

Lecture 6-I

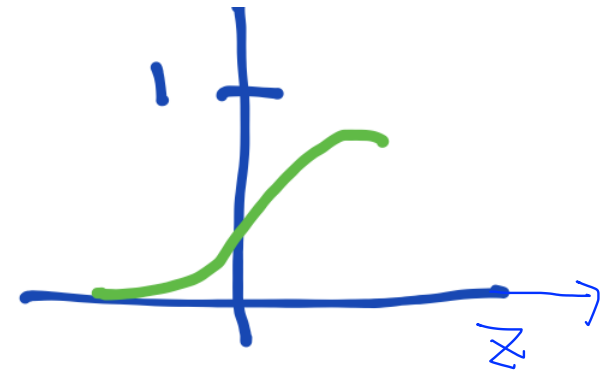
Softmax classification:
Multinomial classification

Sung Kim <hunkim+mr@gmail.com>

Logistic regression

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

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

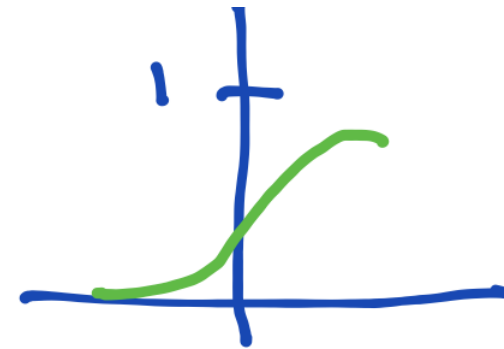


Logistic regression

$$H_L(x) = Wx$$

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

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



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

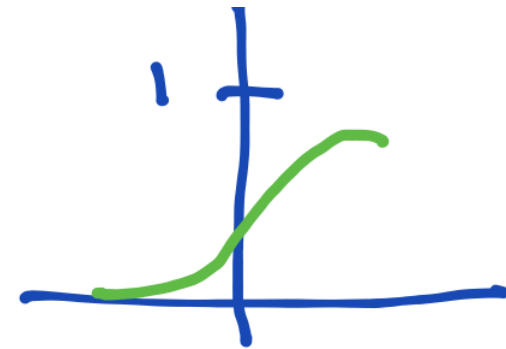
Logistic regression

$$H_L(x) = Wx$$

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

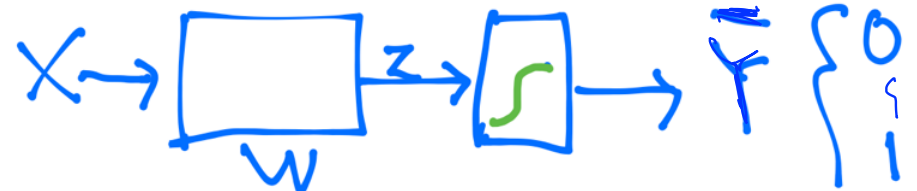
$$g(z) = \frac{1}{1 + e^{-z}}$$

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

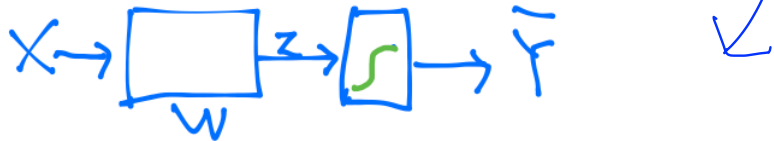


y : real

\hat{y} : prediction
($H(x)$)



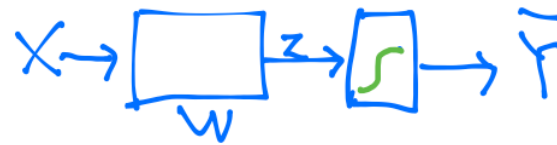
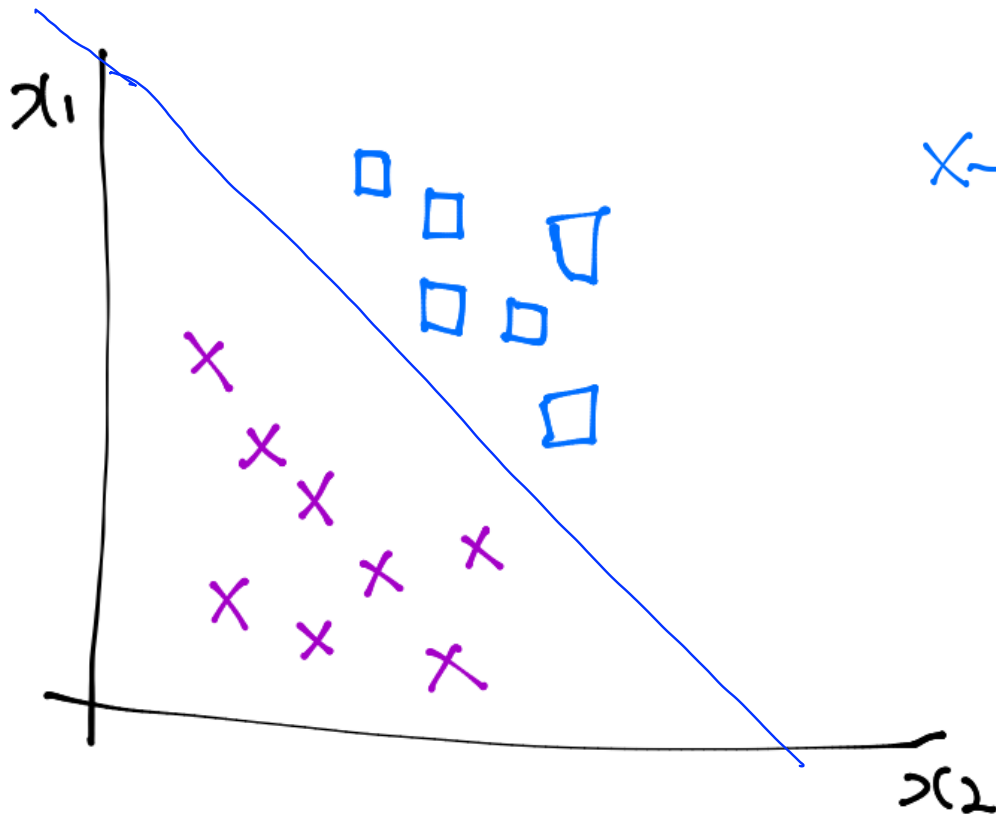
Logistic regression

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


The diagram illustrates the architecture of a logistic regression model. It shows an input X being processed by a linear layer (represented by a box with weight w below it) to produce an intermediate value z . This value z is then passed through a sigmoid activation function (represented by a box with a green S-curve) to produce the final output \hat{Y} . A blue arrow points from the equation $H_R(x) = g(H_L(x))$ down to the diagram, indicating the relationship between the mathematical formulation and the model structure.

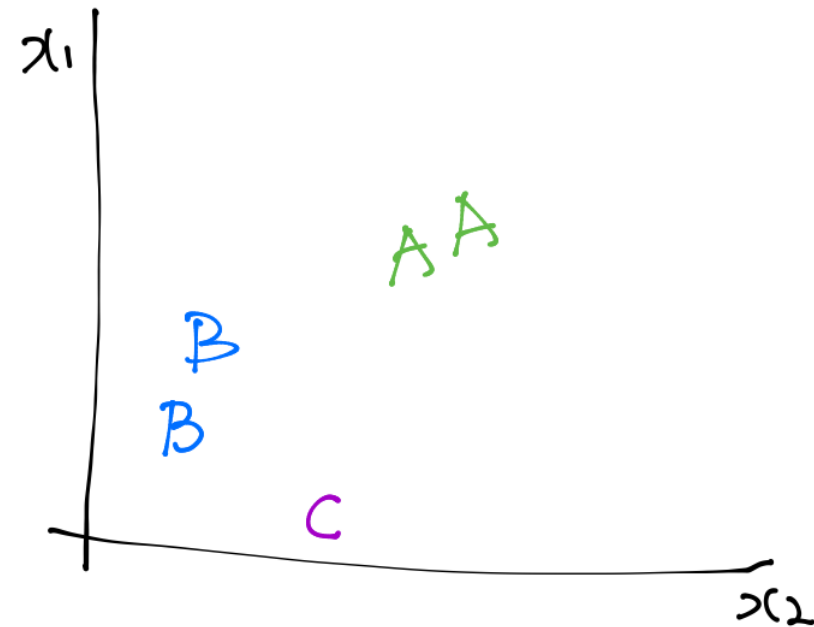
Logistic regression

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

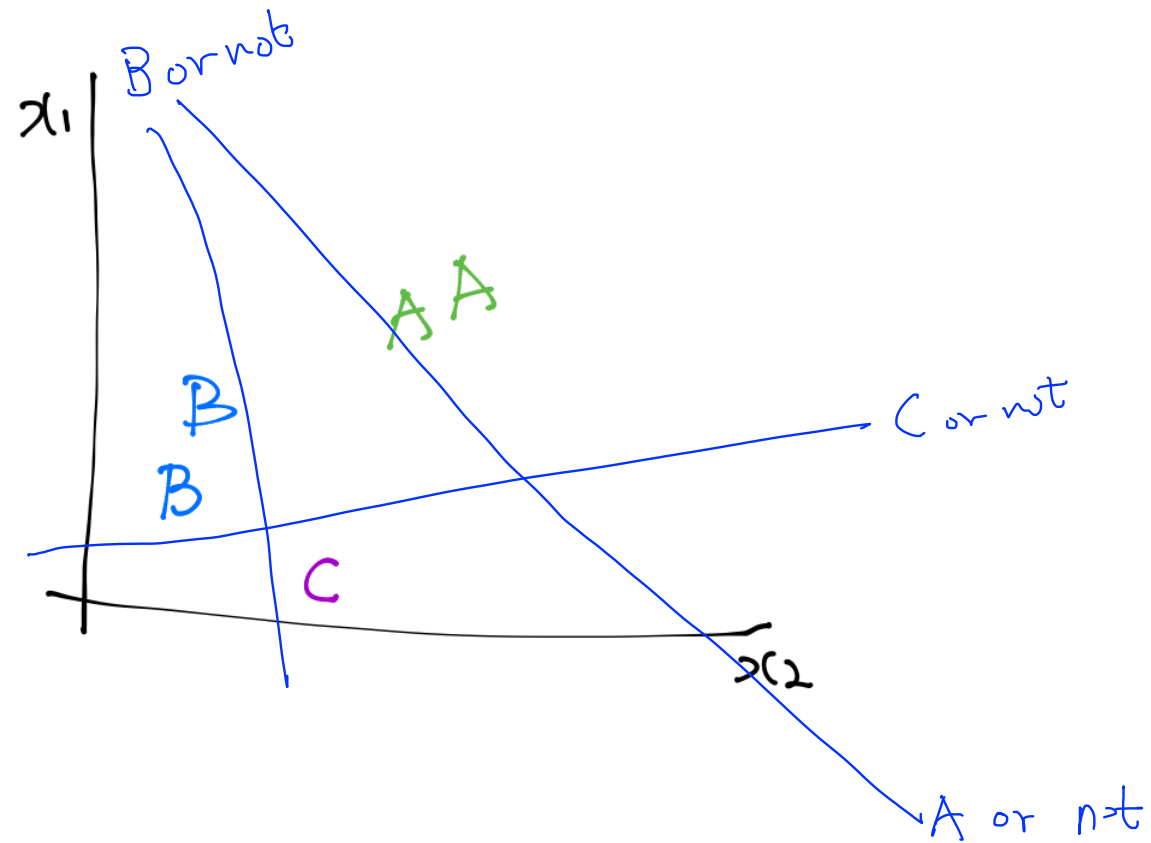


Multinomial classification

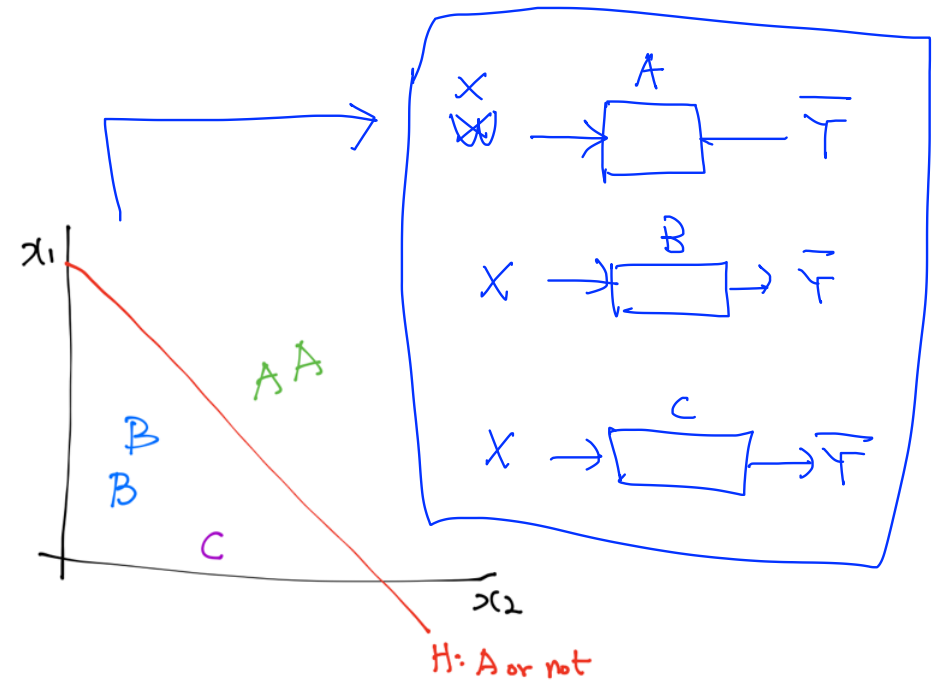
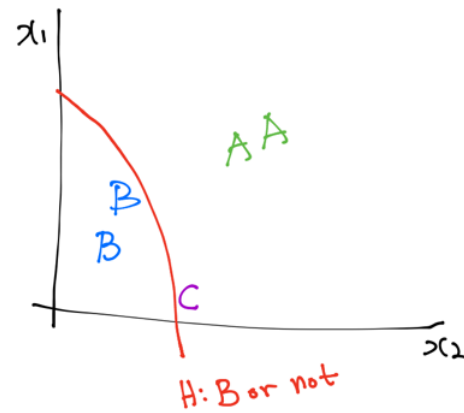
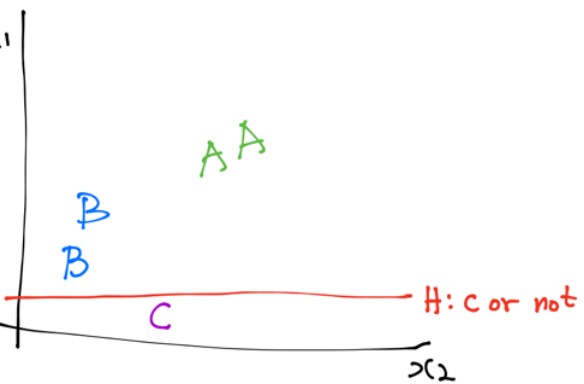
x1 (hours)	x2 (attendance)	y (grade)
10	5	A
9	5	A
3	2	B
2	4	B
11	1	C



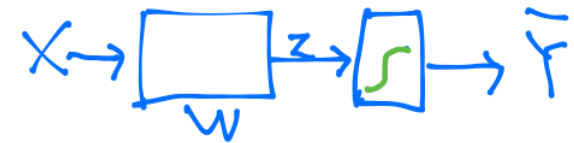
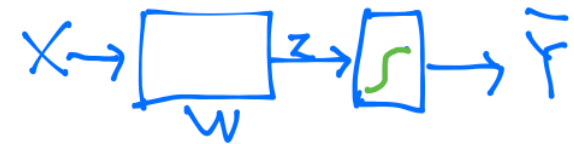
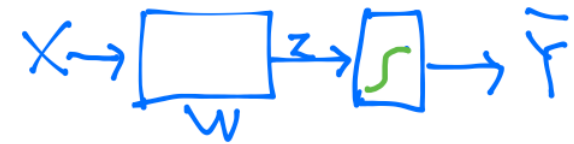
Multinomial classification



Multinomial classification

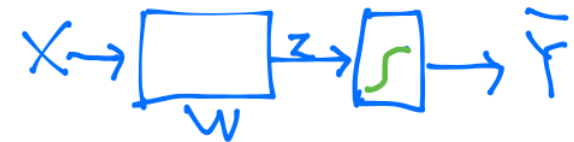
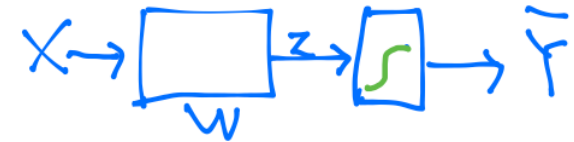
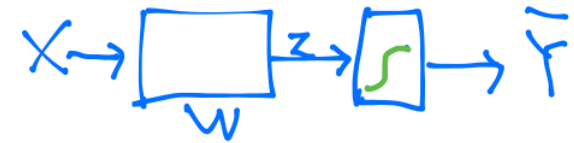


Multinomial classification



Multinomial classification

$$\begin{array}{c}
 [w_1 \quad w_2 \quad w_3] \\
 \underbrace{\hspace{1.5cm}}_w
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\
 \underbrace{\hspace{1cm}}_X
 \end{array}
 = \underbrace{[w_1 x_1 + w_2 x_2 + w_3 x_3]}_{H(X)}$$

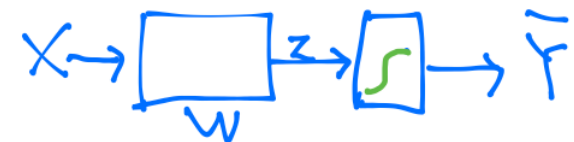
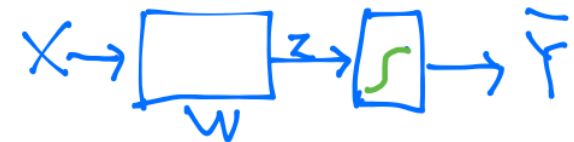
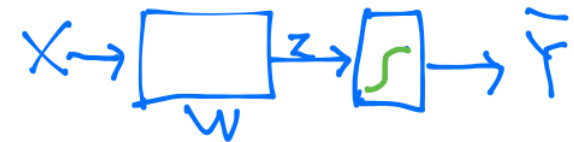


Multinomial classification

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

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

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

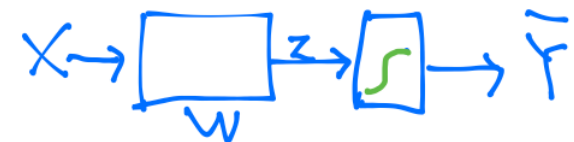
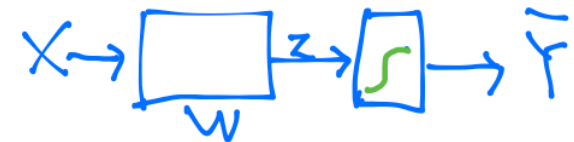
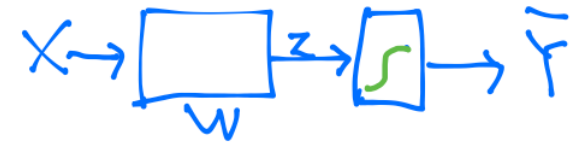


Multinomial classification

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



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



Matrix multiplication

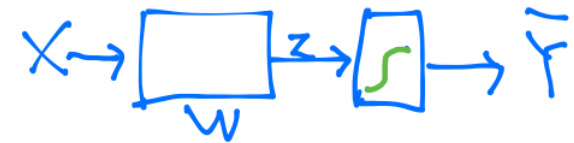
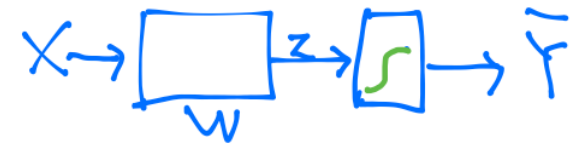
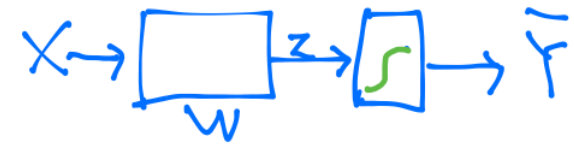
The diagram illustrates the dot product of two matrices. The first matrix is $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ and the second matrix is $\begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}$. A yellow curved arrow labeled "Dot Product" connects the first row of the first matrix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

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



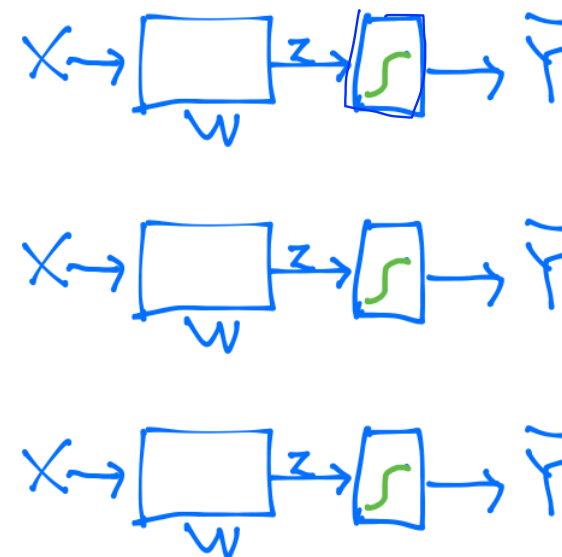
Multinomial classification

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

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

$X \rightarrow \boxed{w} \xrightarrow{z} \boxed{\int} \rightarrow \hat{Y}$
 $X \rightarrow \boxed{w} \xrightarrow{z} \boxed{\int} \rightarrow \hat{Y}$
 $X \rightarrow \boxed{w} \xrightarrow{z} \boxed{\int} \rightarrow \hat{Y}$

Where is sigmoid?

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


The diagram illustrates the process of a linear layer followed by a sigmoid activation function. It shows three identical rows of the process:

- Input X is fed into a box labeled W (representing the weight matrix).
- The output of the box is labeled z .
- The value z is then passed into a box containing a green sigmoid function symbol σ .
- The final output of the sigmoid function is labeled \hat{Y} .

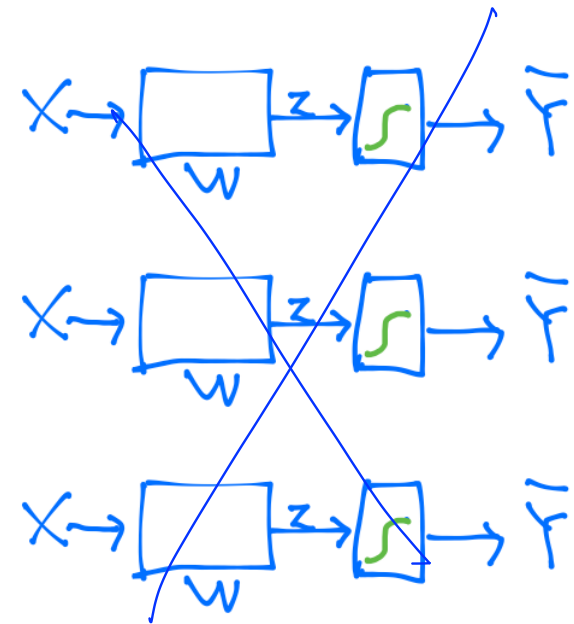
Lecture 6-2

Softmax classification:
softmax and cost function

Sung Kim <hunkim+mr@gmail.com>

Where is sigmoid?

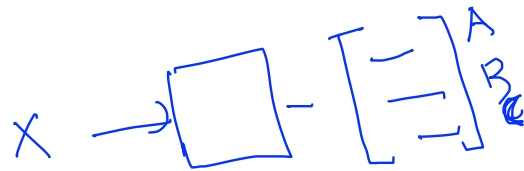
$$\begin{bmatrix} w_{A1} & w_{A2} & w_{A3} \\ w_{B1} & w_{B2} & w_{B3} \\ w_{C1} & w_{C2} & w_{C3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} w_{A1}x_1 + w_{A2}x_2 + w_{A3}x_3 \\ w_{B1}x_1 + w_{B2}x_2 + w_{B3}x_3 \\ w_{C1}x_1 + w_{C2}x_2 + w_{C3}x_3 \end{bmatrix}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



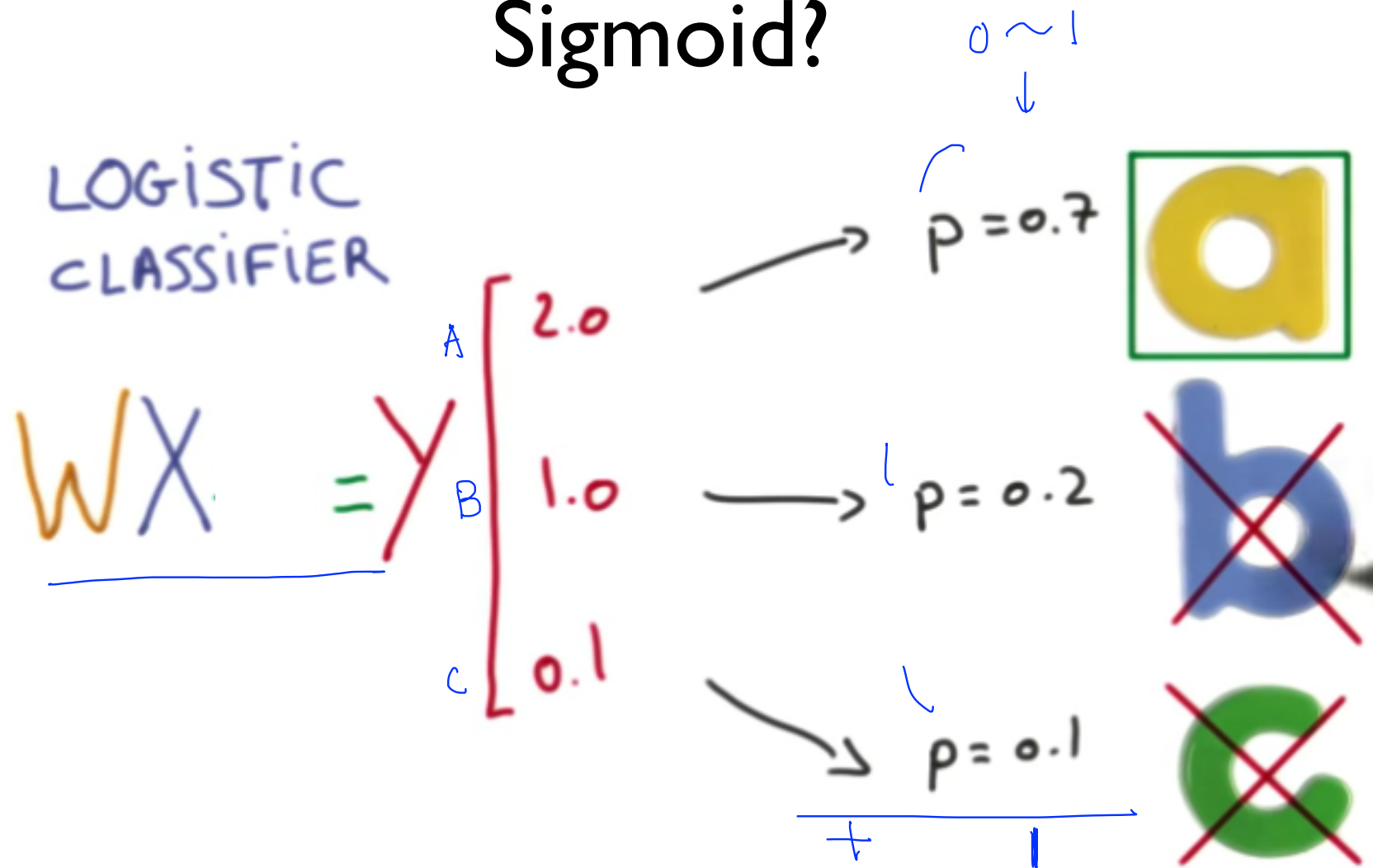
Where is sigmoid?

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

$\begin{matrix} 0 \sim 1 \\ \hline \downarrow \end{matrix}$
 $\begin{bmatrix} 2.0 \\ 1.0 \\ 0.1 \end{bmatrix}$

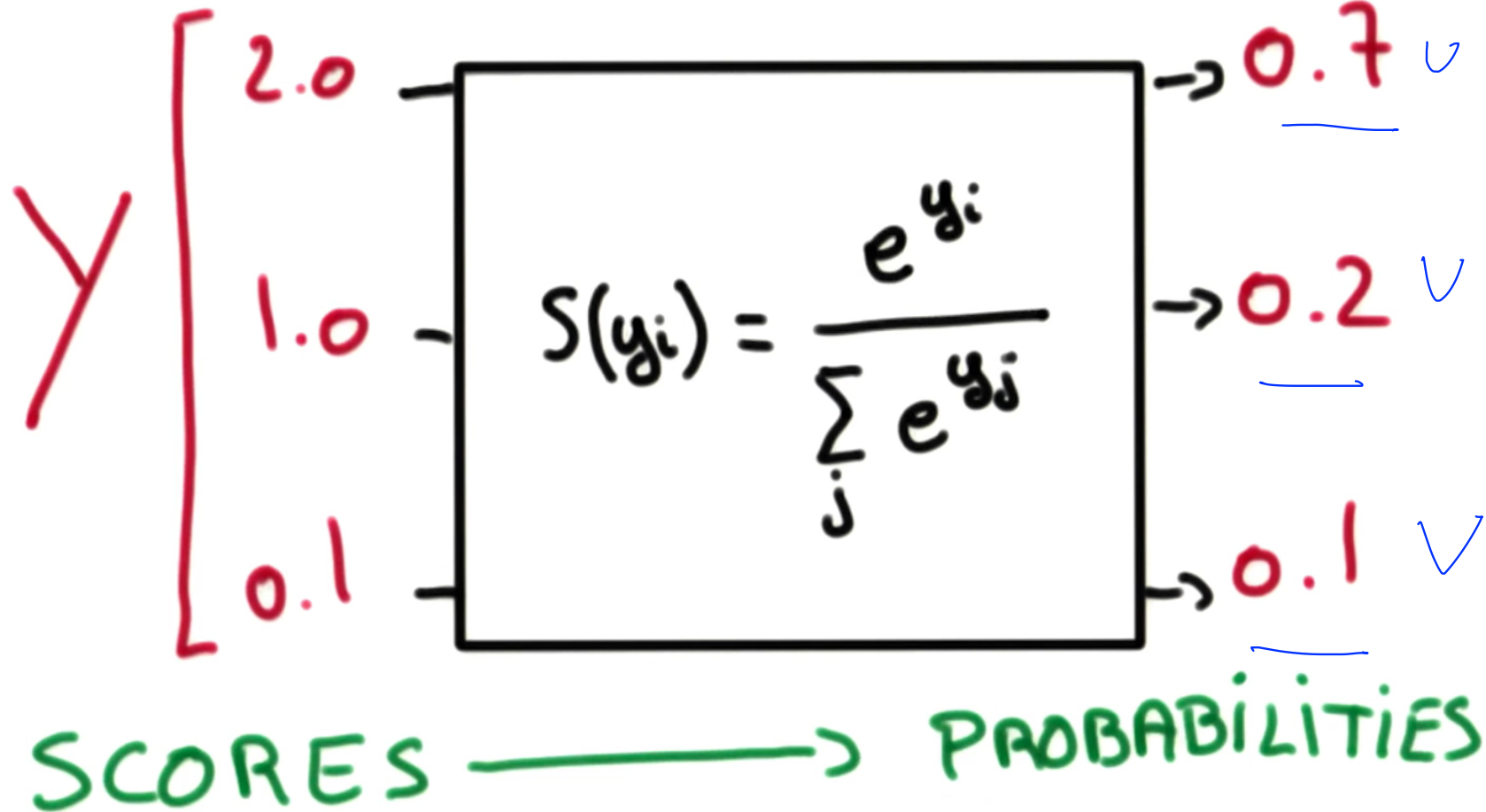


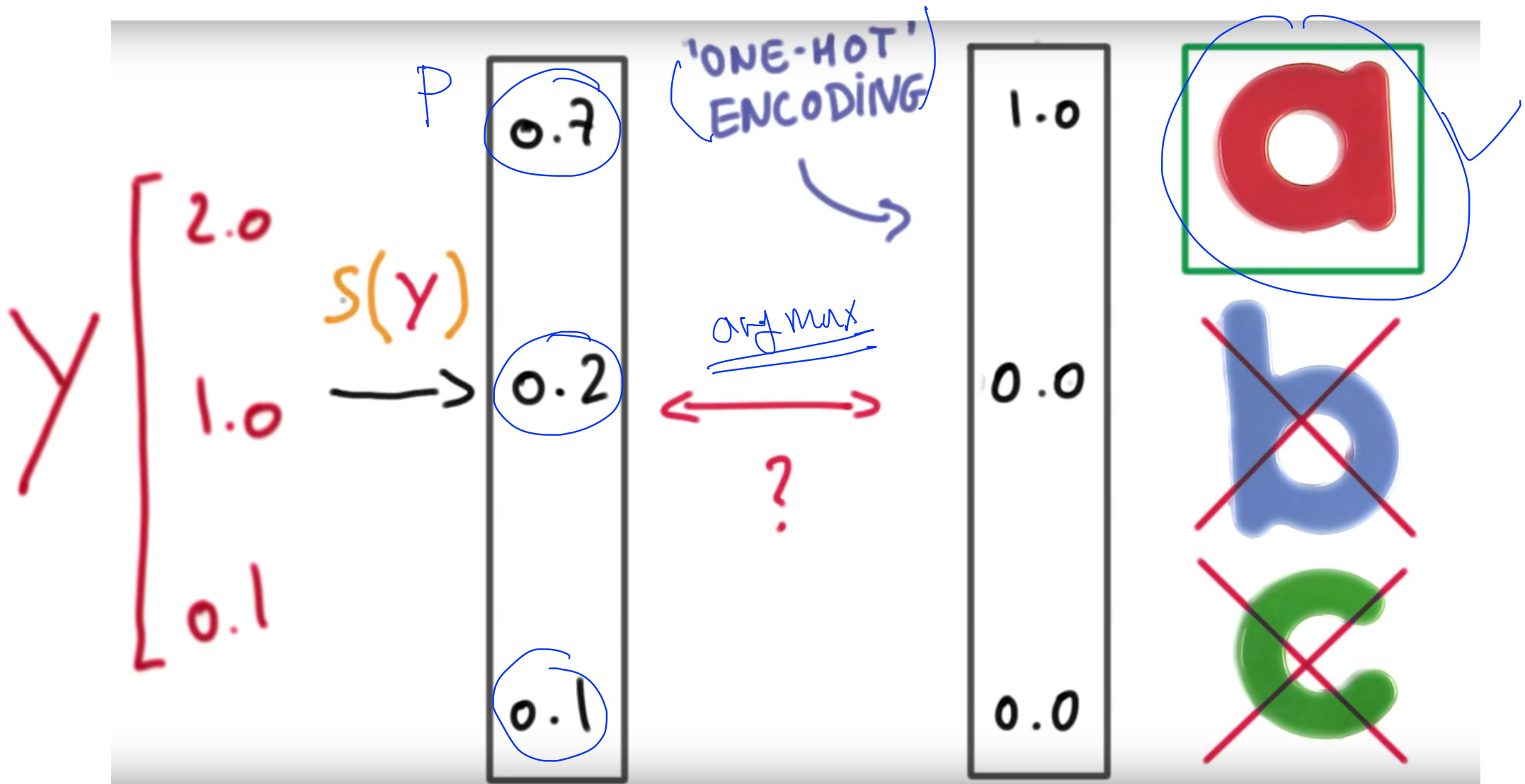
Sigmoid?



SOFTMAX

- ① $0 \sim 1$
- ② $\sum = 1$





Cost function

CROSS-ENTROPY

$S(Y) = Y$

0.7
0.2
0.1

$L = Y$

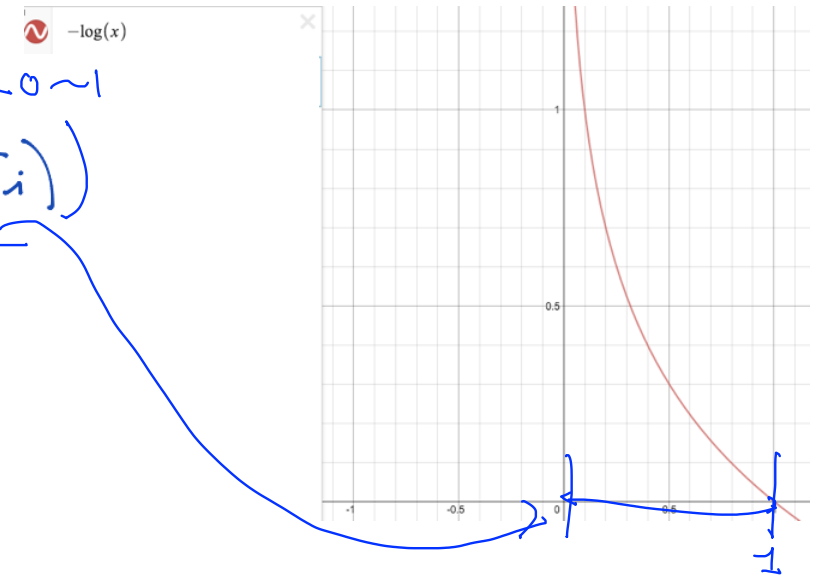
1.0
0.0
0.0

$$D(S, L) = - \sum_i L_i \log(S_i)$$

Cross-entropy cost function

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

$$-\sum_i L_i \log(\bar{y}_i) = \sum_i (L_i * \underline{\underline{-\log(\bar{y}_i)}})$$



Cross-entropy cost function

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

$$-\sum_i L_i \log(\bar{y}_i) = \sum_i \underline{L_i} * \underline{-\log(\bar{y}_i)}$$

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

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

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



Cross-entropy cost function

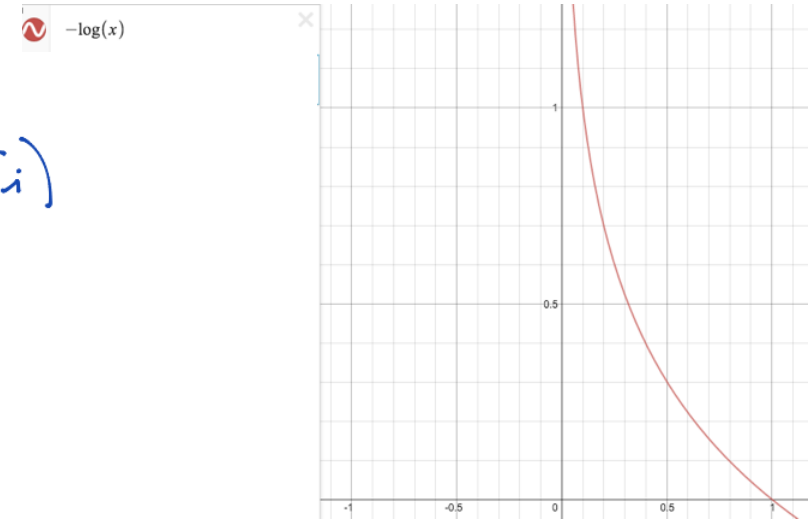
$$-\sum_i L_i \log(s_i)$$

$$-\sum_i L_i \log(\bar{y}_i) = \sum_i L_i * \underline{-\log(\bar{y}_i)}$$

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

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

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



Logistic cost VS cross entropy

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

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

// ?

Cost function

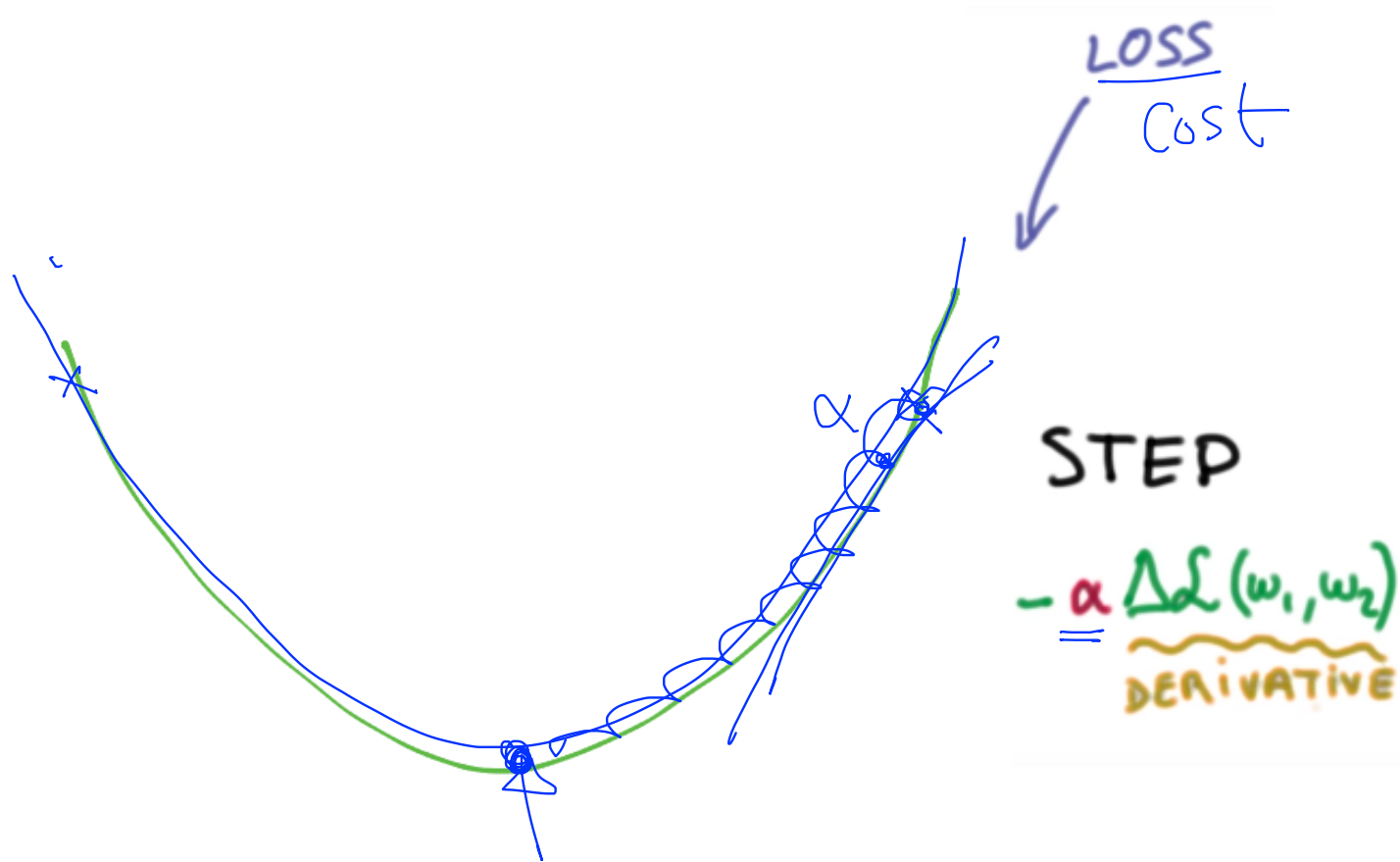
A handwritten diagram illustrating the cost function formula. The formula is $J = \frac{1}{N} \sum_i \mathcal{D}(s(wx_i + b), L_i)$. The word "LOSS" is written in blue above the formula, with an arrow pointing to the \mathcal{D} term. The term $\frac{1}{N}$ is circled in blue. The summation symbol \sum is green, and the index i is red. The function s is orange, w is blue, x_i is green, b is orange, and L_i is blue. The word "TRAINING SET" is written in blue below the formula, with two arrows pointing to the x_i and L_i terms.

LOSS

$$J = \frac{1}{N} \sum_i \mathcal{D}(s(wx_i + b), L_i)$$

TRAINING SET

Gradient descent



Next Applications & Tips

