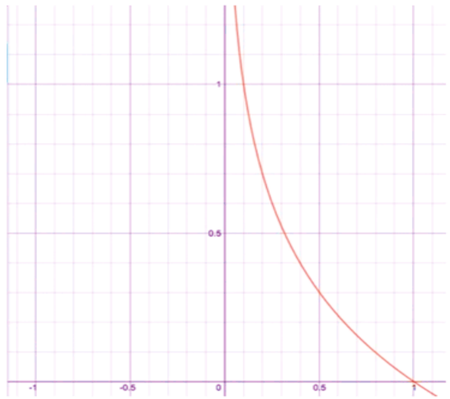
$$\hat{Y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = B \quad (O)$$

$$\hat{Y} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = A \quad (X)$$



Cross-entropy cost function

$$-\sum_i Li \ log(Si) = -\sum_i Li \ log(\hat{Y}_i) = \sum_i Li \cdot \left(-log(\hat{Y}_i)\right)$$



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

$$\hat{Y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = B \quad (O)$$

$$\hat{Y} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = A \quad (X)$$

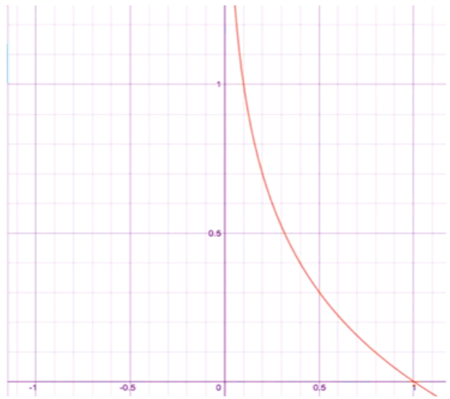
$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot -log \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{bmatrix} \infty \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0$$

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot -log \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{bmatrix} 0 \\ \infty \end{bmatrix} = \begin{bmatrix} 0 \\ \infty \end{bmatrix} = \infty$$

cost function

Cross-entropy cost function

$$-\sum_i Li\ log(Si) = -\sum_i Li\ log(\hat{Y}_i) = \sum_i Li\ \cdot \left(-log(\hat{Y}_i)\right)$$



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

(A: 1, B: 0)

$$\hat{Y} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = A \quad (O)$$

$$\hat{Y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = B \quad (X)$$

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \odot -log\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \odot \begin{bmatrix} 0 \\ \infty \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0$$

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \odot -log\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \odot \begin{bmatrix} \infty \\ 0 \end{bmatrix} = \begin{bmatrix} \infty \\ 0 \end{bmatrix} = \infty$$

cost function

Logistic cost VS cross entropy

$$cost(W) = \frac{1}{m}\sum \quad c(H(x), y)$$

$$c(H(x), y) = \begin{cases} -log(H(x)) & : y = 1 \\ -log(1 - H(x)) & : y = 0 \end{cases}$$

=

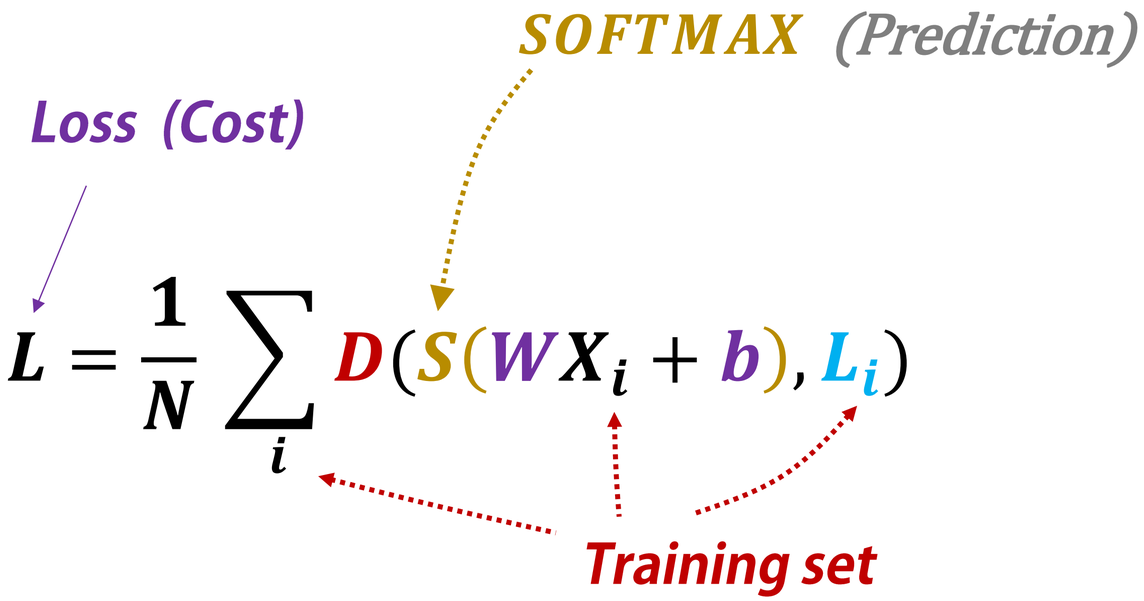
$$c(H(x), y) = -\,ylog(H(x)) - (1 - y)log(1 - H(x))$$

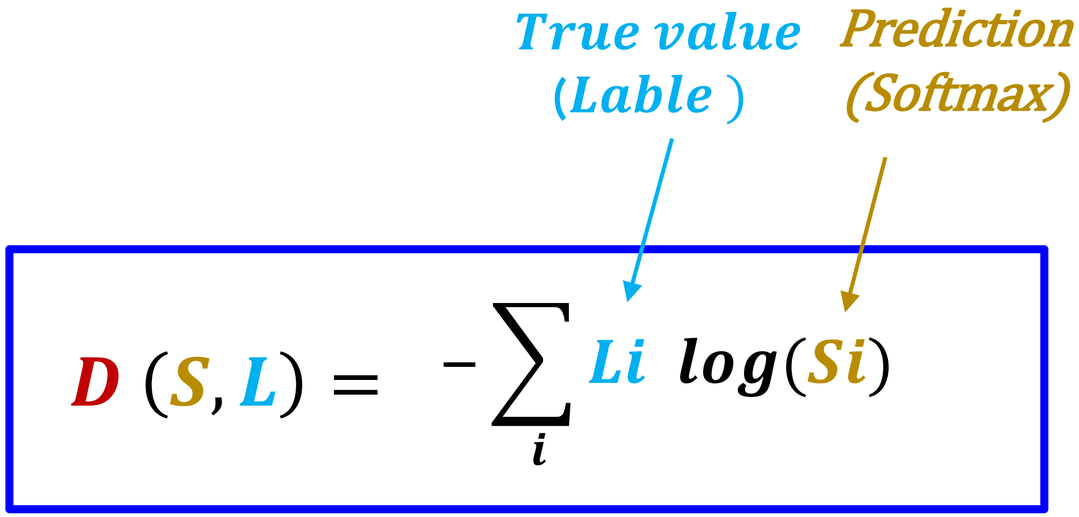
$$\boldsymbol{D}\,(\boldsymbol{S}, \boldsymbol{L}) = -\sum_{\boldsymbol{i}} \boldsymbol{Li}\ \boldsymbol{log}(\boldsymbol{Si})$$

Difference (Prediction of Softmax , True value)

H(x) Label

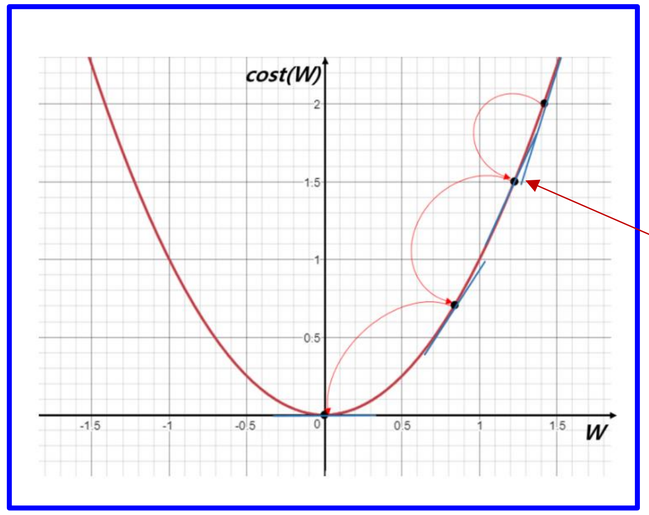
Cost function





Gradient descent

Cost (Loss)



STEP (learning rate)

Loss (Cost)

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

Training set