Classification (Short Version)

Hung-yi Lee 李宏毅

Classification

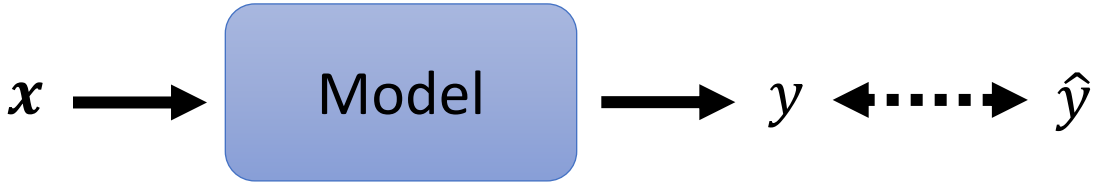
- To learn more ……

https://youtu.be/fZAZUYEeIMg
(in Mandarin)

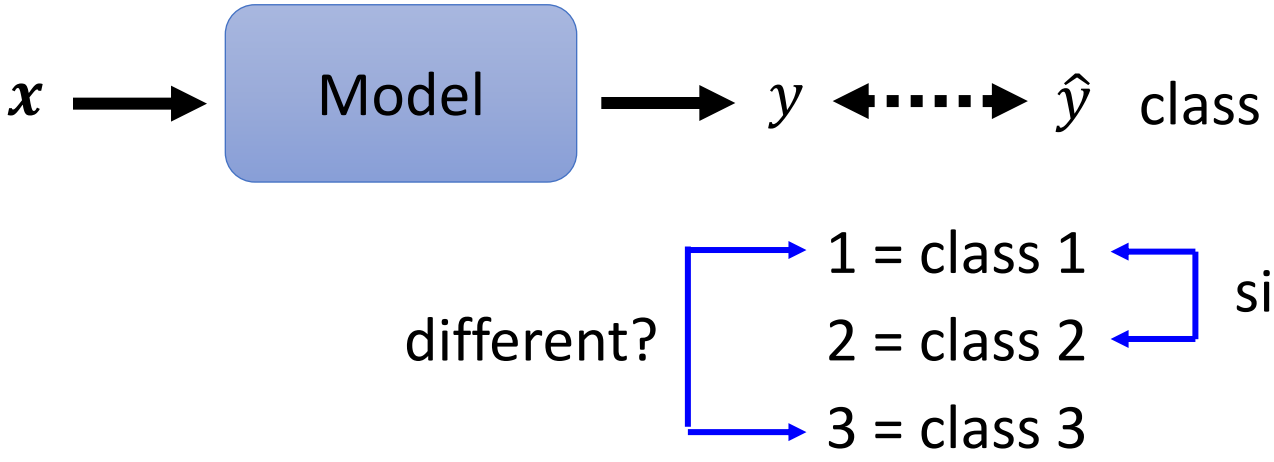
https://youtu.be/hSXFuypLukA
(in Mandarin)

Classification as Regression?

- Regression



- Classification as regression?



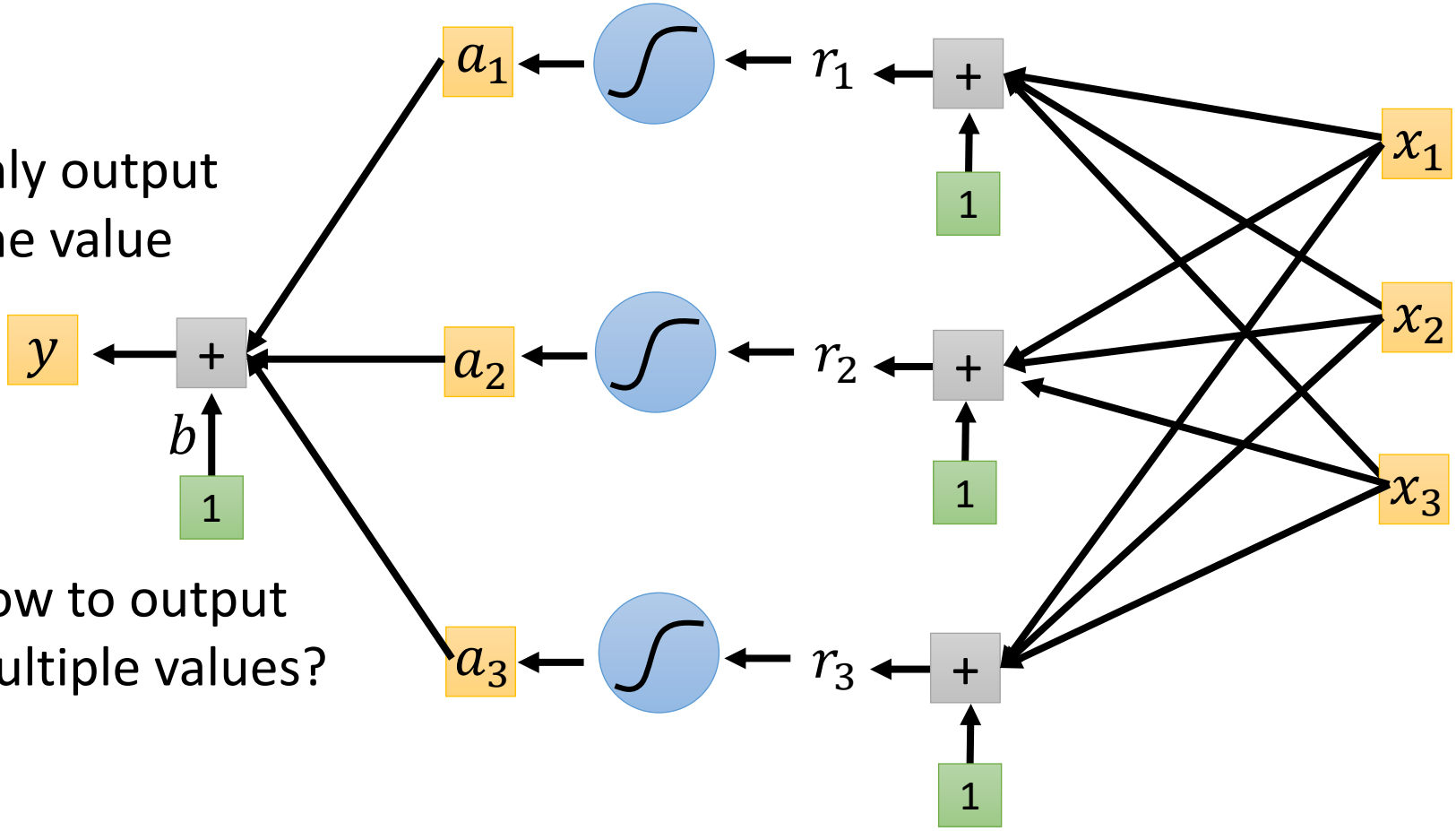
Class as one-hot vector

Class 1 Class 2 Class 3

$$\hat{y} = \begin{bmatrix}1\\0\\0\end{bmatrix} \text{ or } \begin{bmatrix}0\\1\\0\end{bmatrix} \text{ or } \begin{bmatrix}0\\0\\1\end{bmatrix}$$

only output one value

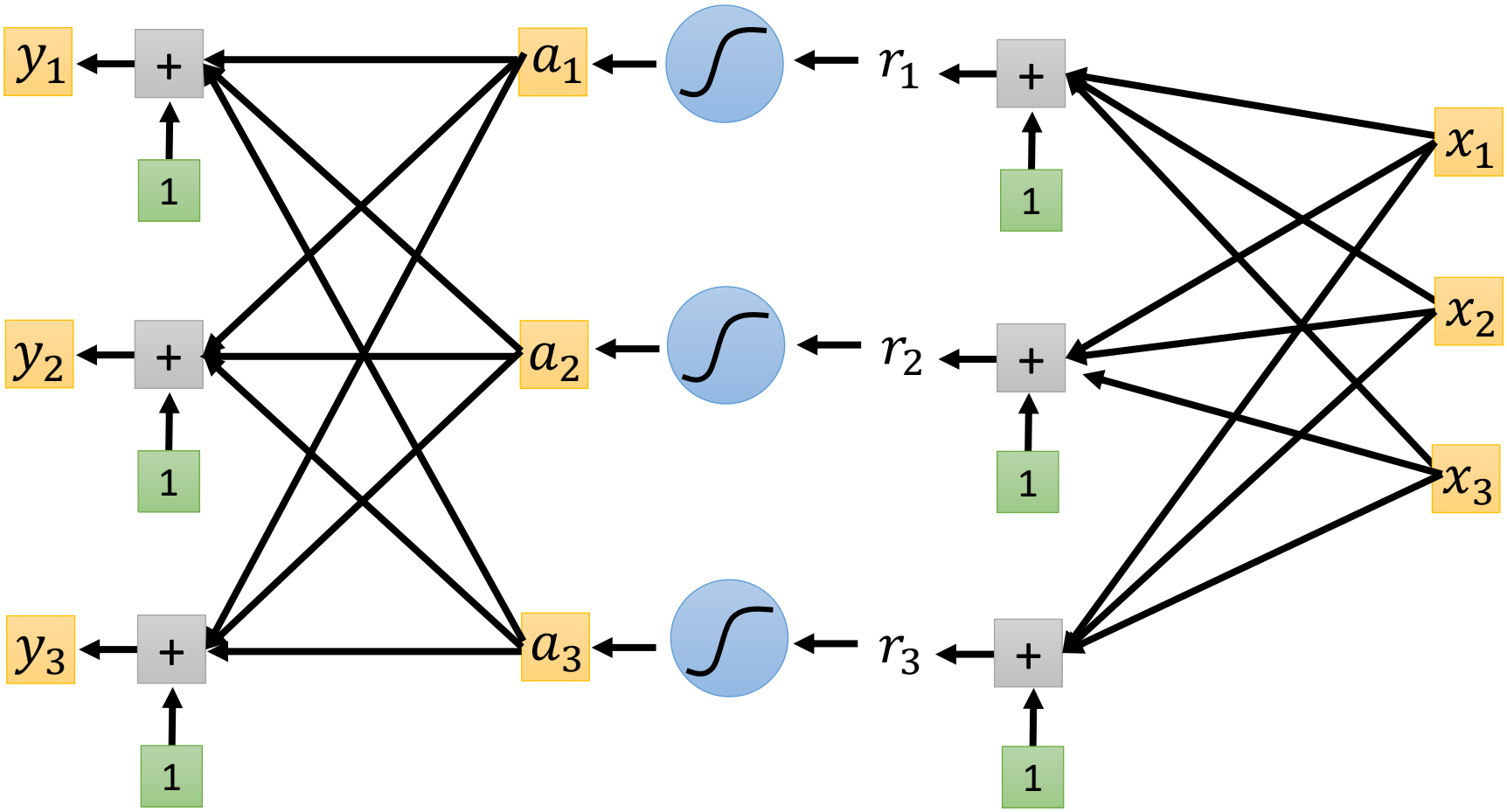
How to output multiple values?



Class as one-hot vector

Class 1 Class 2 Class 3

$$\hat{y} = \begin{bmatrix}1\\0\\0\end{bmatrix} \text{ or } \begin{bmatrix}0\\1\\0\end{bmatrix} \text{ or } \begin{bmatrix}0\\0\\1\end{bmatrix}$$



Regression

label

feature

$$\hat{y} \longleftrightarrow y = b + c^T \sigma(b + W x)$$

Classification

feature

$$y = b' + W' \sigma(b + W x)$$

label

$$\hat{y} \longleftrightarrow y' = softmax(y)$$

0 or 1

Make all values between 0 and 1

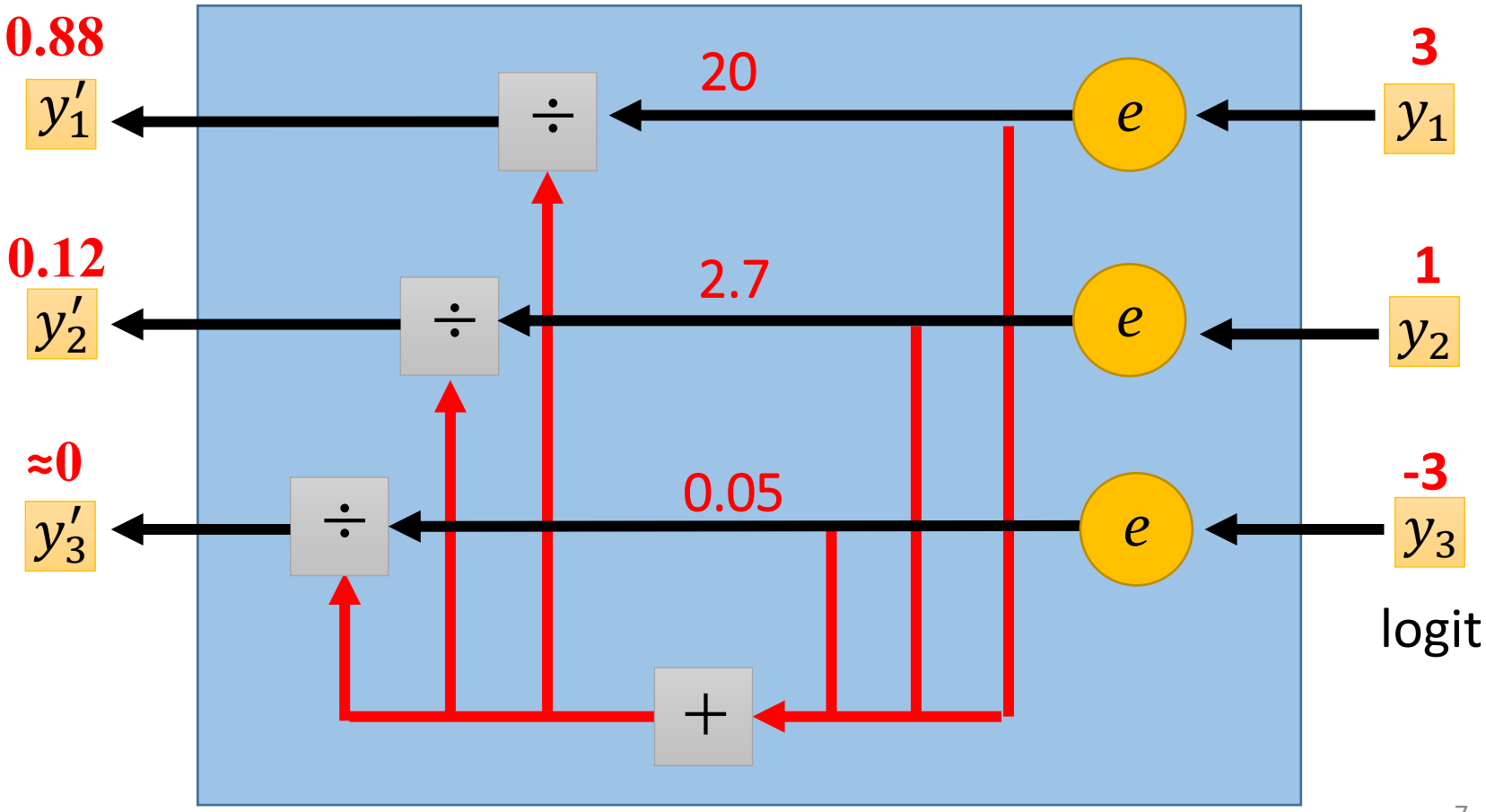
Can have any value

Soft-max

$$y_i' = \frac{exp(y_i)}{\sum_j exp(y_i)}$$

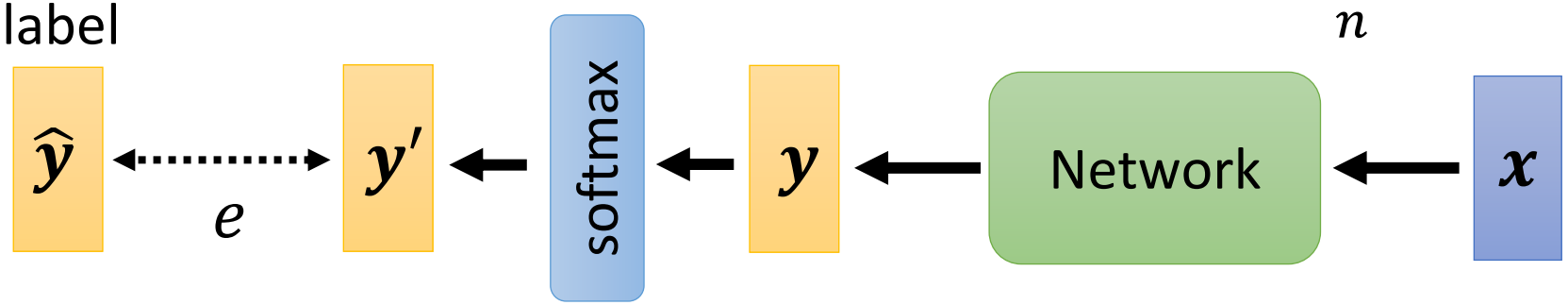
- $1 > y_i' > 0$
- $\sum_i y_i' = 1$

How about **binary classification**? ☺



Loss of Classification

$$L = \frac{1}{N}\sum_{n} e_n$$



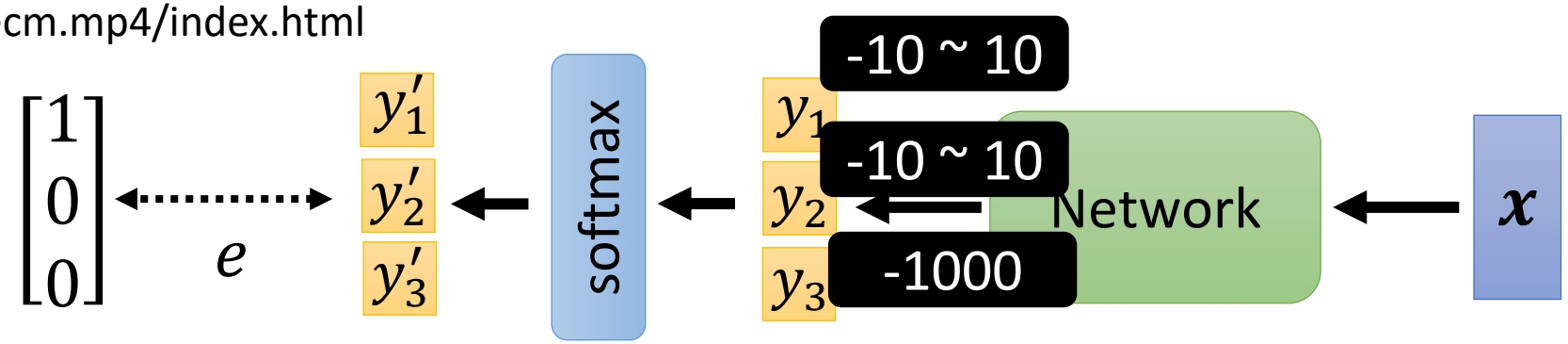
Mean Square Error (MSE) $e = \sum_{i} (\hat{y}_i - y'_i)^2$

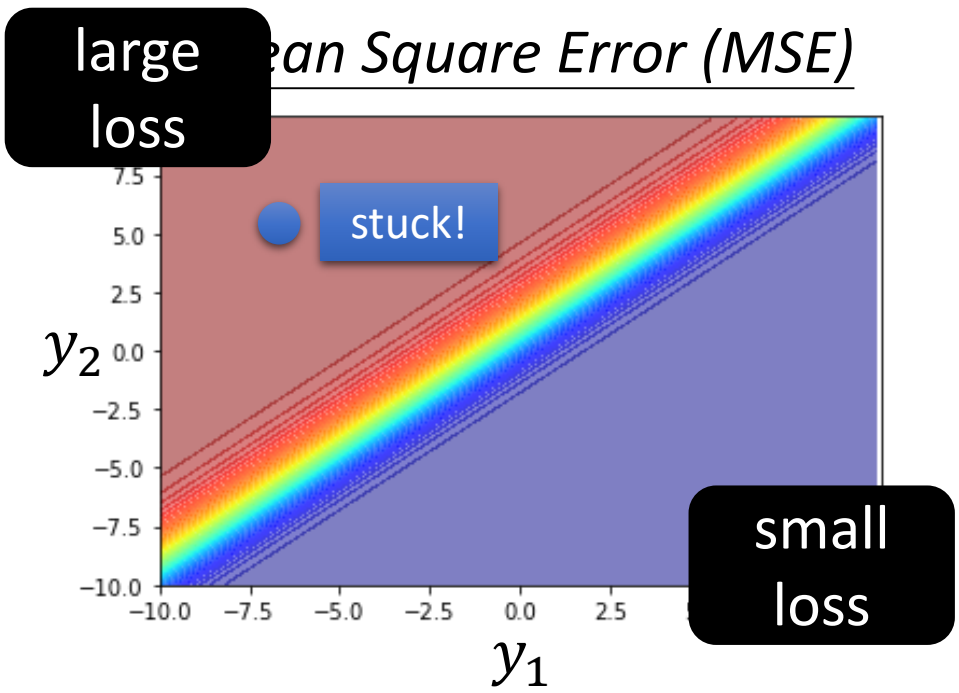
Cross-entropy 

$e = -\sum_{i} \hat{y}_i ln y'_i$

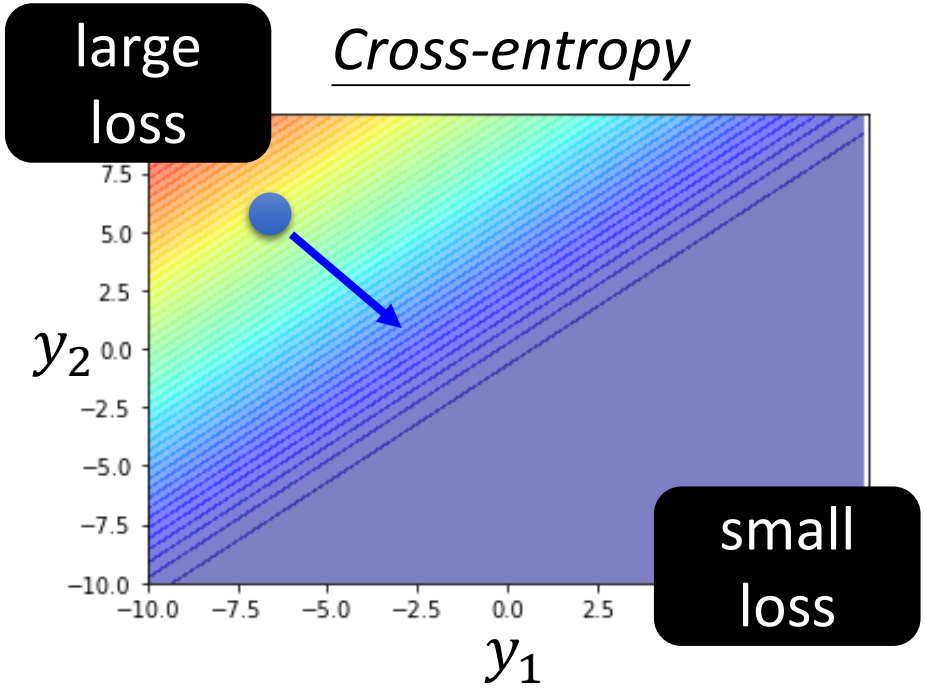
Minimizing cross-entropy is equivalent to **maximizing likelihood**.

http://speech.ee.ntu.edu.tw/~tlkagk/courses/MLDS_2015_2/Lecture/Deep%20More%20(v2).ecm.mp4/index.html





ean Square Error (MSE)



Cross-entropy

Changing the loss function can change the difficulty of optimization.