Classification (Short Version)

Hung-yi Lee 李宏毅

Classification

- To learn more ……

https://youtu.be/fZAZUYEeIMg

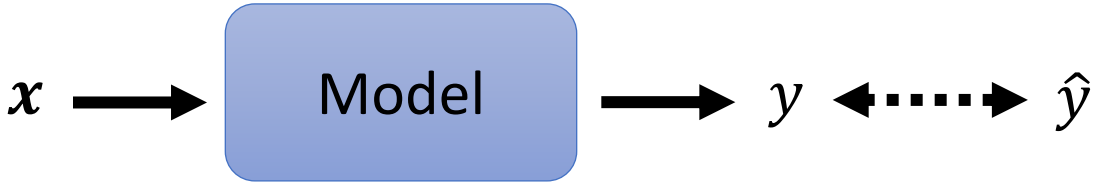
(in Mandarin)

https://youtu.be/hSXFuypLukA

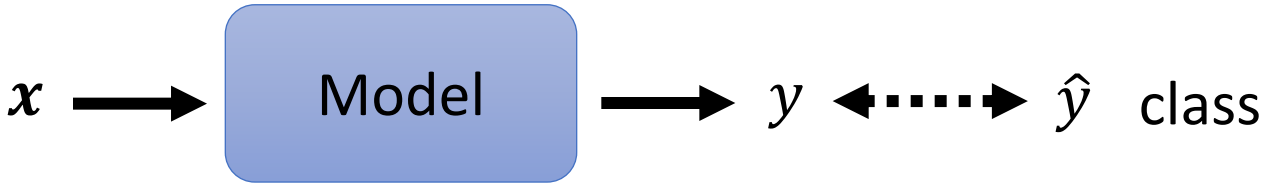
(in Mandarin)

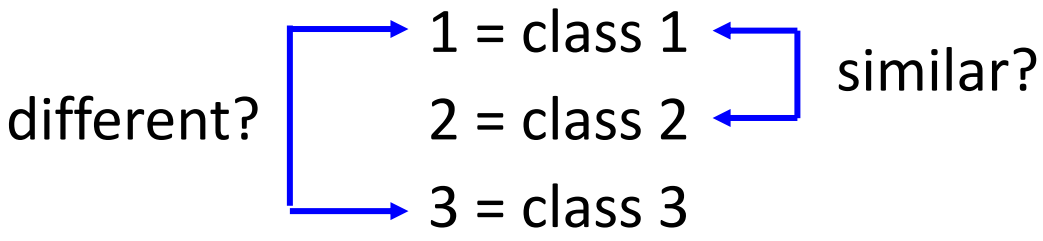
Classification as Regression?

- Regression



- Classification as regression?

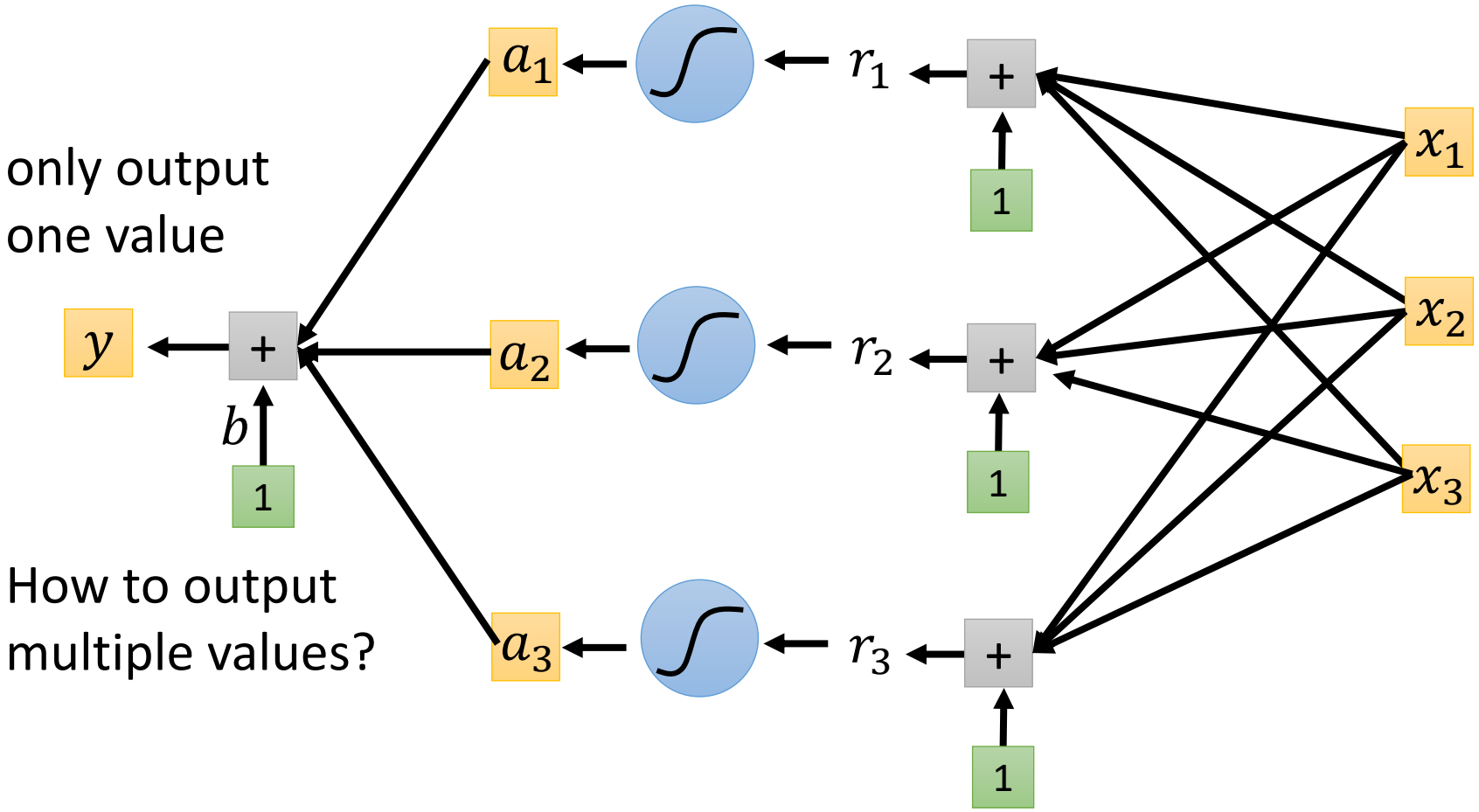




Class as one-hot vector

Class 1 Class 2 Class 3

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



为什么说要用 one-hot 形式？

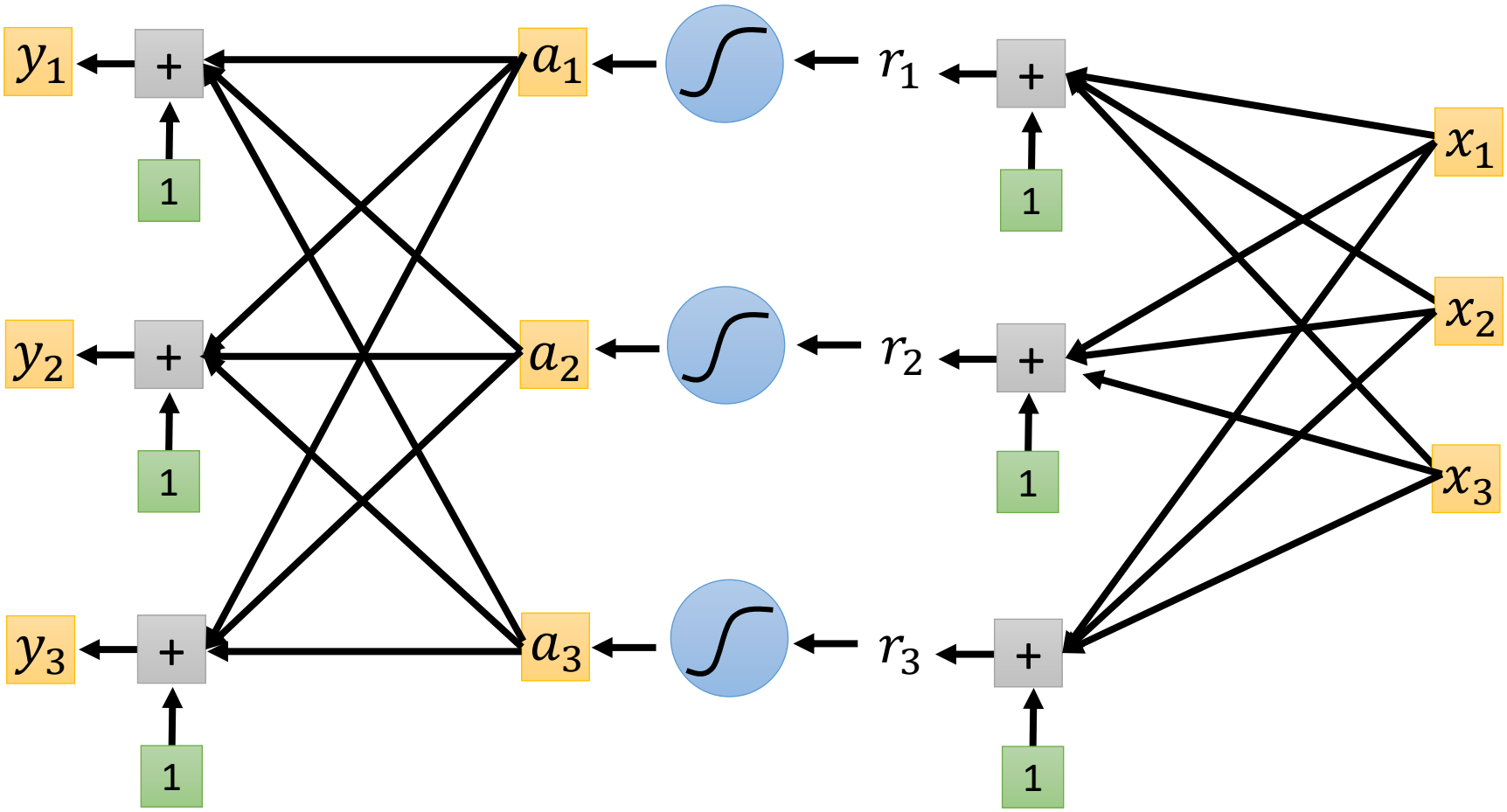
一是前文提到欧氏距离问题

二是 one-hot 本身就是真实概率分布，可用于交叉熵损失

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

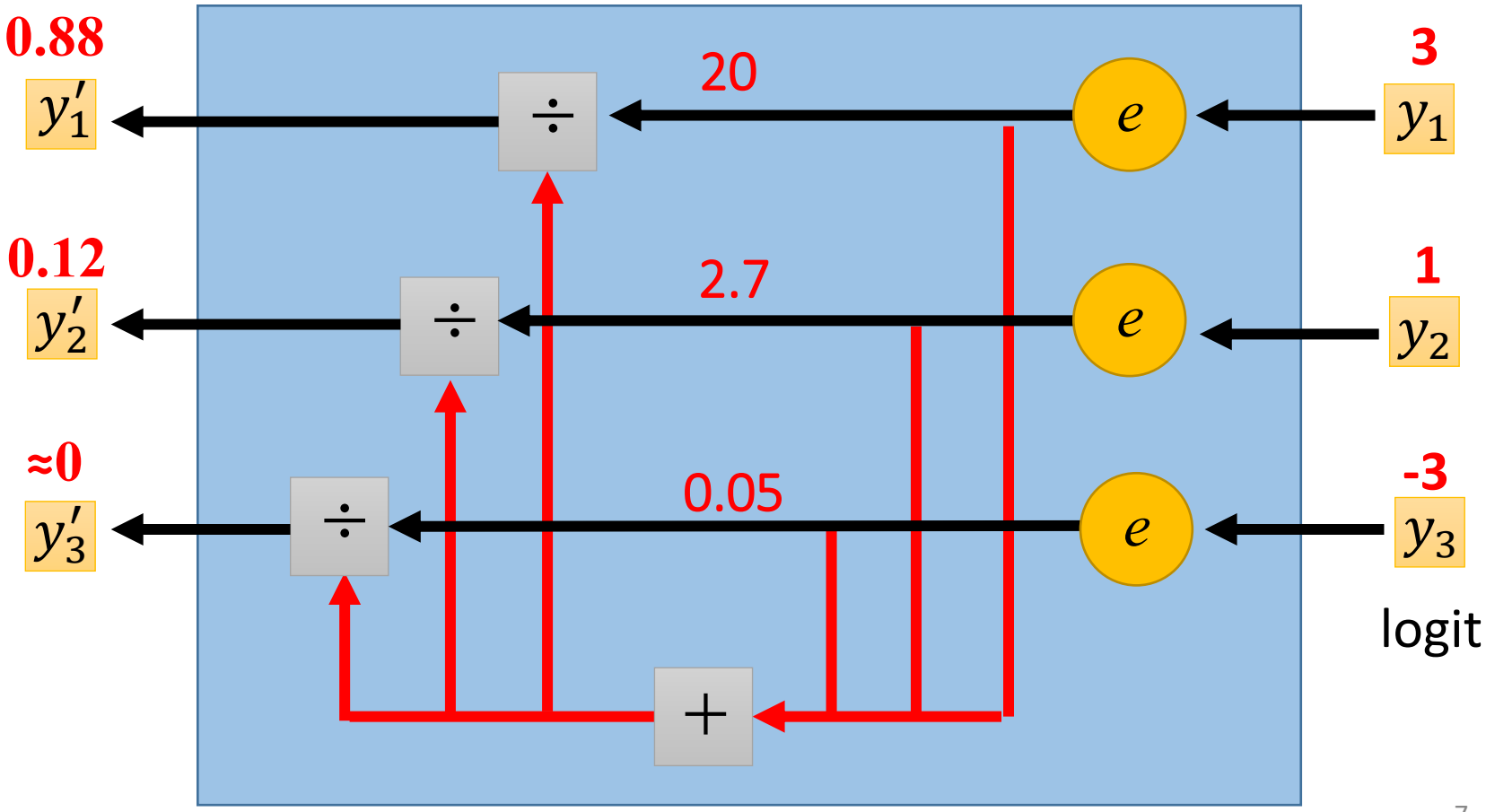
Can have any value

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**? ☺



为什么两个类的 softmax 和直接用 sigmoid 等价？

二分类时 $S(x) = \frac{1}{e^{z_1} + e^{z_2}} \cdot \begin{bmatrix} e^{z_1} \\ e^{z_2} \end{bmatrix}$

$= \frac{1}{1 + e^{z_2 - z_1}} \cdot \begin{bmatrix} 1 \\ e^{z_2 - z_1} \end{bmatrix}$

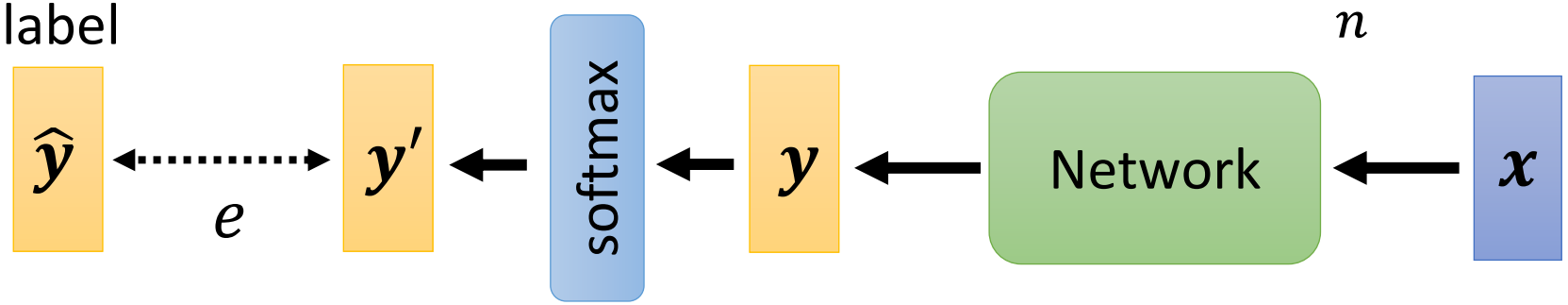
令 $z = z_1 - z_2$

则 $S(x) = \frac{1}{1 + e^{-z}} \begin{bmatrix} 1 \\ e^{-z} \end{bmatrix}$

$= \begin{bmatrix} \frac{1}{1+e^{-z}} \\ 1 - \frac{1}{1+e^{-z}} \end{bmatrix}$ = Logistic 回归

Loss of Classification

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



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

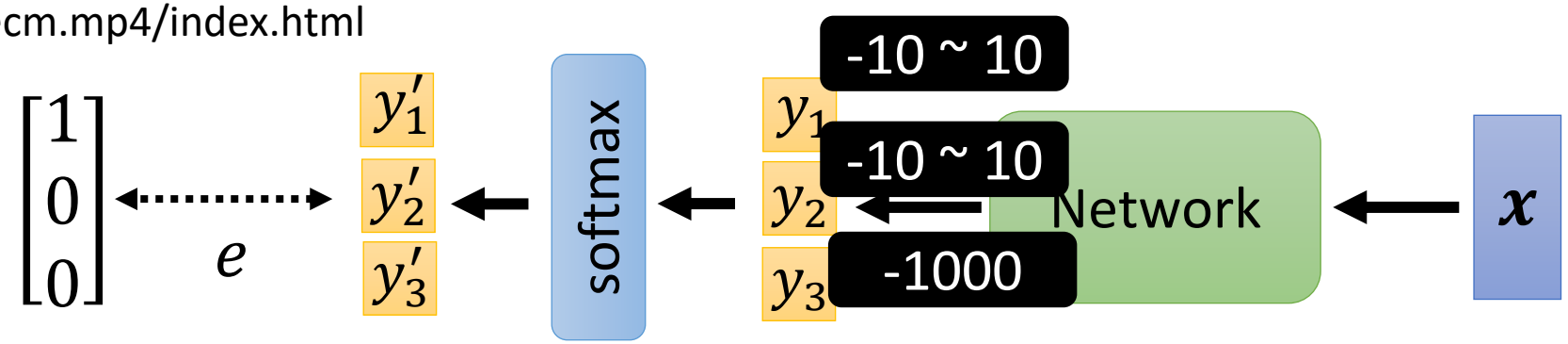
Cross-entropy 

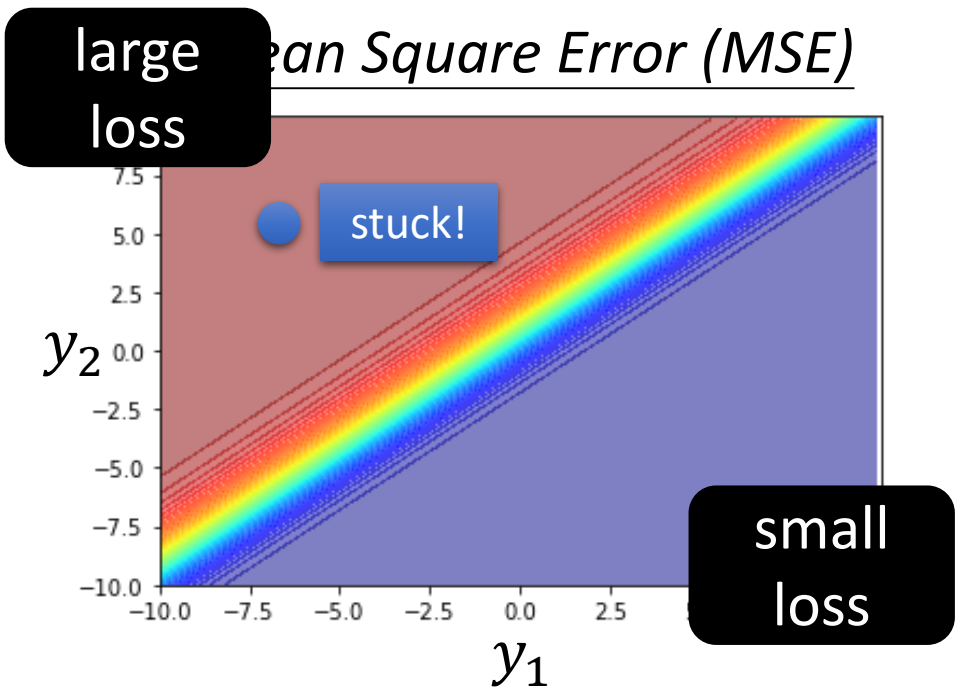
$e = -\sum_i \hat{y}_i ln y'_i$

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

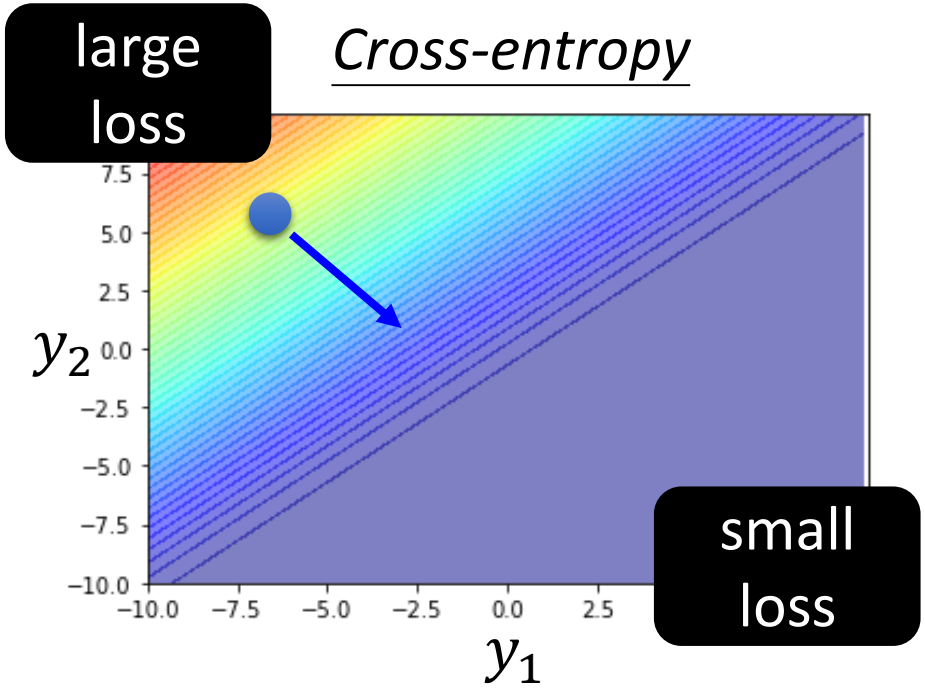
pytorch 中选用 cross-entropy 默认使用 softmax

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.