Lecture 06

Softmax classification: Multinomial classification

Logistic regression

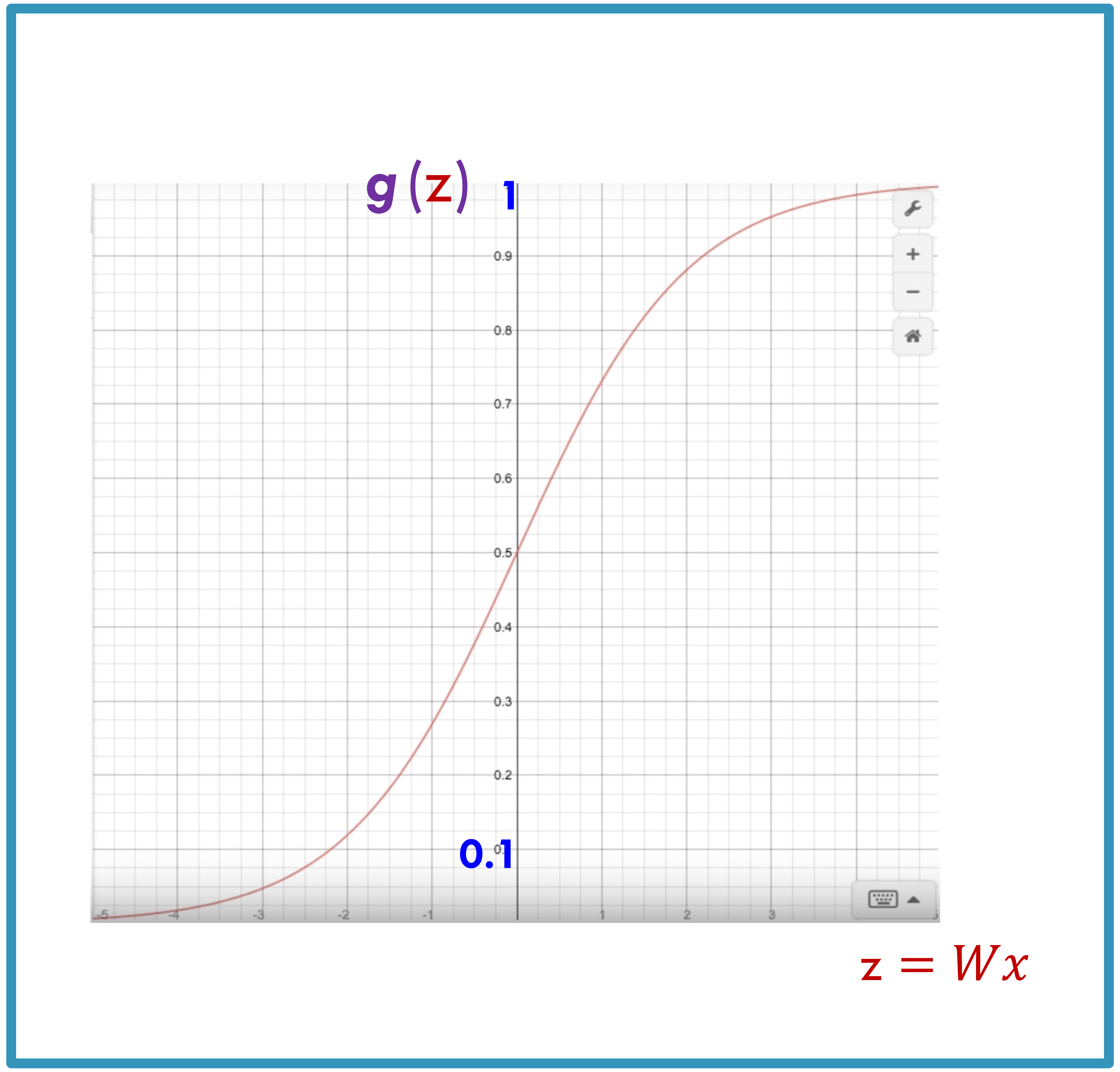
$$H_L(x) = Wx$$

$$z = H_L(x), \quad g(z) \rightarrow 0\sim1$$

$$Z$$

$$\downarrow$$

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



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

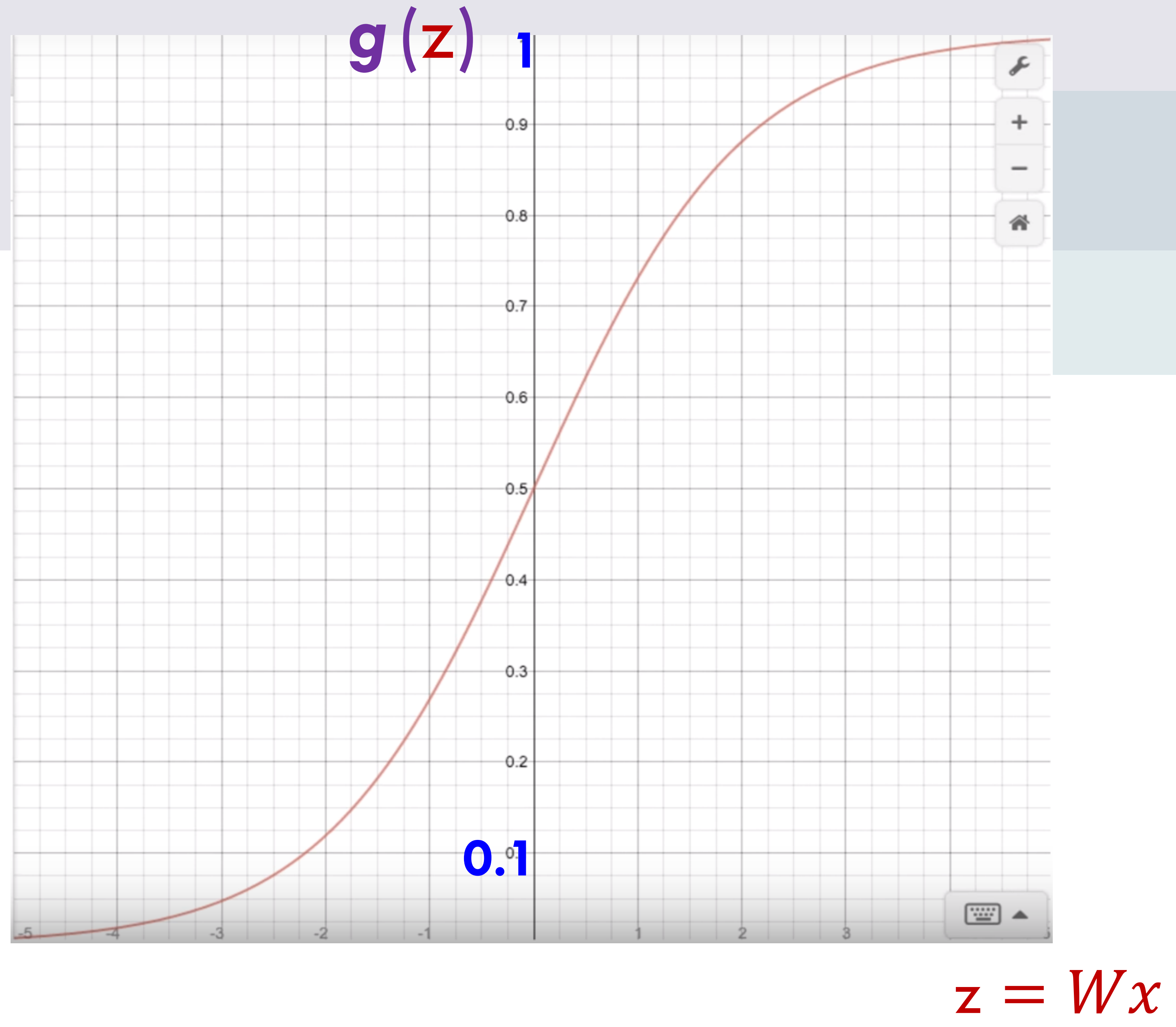
Logistic regression

$$H_L(x) = Wx$$

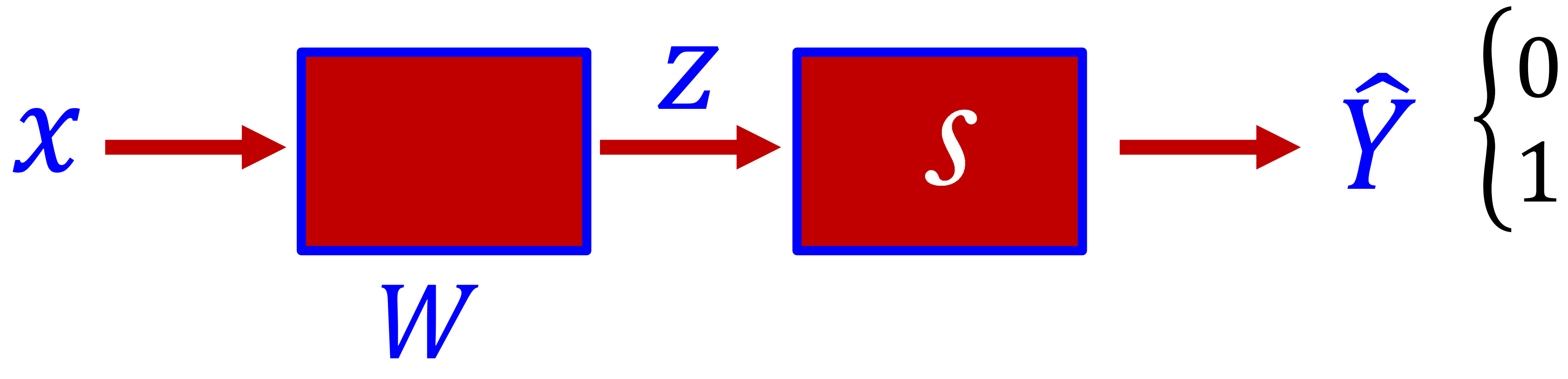
$$z = H_L(x), \quad g(z) \rightarrow 0\sim1$$

$$Z \downarrow g(Z)$$

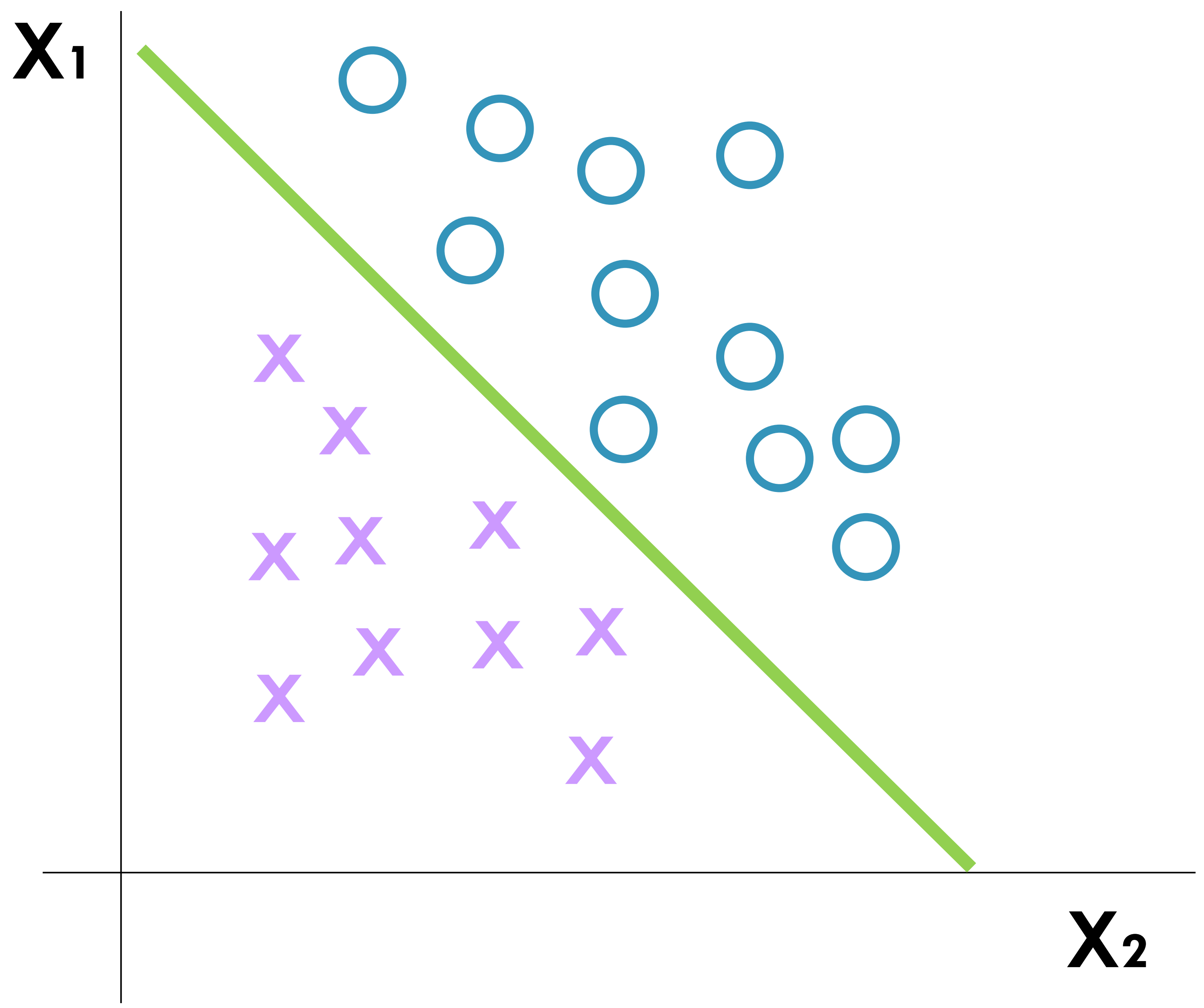
$$g(z) = \frac{1}{(1+e^{-z})}$$



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

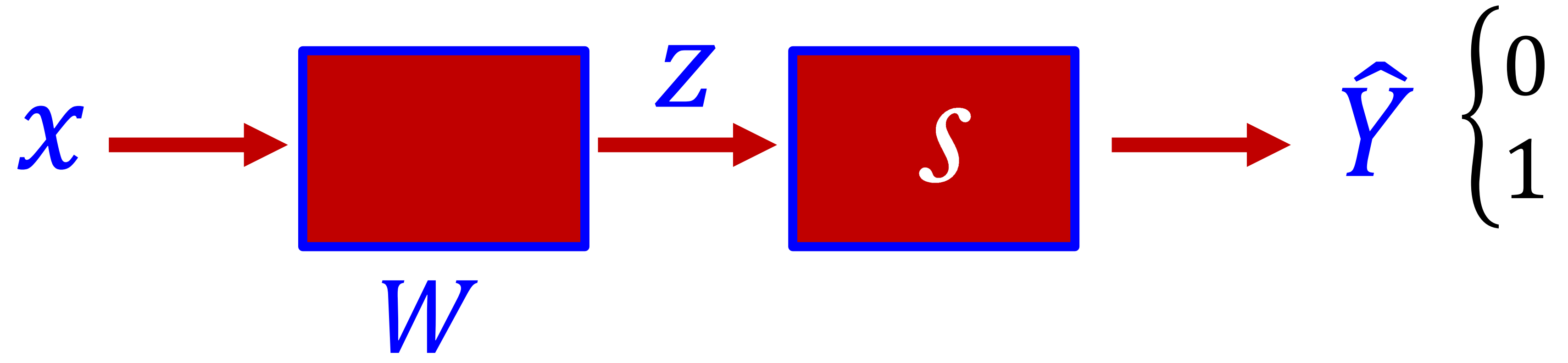


Logistic regression



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

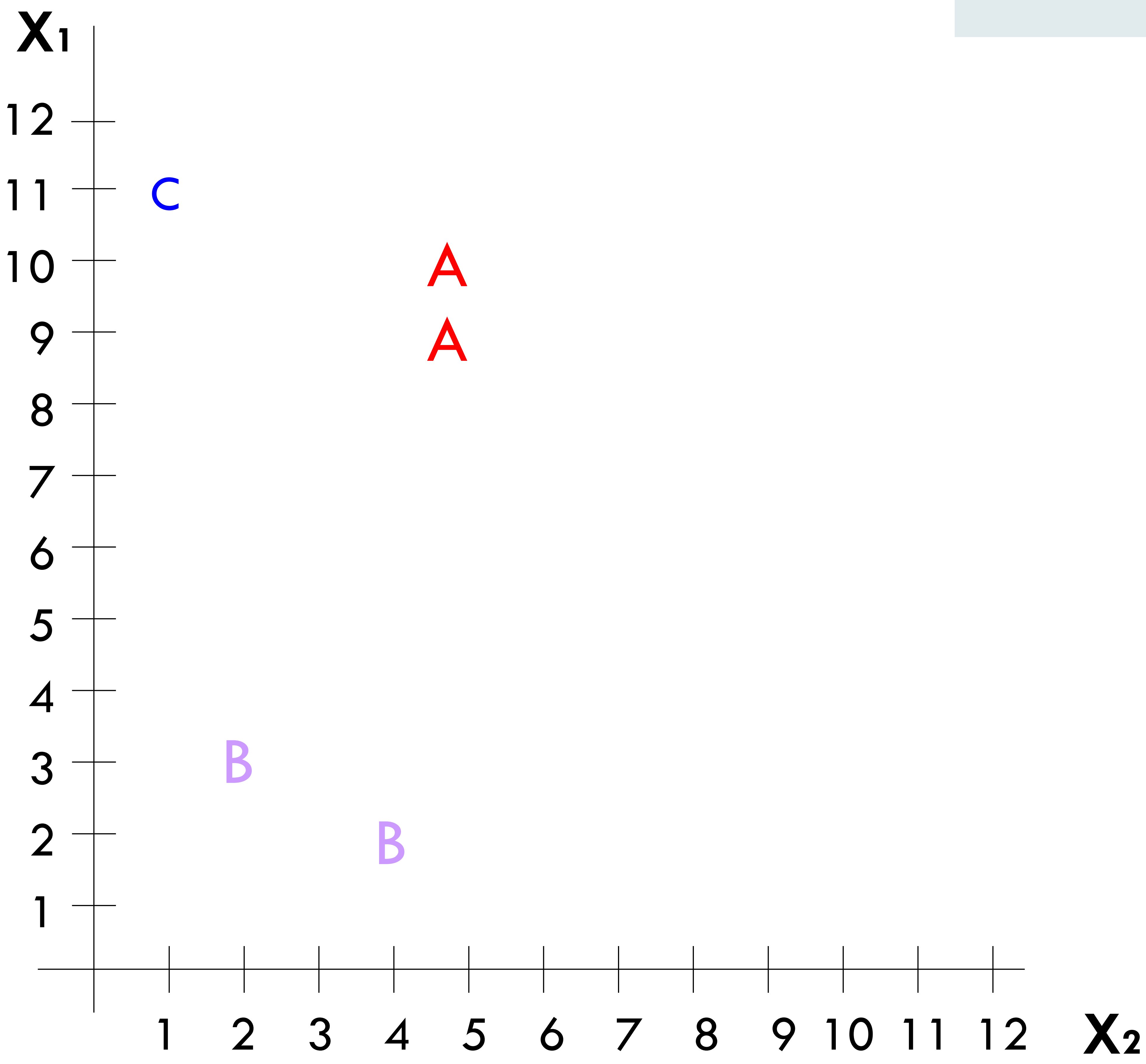
$$H_R(x) = g(H_L(x))$$



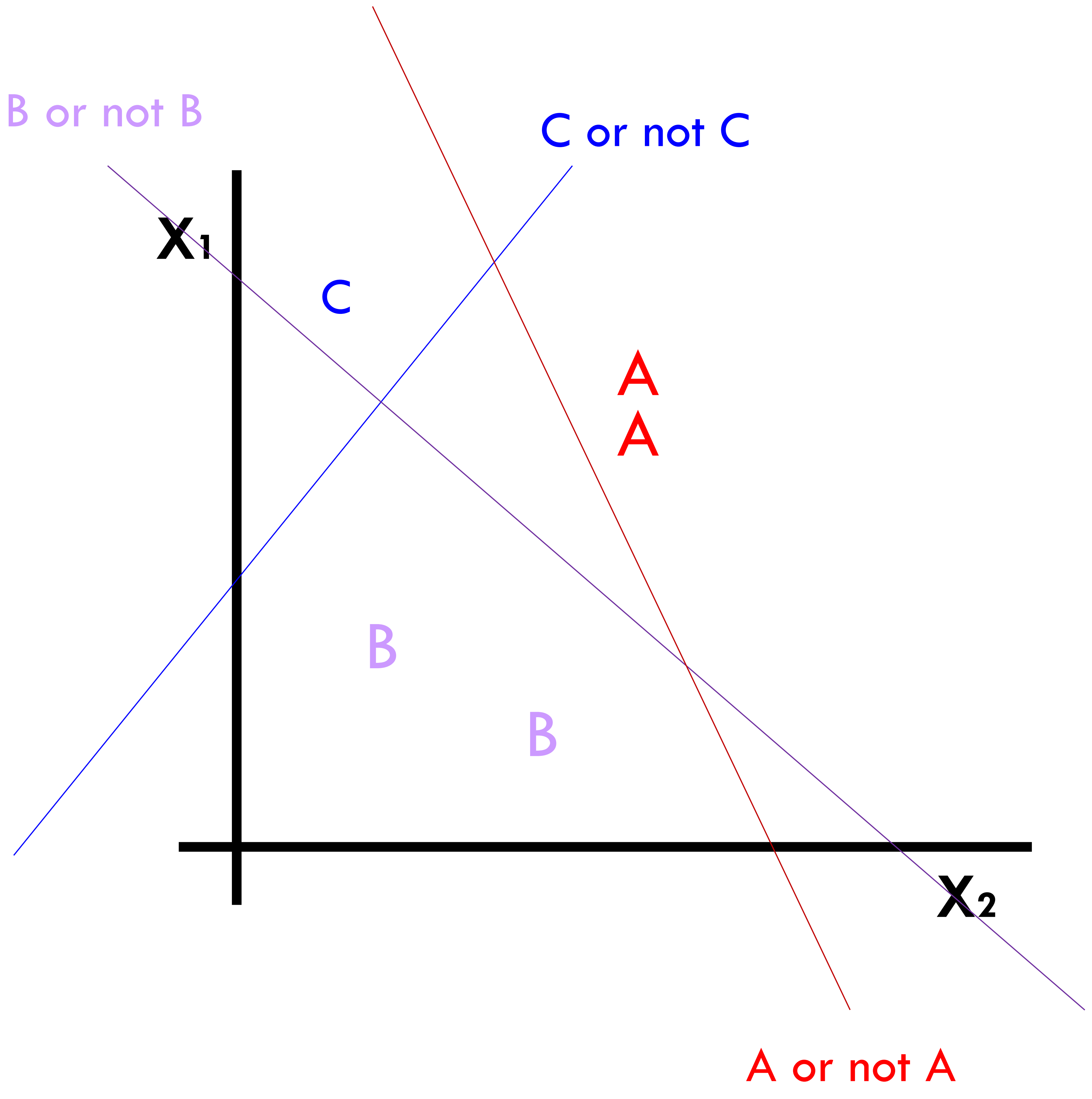
if hypothesis>0.5 :
then 1
else 0

Multinomial classification

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

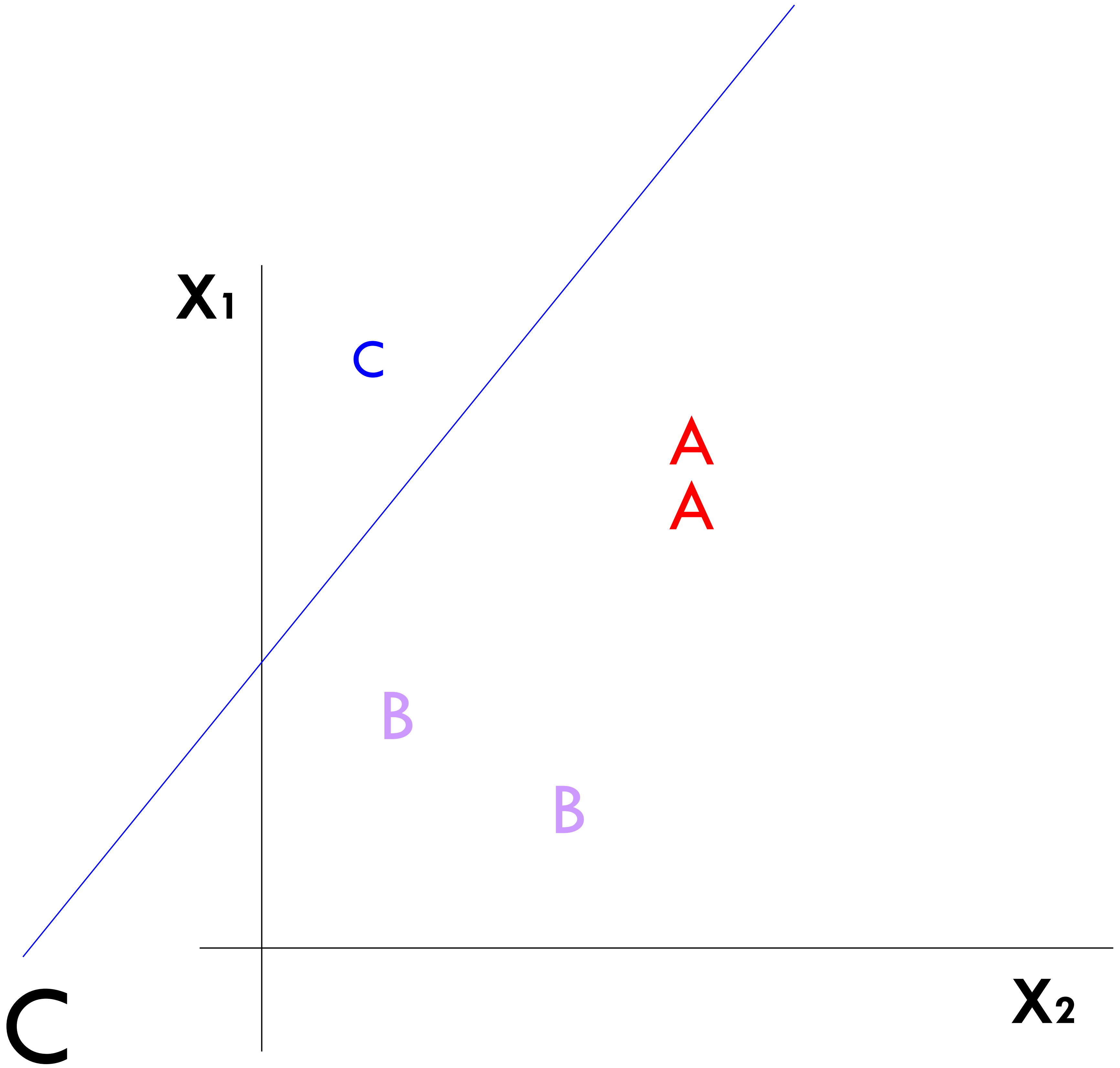


Multinomial classification



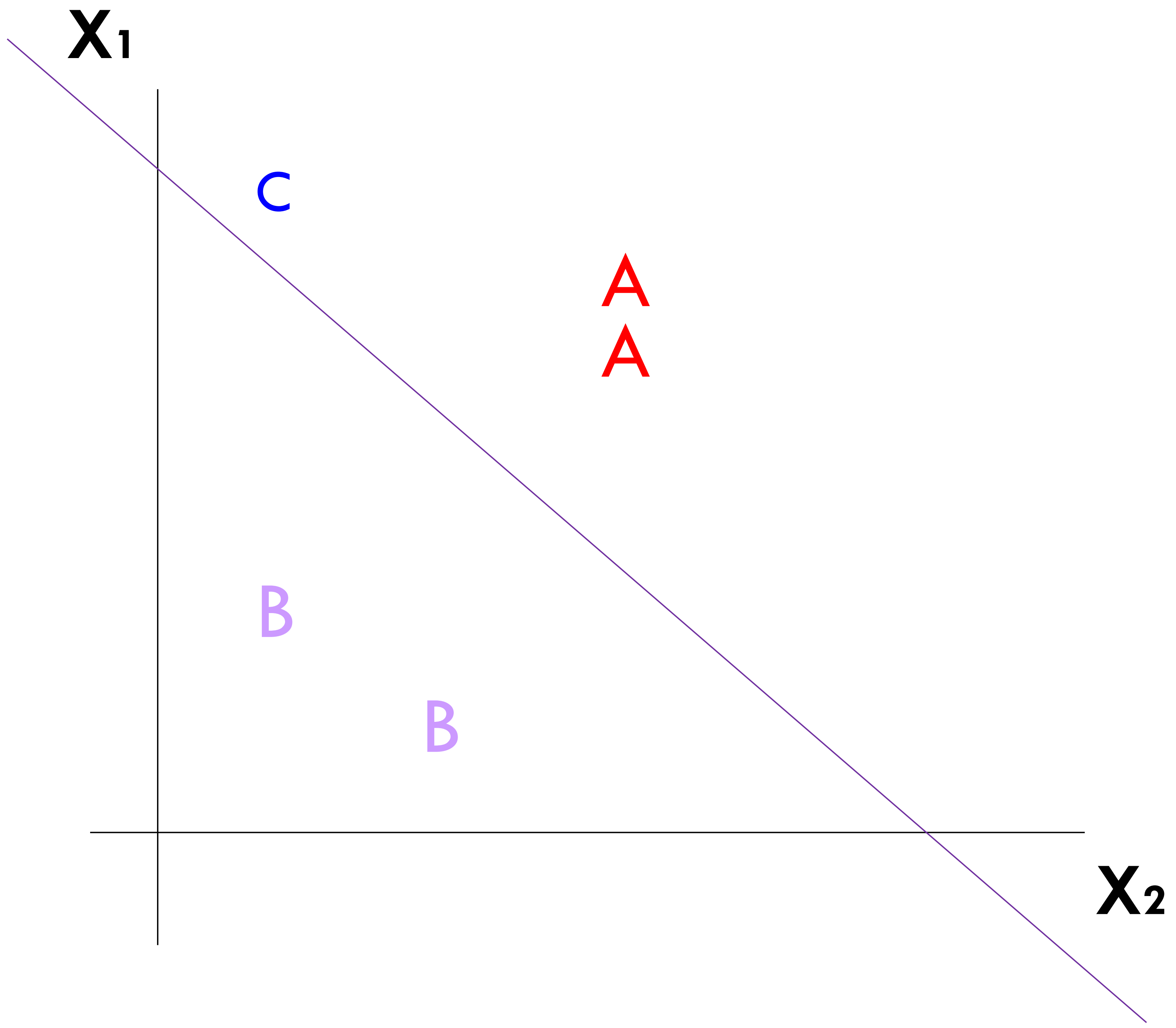
Binary classification

Multinomial classification



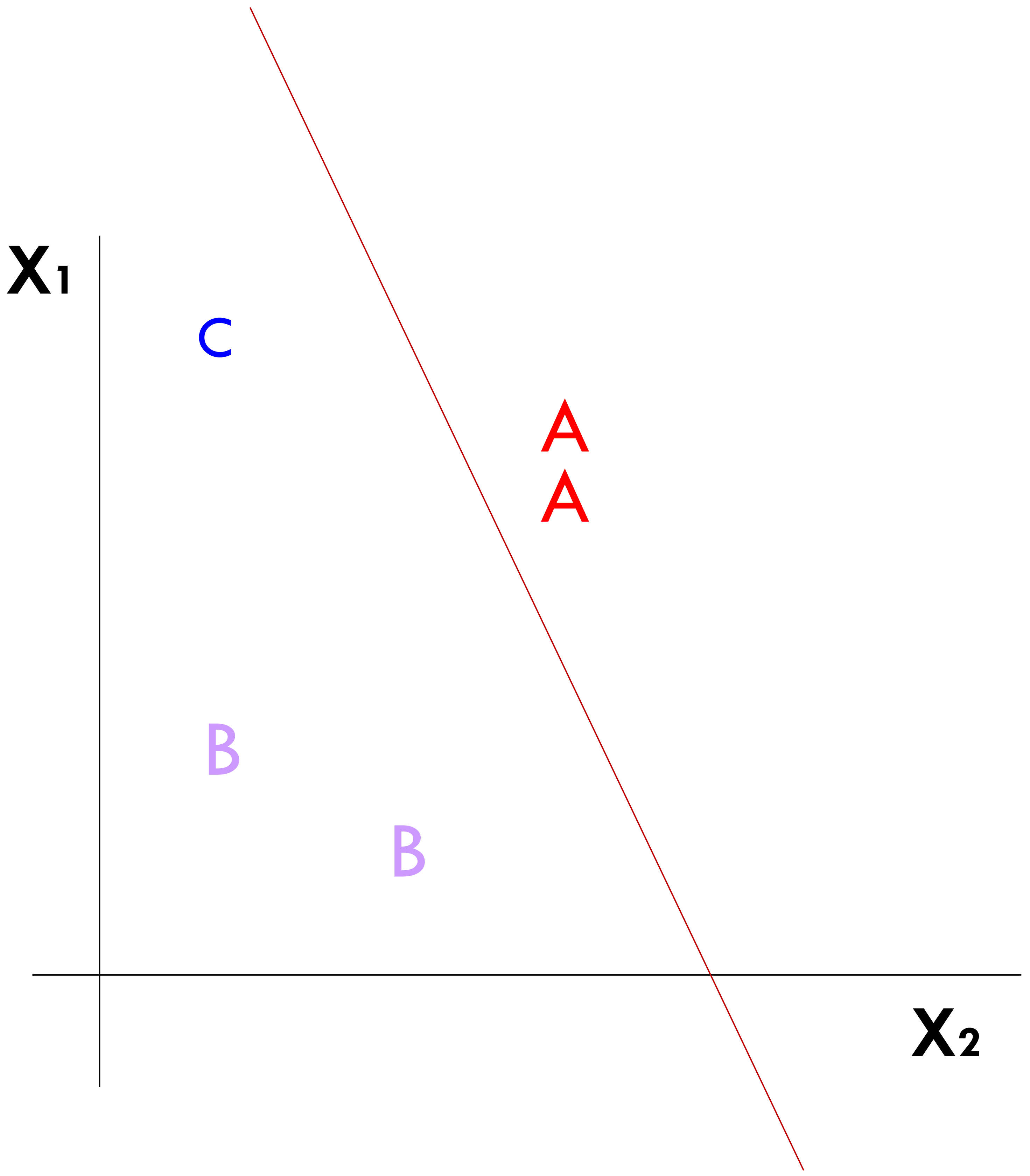
C or not C

Multinomial classification



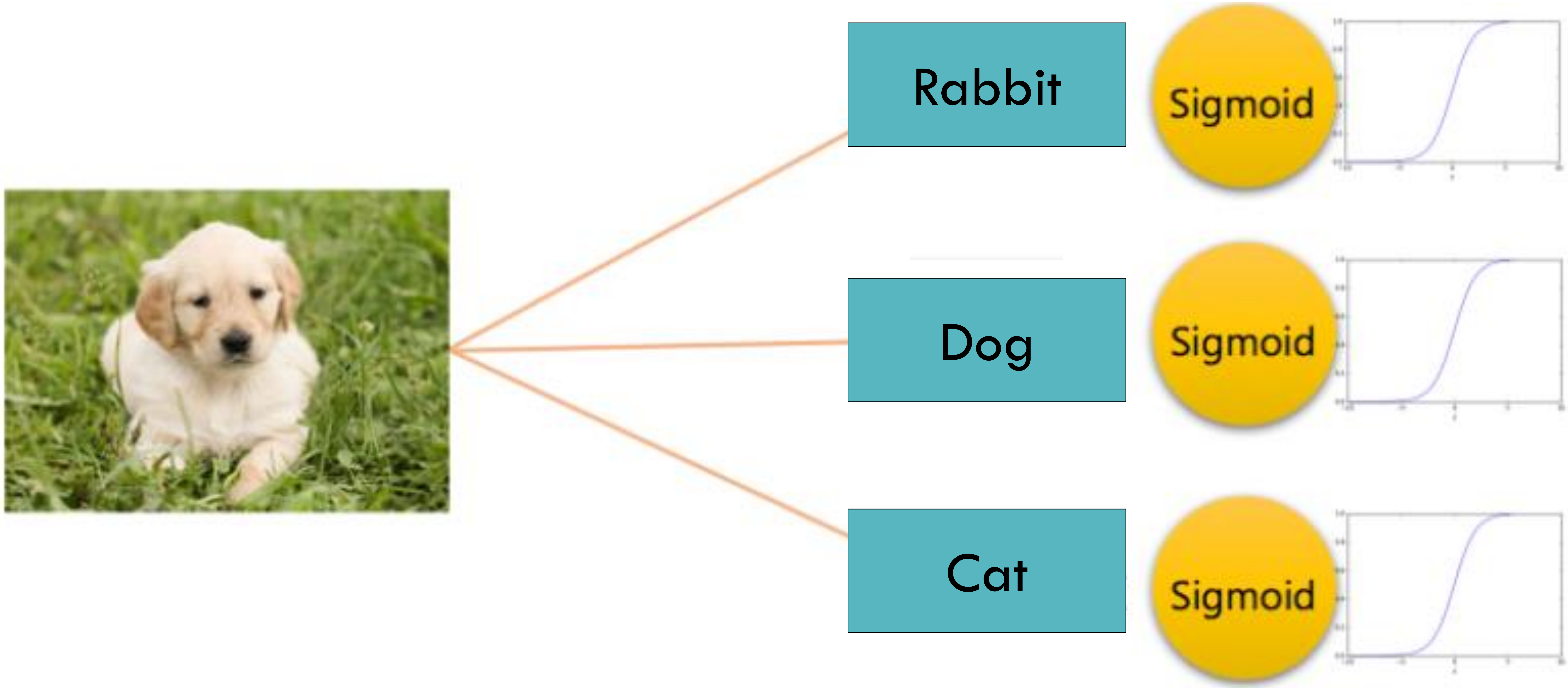
B or not B

Multinomial classification



A or not A

Multinomial classification



Rabbit

Sigmoid

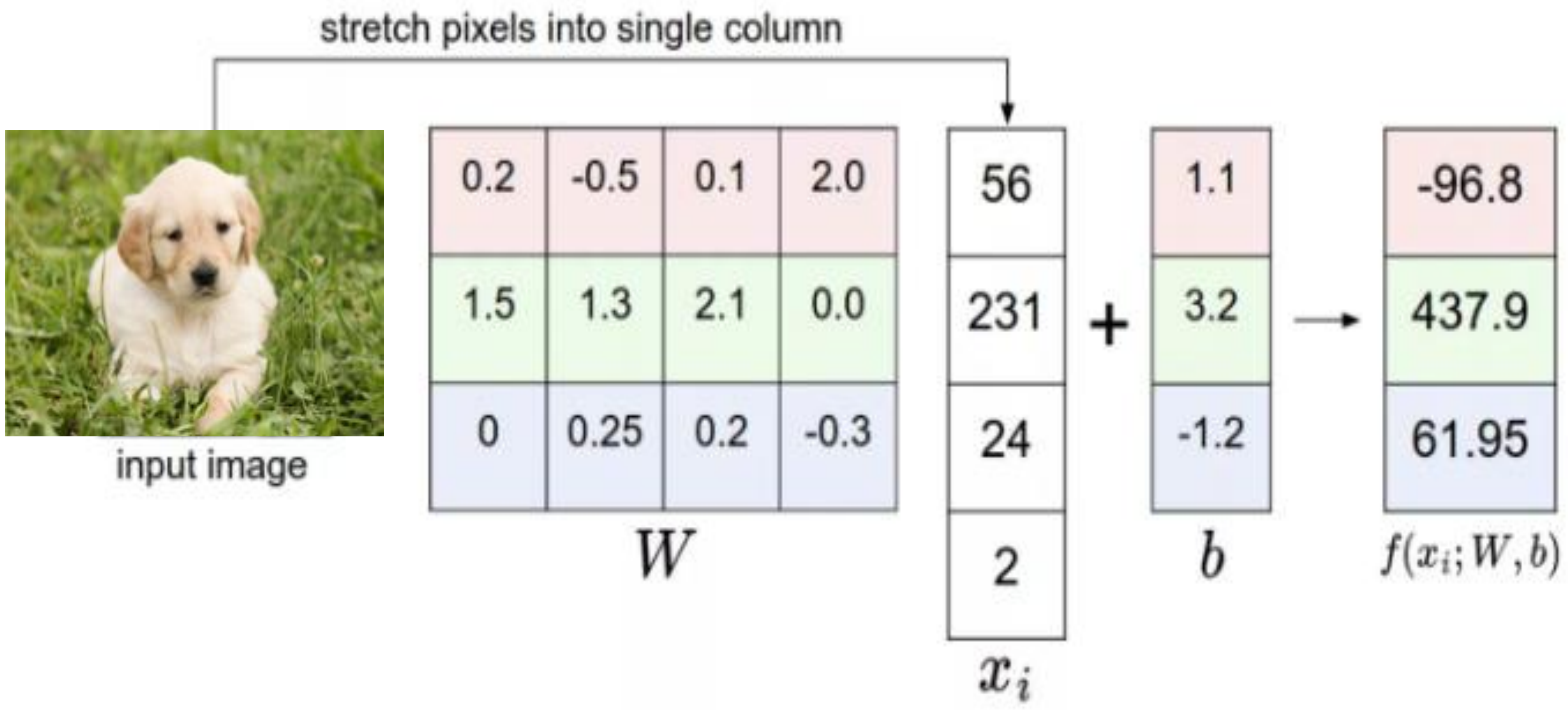
Dog

Sigmoid

Cat

Sigmoid

Multinomial classification

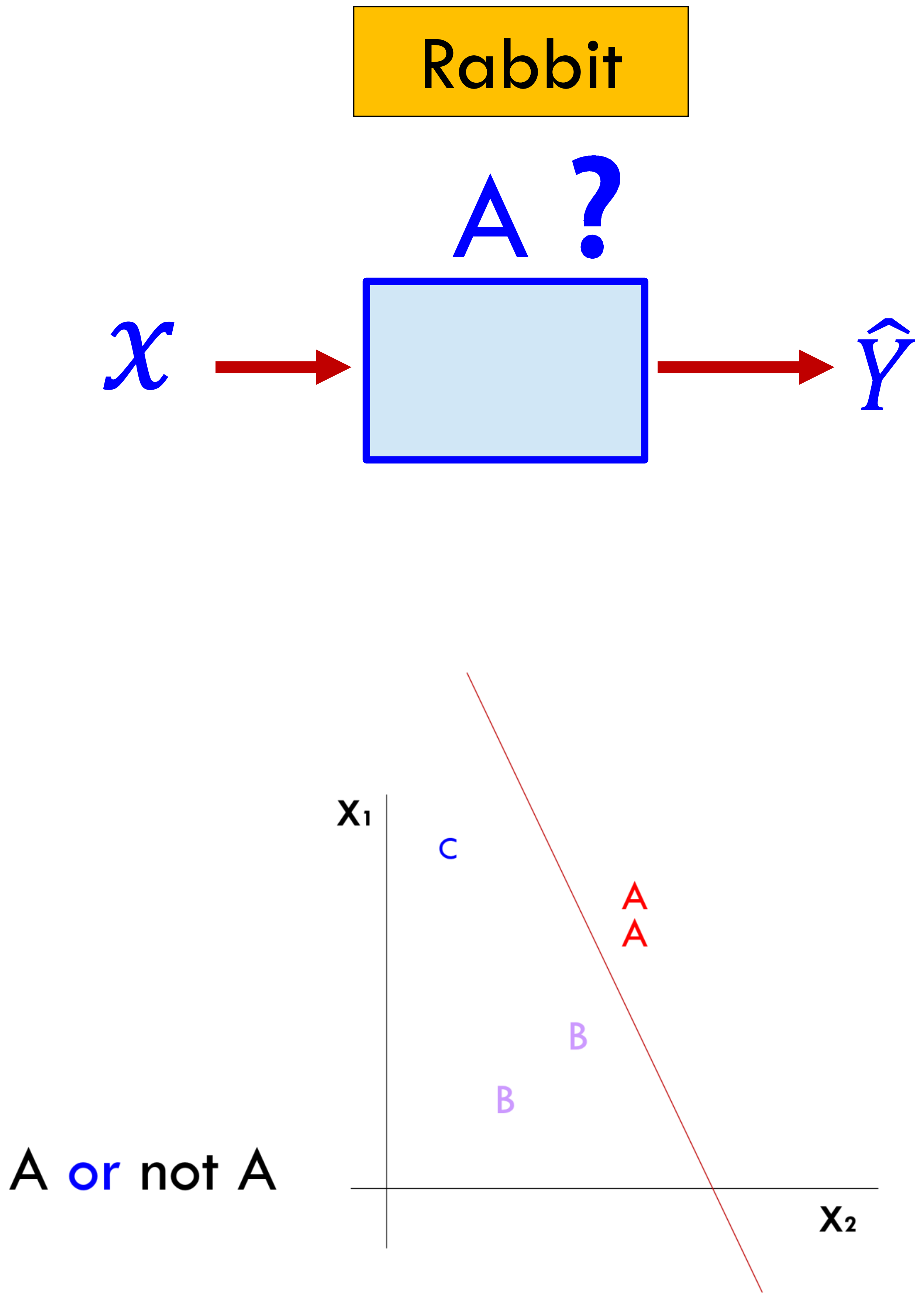


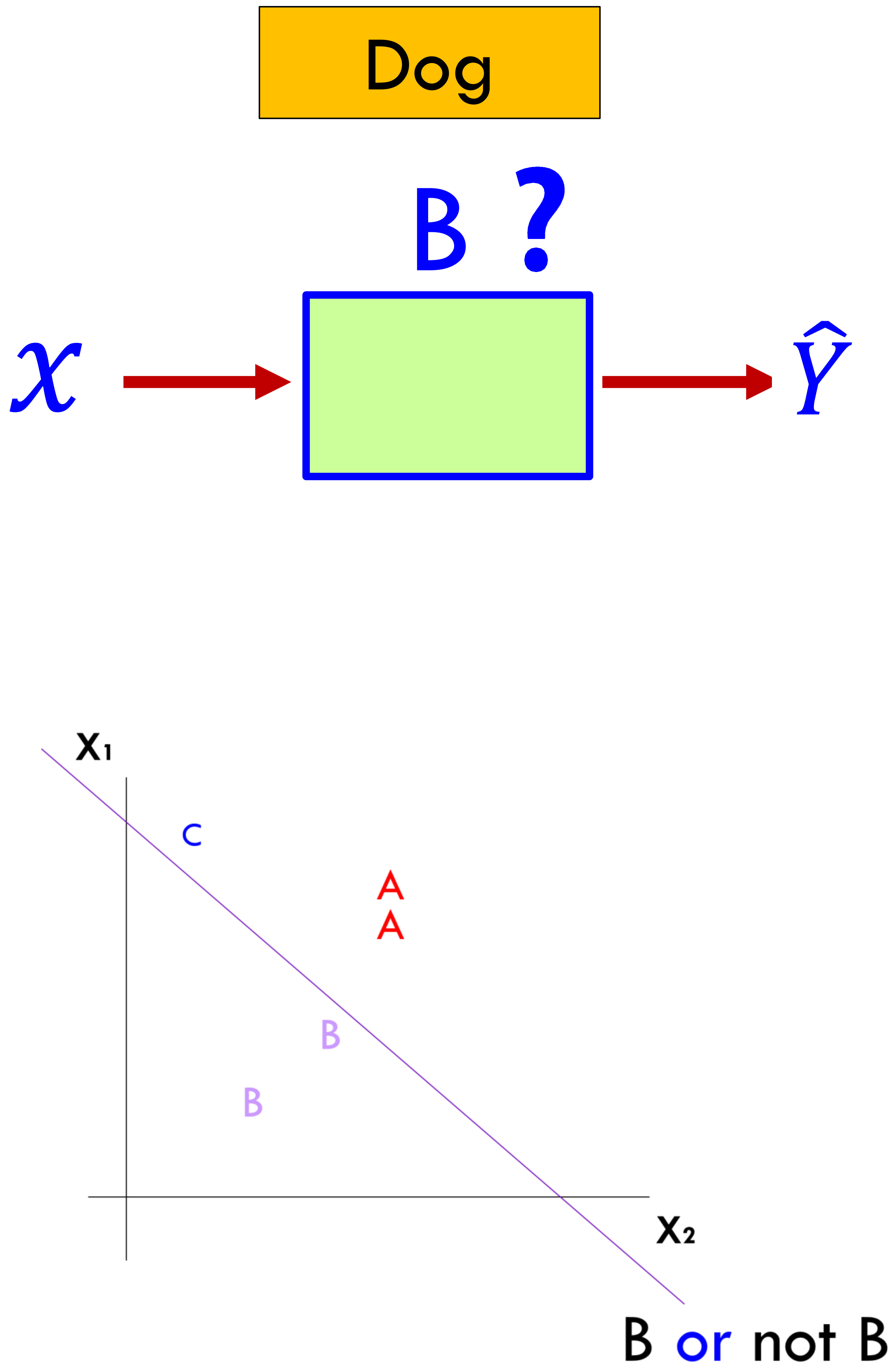
Rabbit

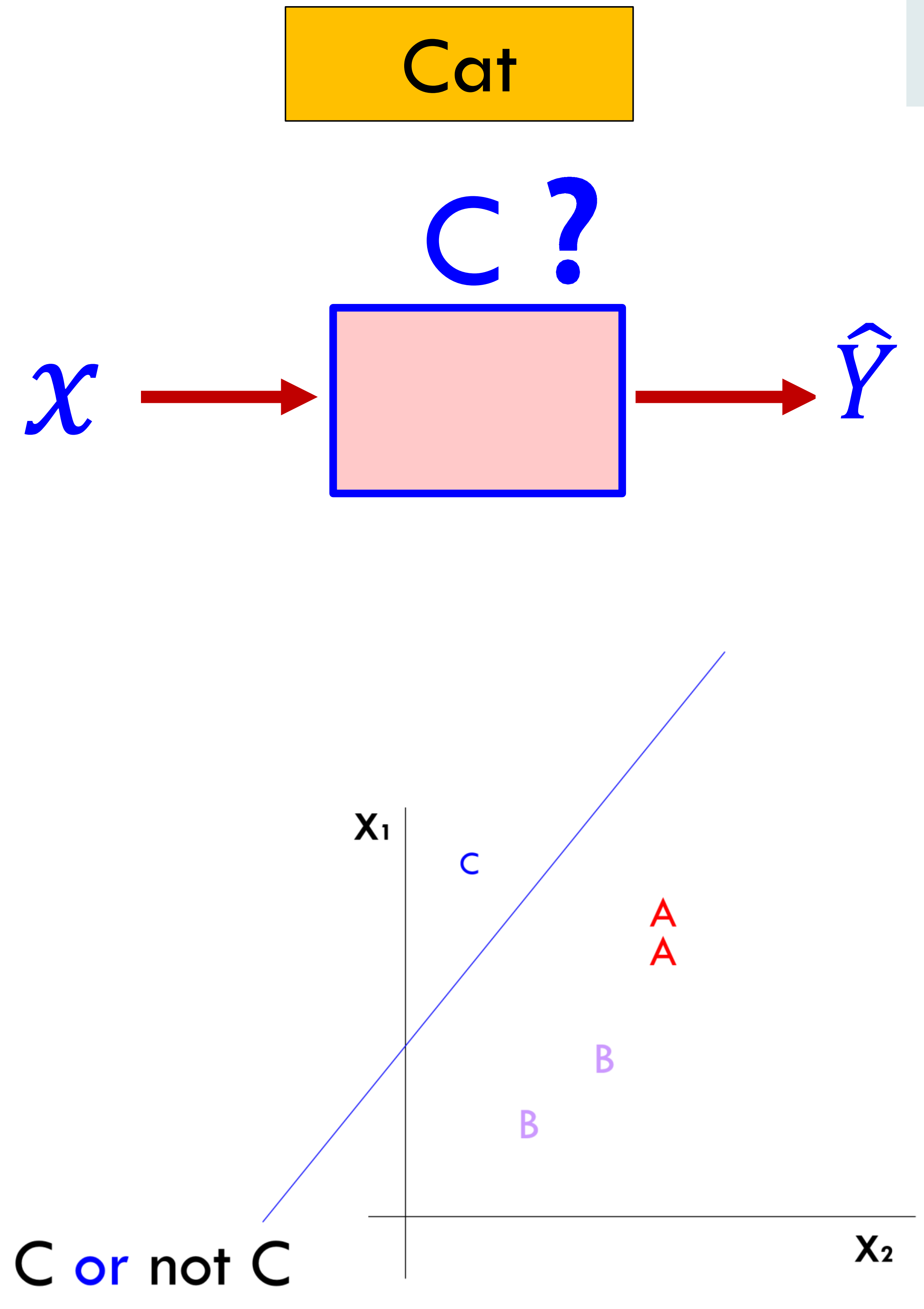
Dog

Cat

Multinomial classification



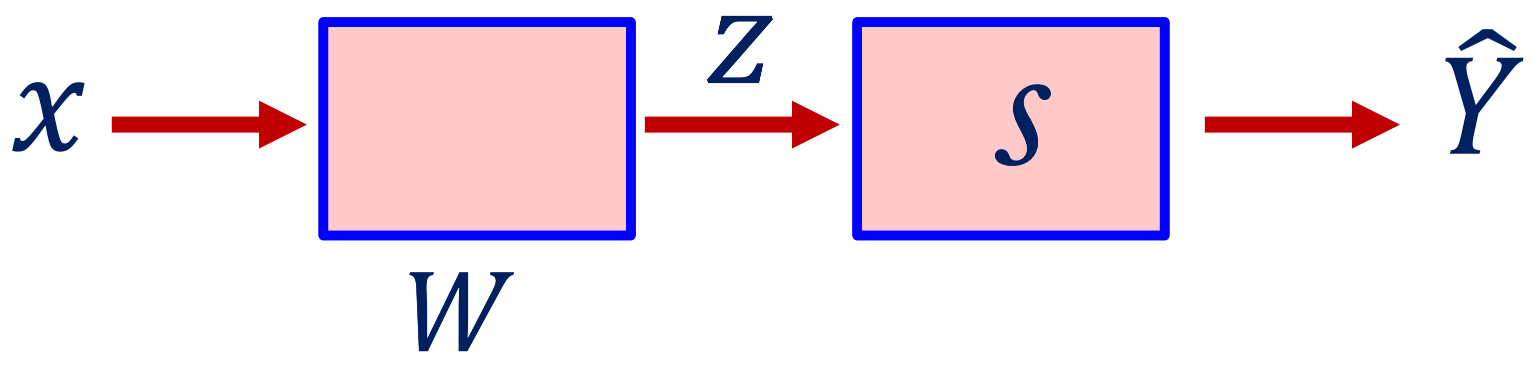




Multinomial classification

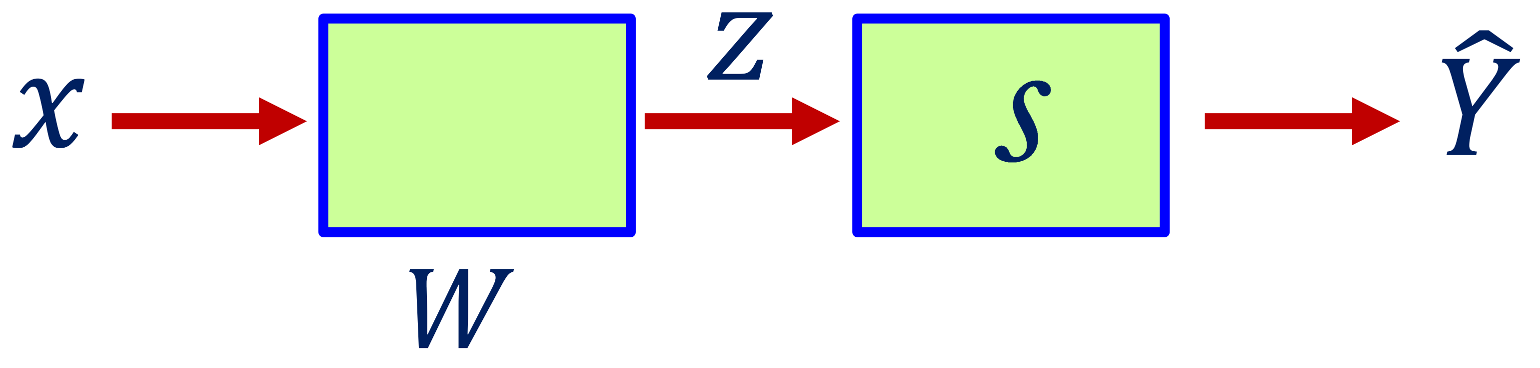
Rabbit

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



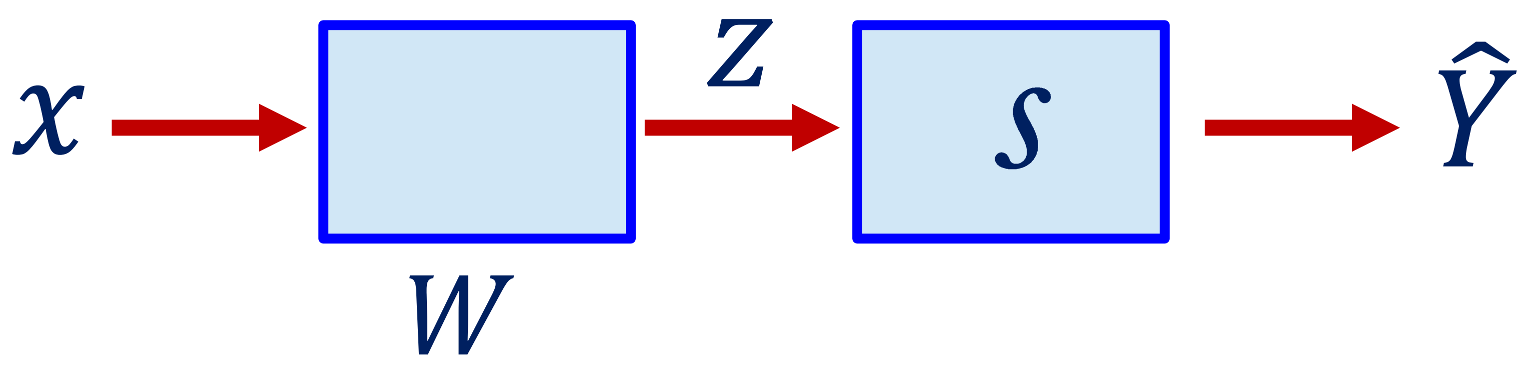
Dog

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



Cat

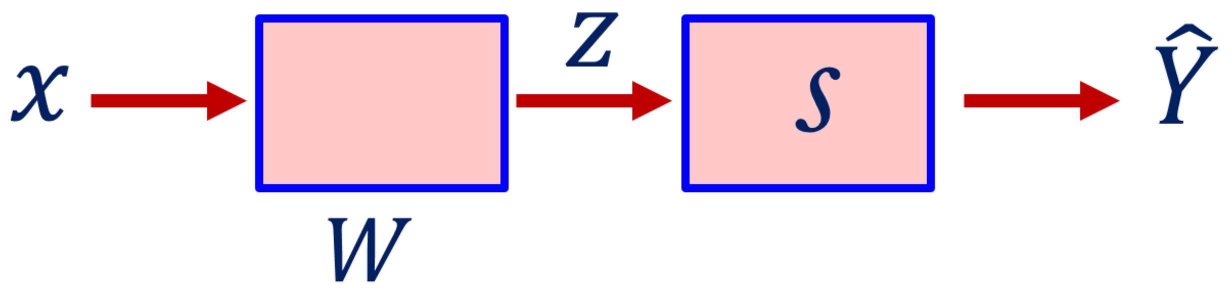
$$[WC1 \;\; WC2 \;\; WC3]\begin{bmatrix}x1\\x2\\x3\end{bmatrix} = [WC1x1 + WC2x2 + WC3x3]$$



Multinomial classification

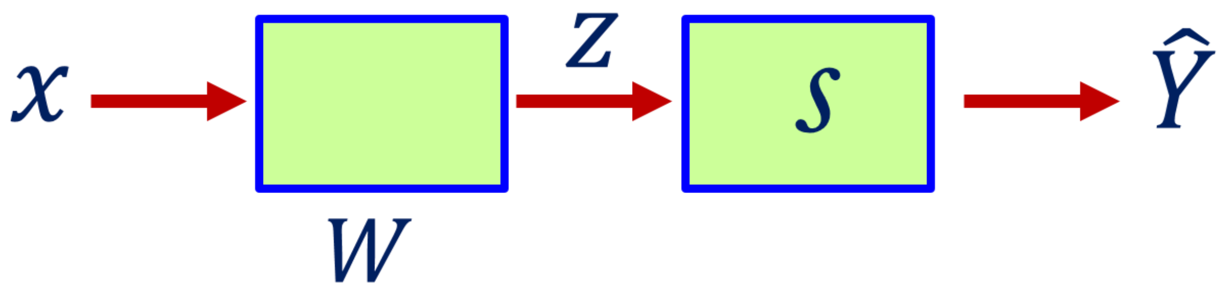
Rabbit

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



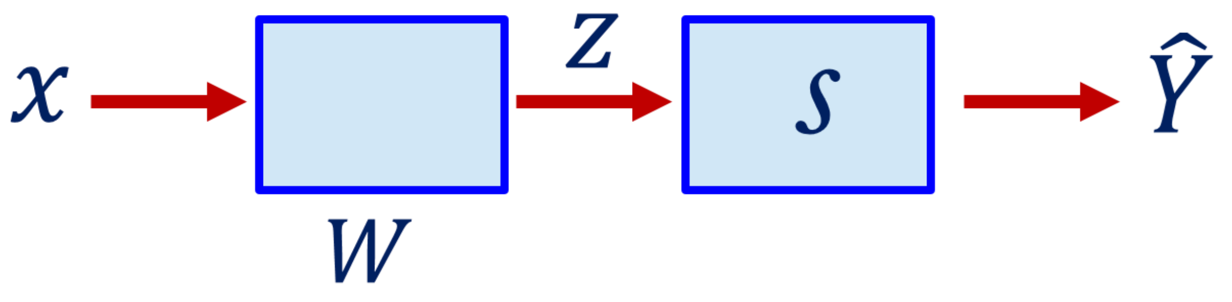
Dog

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



Cat

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

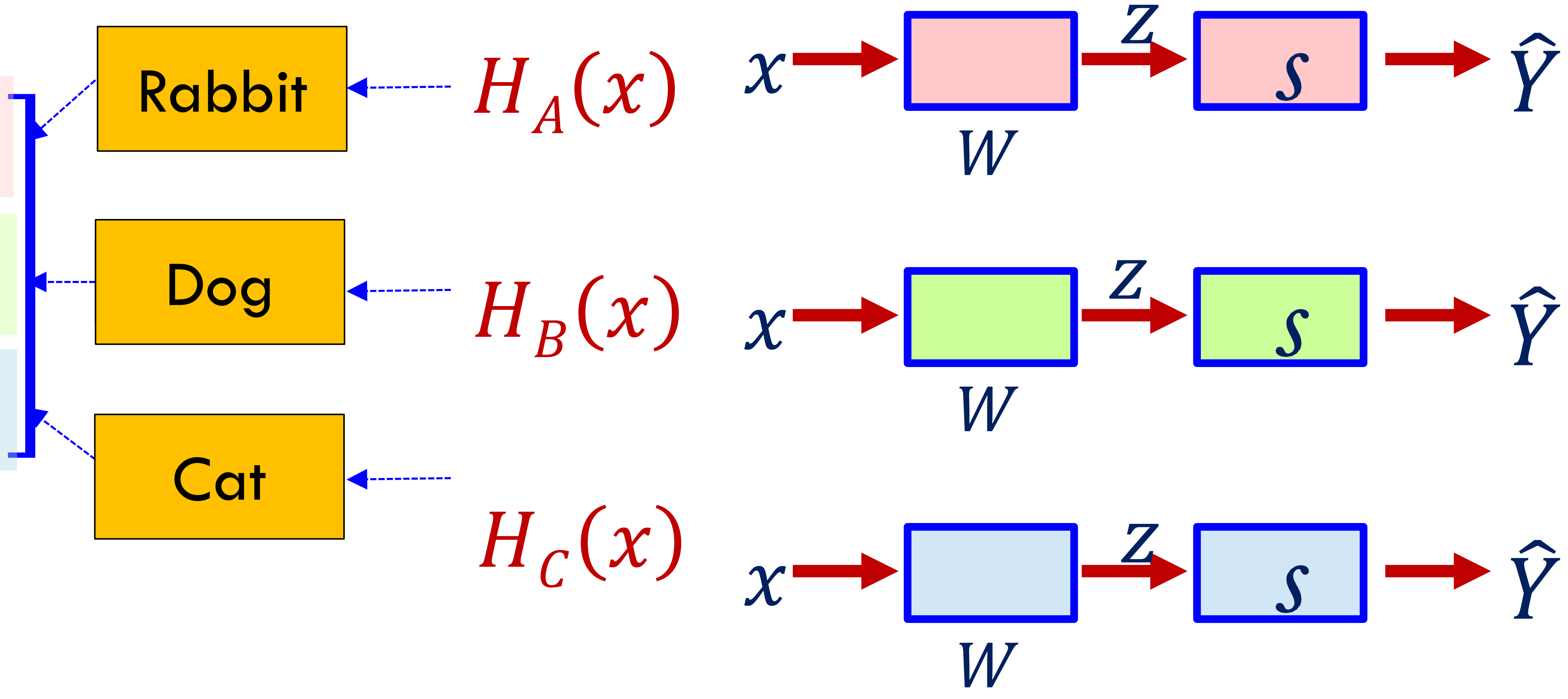


$$\begin{array}{l} \text{Rabbit} \\ \text{Dog} \\ \text{Cat} \end{array} \begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix}x1\\x2\\x3\end{bmatrix} =$$

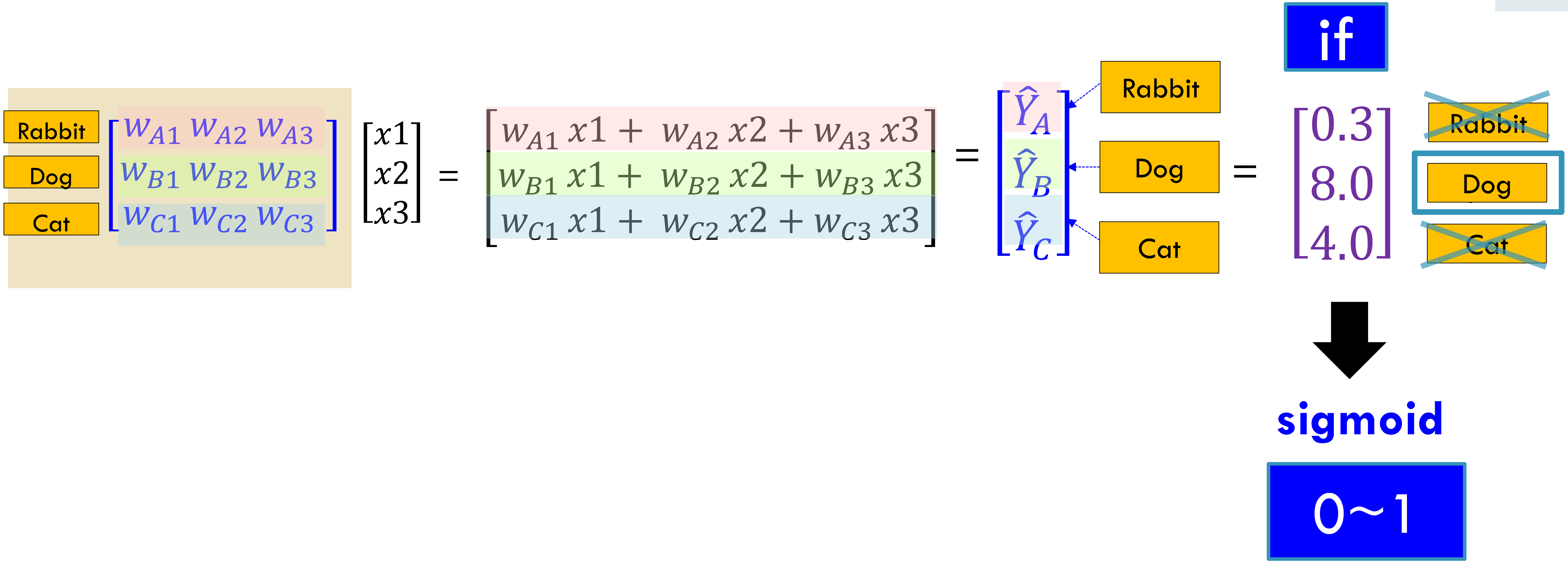
Multinomial classification

$$\begin{array}{l|c} \text{Rabbit} & \\ \text{Dog} & \begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x1 \\ x2 \\ x3 \end{bmatrix} = \\ \text{Cat} & \end{array}$$

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



Where is sigmoid?



$$\begin{array}{c} \text{Rabbit} \\ \text{Dog} \\ \text{Cat} \end{array} \begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x1 \\ x2 \\ x3 \end{bmatrix} = \begin{bmatrix} w_{A1}\,x1 + w_{A2}\,x2 + w_{A3}\,x3 \\ w_{B1}\,x1 + w_{B2}\,x2 + w_{B3}\,x3 \\ w_{C1}\,x1 + w_{C2}\,x2 + w_{C3}\,x3 \end{bmatrix} = \begin{bmatrix} \hat{Y}_A \\ \hat{Y}_B \\ \hat{Y}_C \end{bmatrix} \begin{array}{l} \leftarrow \text{Rabbit} \\ \leftarrow \text{Dog} \\ \leftarrow \text{Cat} \end{array} = \begin{bmatrix} 0.3 \\ 8.0 \\ 4.0 \end{bmatrix} \begin{array}{c} \cancel{\text{Rabbit}} \\ \boxed{\text{Dog}} \\ \cancel{\text{Cat}} \end{array}$$

if

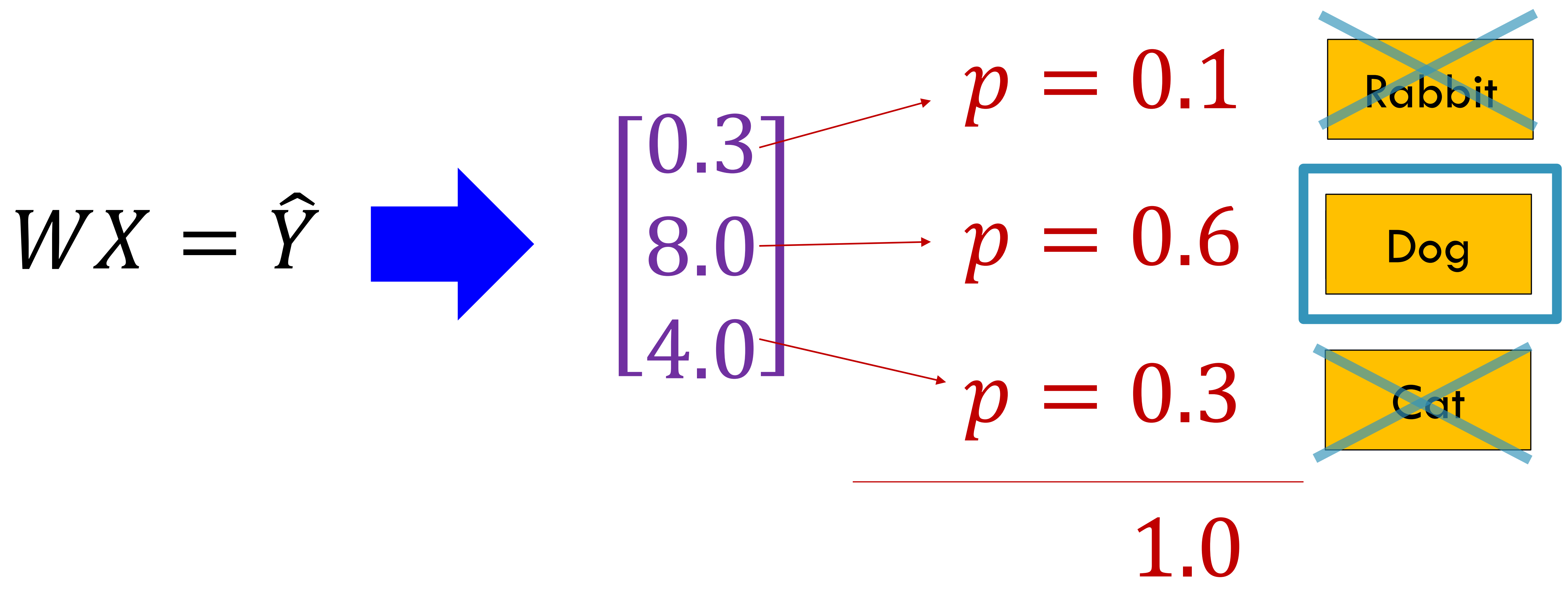
$\downarrow$

sigmoid

0~1

Sigmoid?

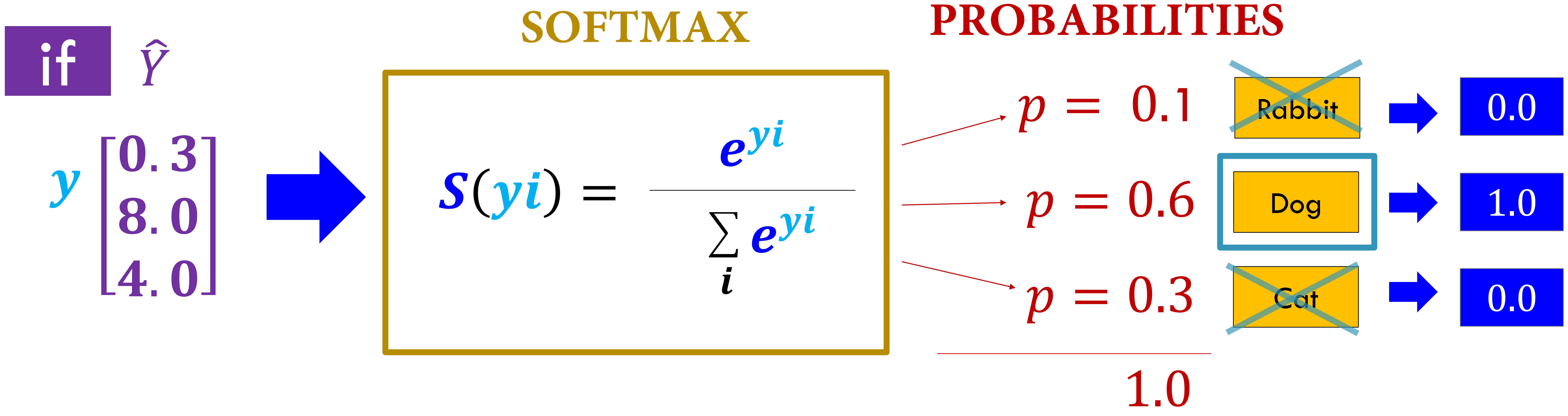
Logistic Classifier ?



$$WX = \hat{Y} \Rightarrow \begin{bmatrix} 0.3 \\ 8.0 \\ 4.0 \end{bmatrix} \begin{array}{l} \rightarrow p = 0.1 \\ \rightarrow p = 0.6 \\ \rightarrow p = 0.3 \\ \hline \quad\quad\quad 1.0 \end{array}$$

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

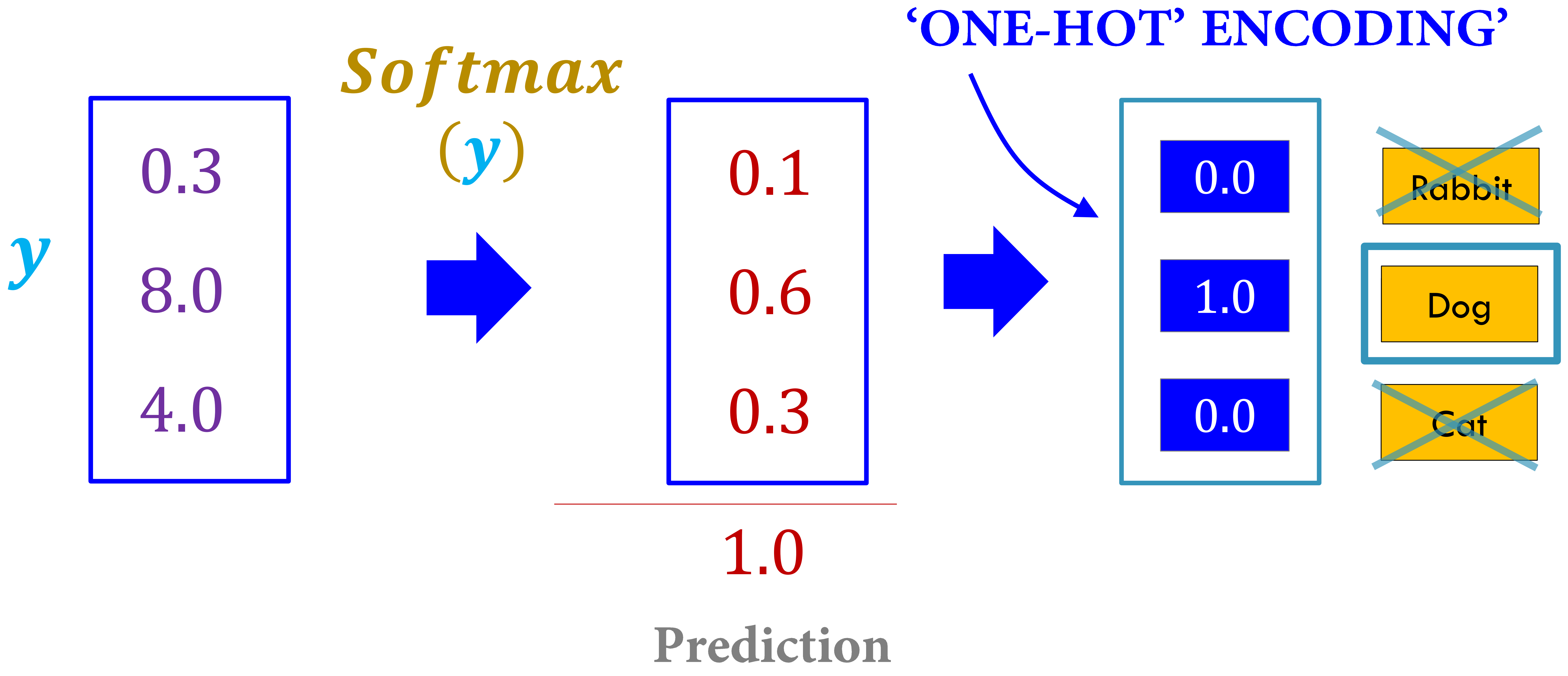
$p = 0.1$ ~~Rabbit~~ 0.0

$p = 0.6$ Dog 1.0

$p = 0.3$ ~~Cat~~ 0.0

1.0

Cost function



y

0.3

8.0

4.0

$Softmax(y)$

0.1

0.6

0.3

1.0

Prediction

'ONE-HOT' ENCODING'

0.0

1.0

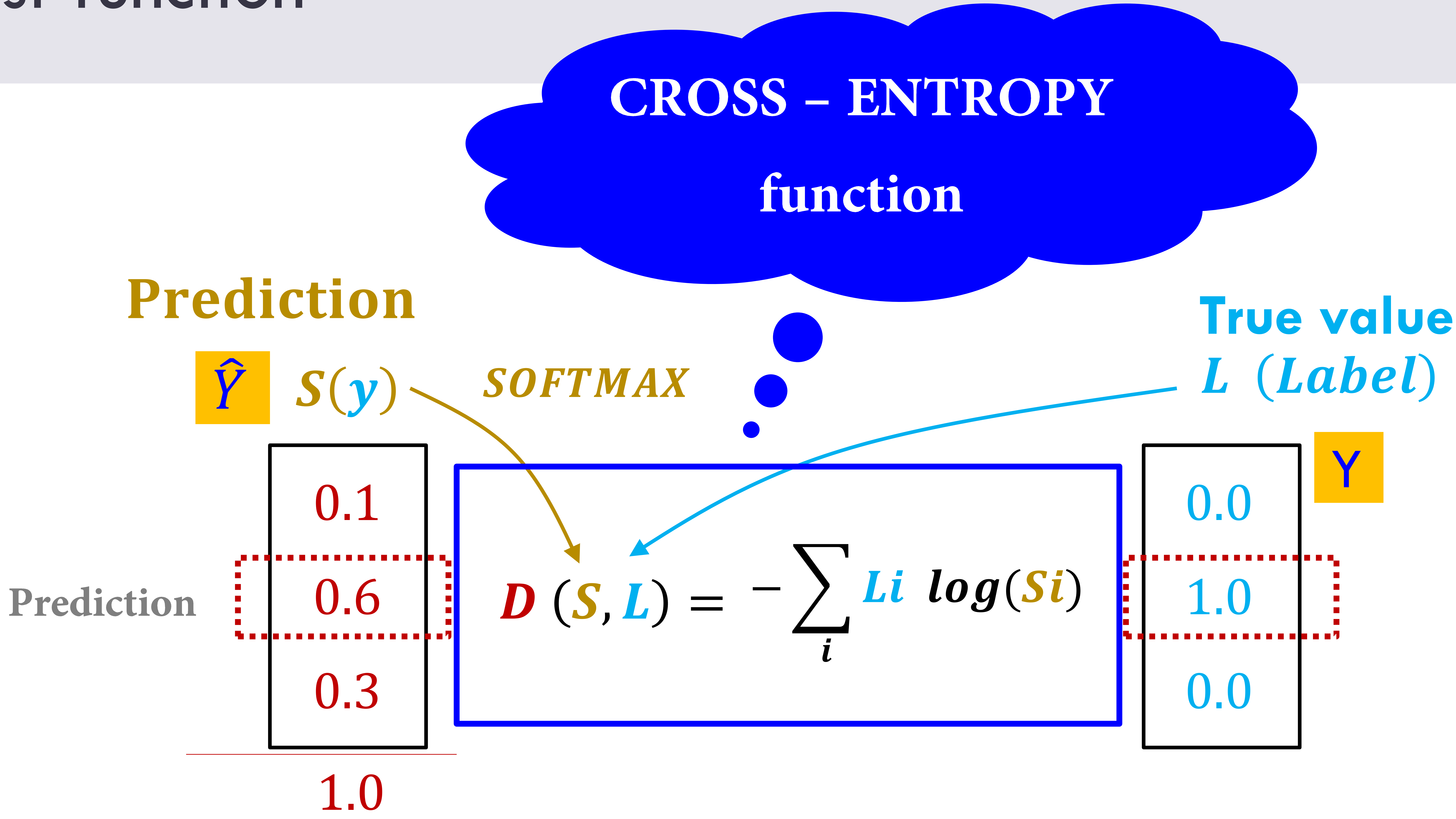
0.0

~~Rabbit~~

Dog

~~Cat~~

Cost function



CROSS – ENTROPY function

Prediction

$\hat{Y}$ $S(y)$

SOFTMAX

True value

L ($Label$)

Y

0.1
0.6
0.3

1.0

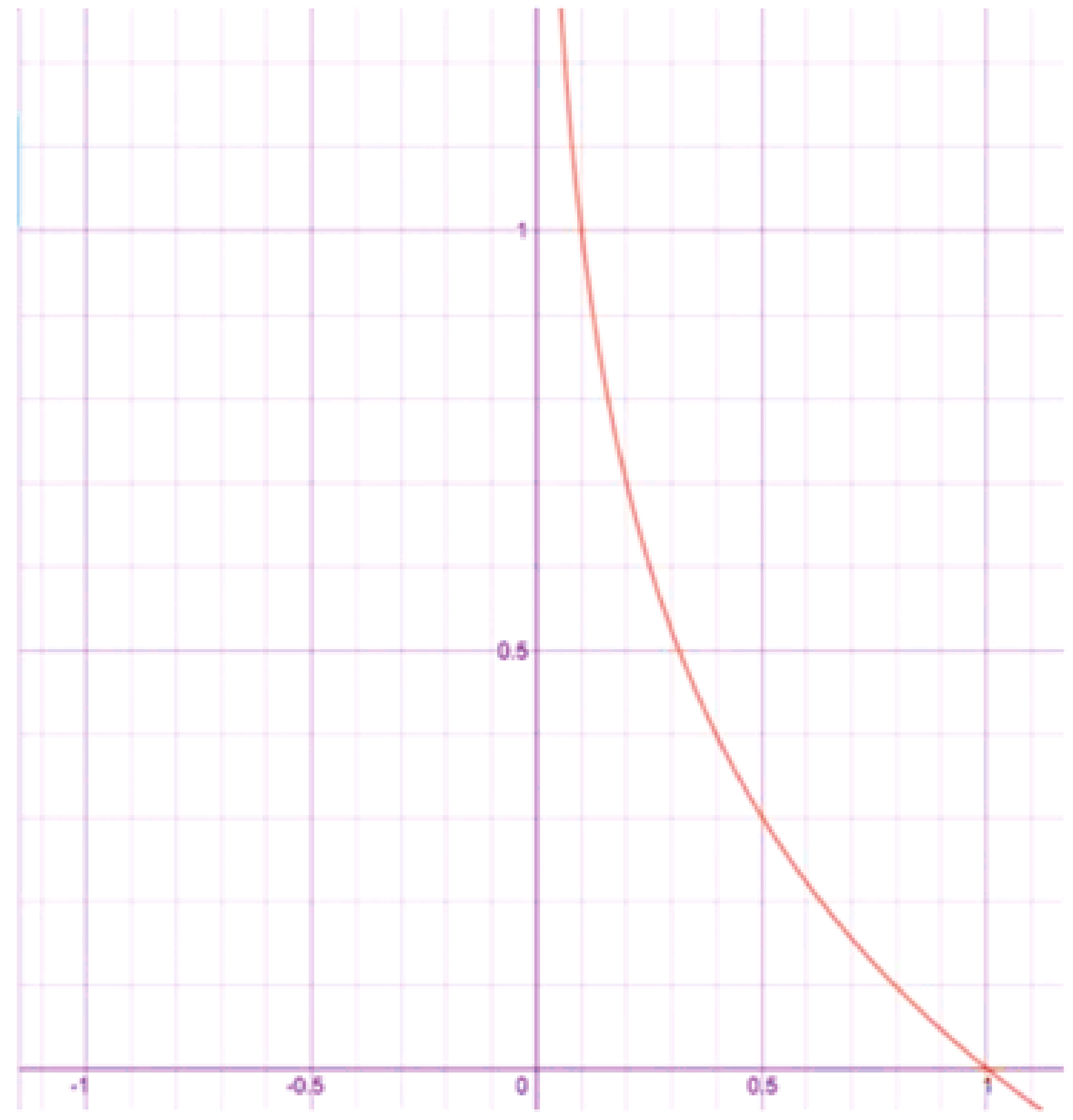
Prediction

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

0.0
1.0
0.0

Cross-entropy cost function

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



A

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

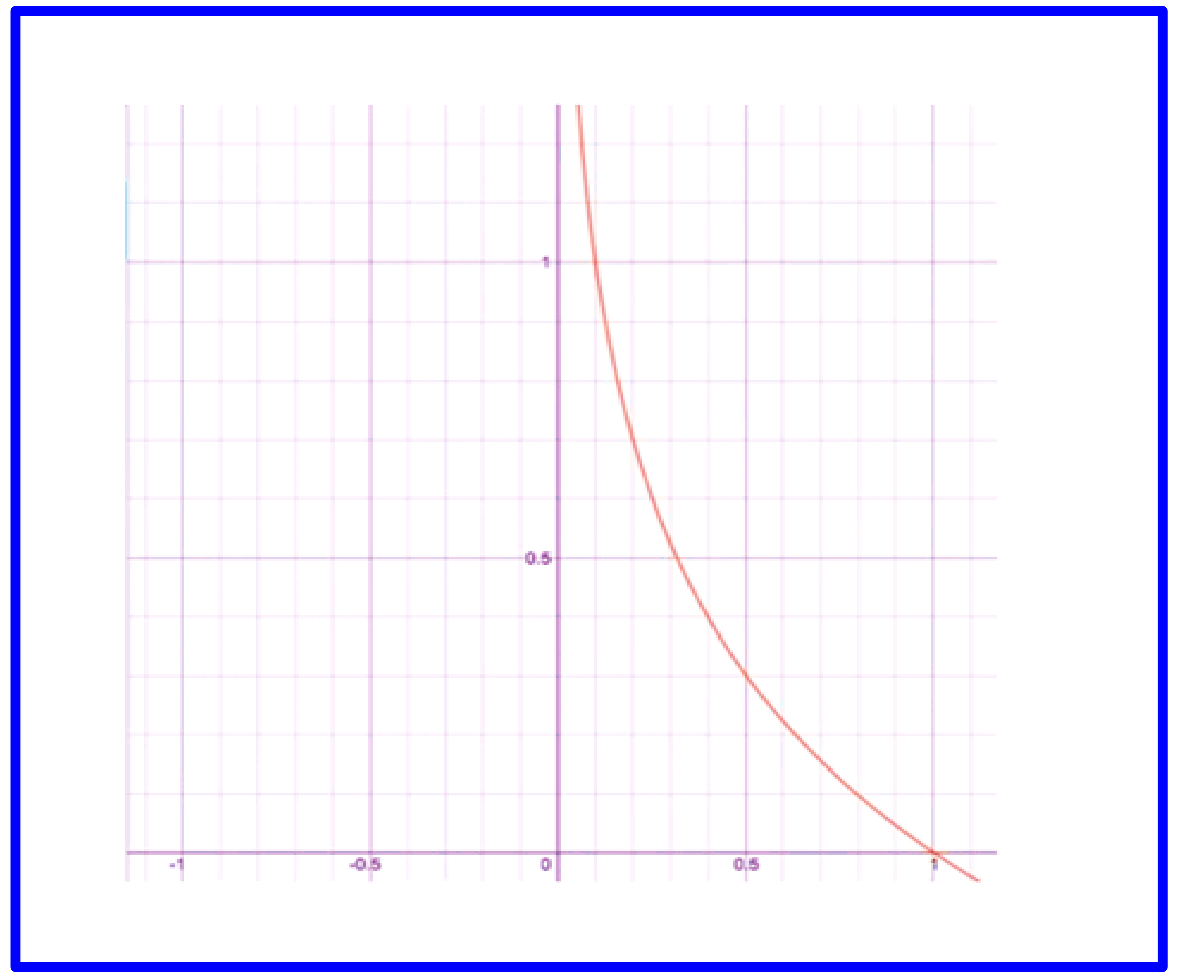
B

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

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

Cross-entropy cost function

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



$$L = \begin{bmatrix} \overset{\boxed{A}}{0} \\ \underset{\boxed{B}}{1} \end{bmatrix} = \boxed{B}$$

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

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

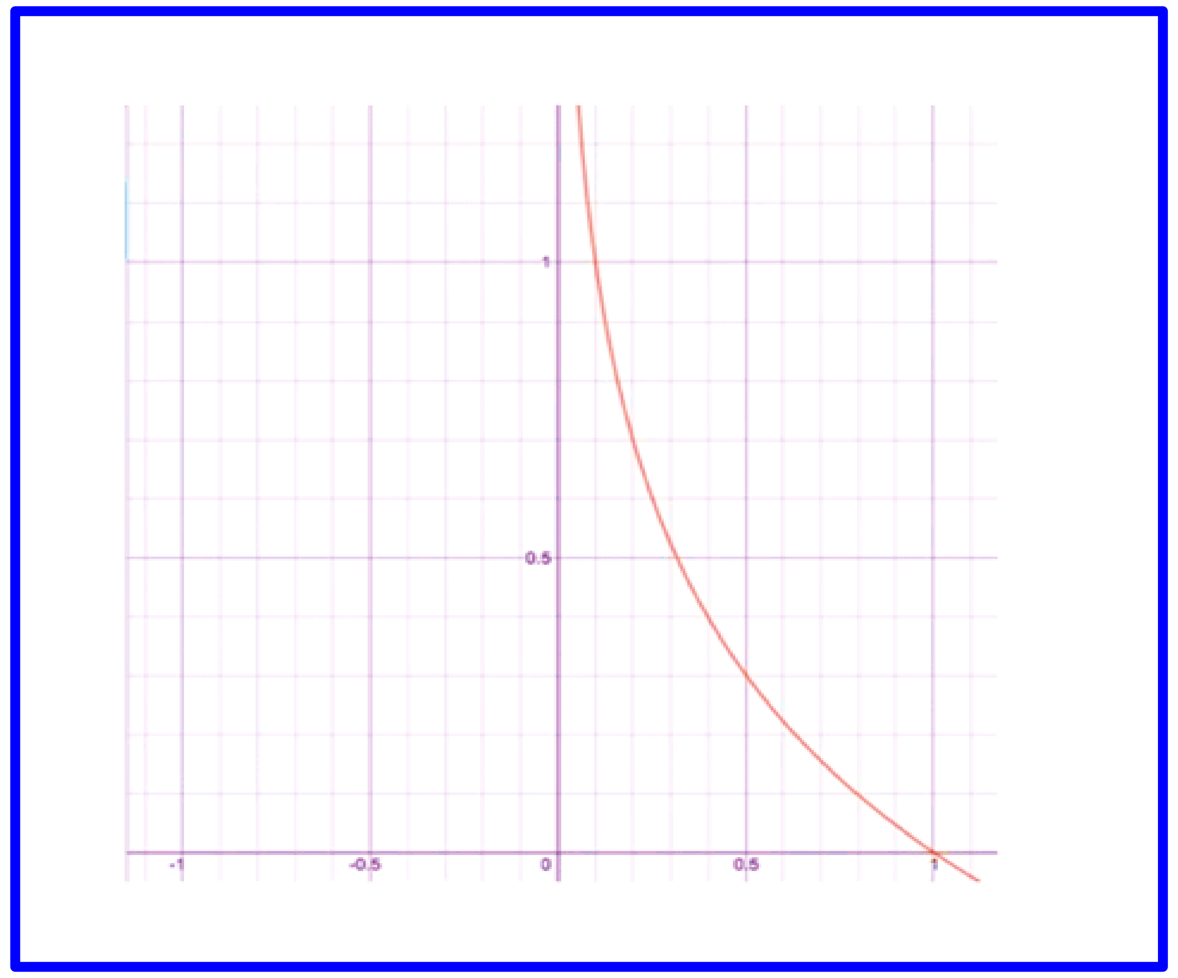
$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot -\log \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{bmatrix} \infty \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0$$

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

cost function

Cross-entropy cost function

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



$$L = \overset{\boxed{A}}{\underset{\boxed{B}}{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}} = \boxed{A}$$

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

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

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

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

cost function

Logistic cost VS cross entropy

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

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

=

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

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

Difference (Prediction of Softmax , True value)

H(x) Label

Cost function

SOFTMAX (Prediction)

Loss (Cost)

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

Training set

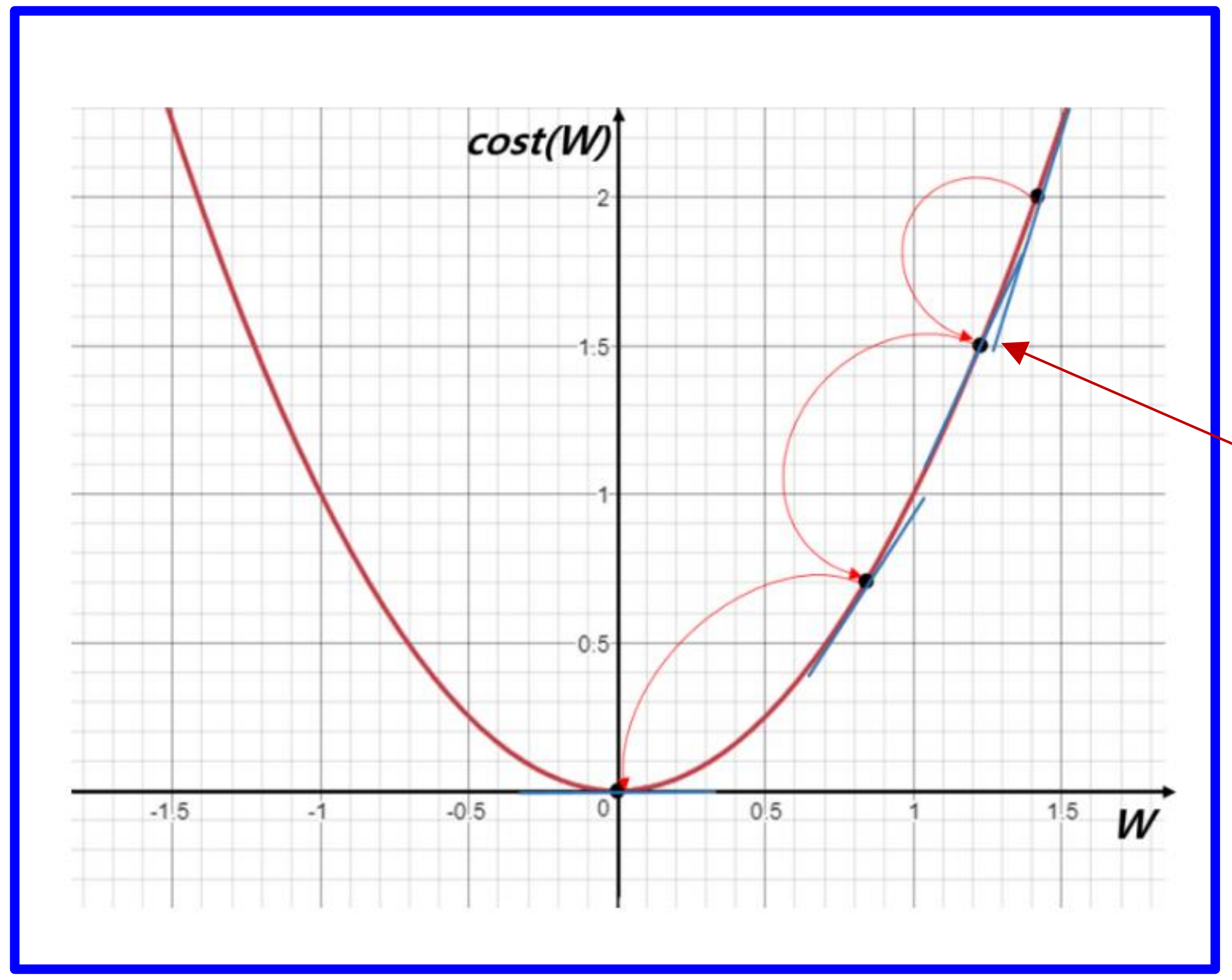
True value (Lable)

Prediction (Softmax)

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

Gradient descent

Cost (Loss)



STEP (learning rate)

Loss (Cost)

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

Training set