



02.3-DeepLearning-Loss of Classification

1. Classification as Regression

2. Class as one-hot vector

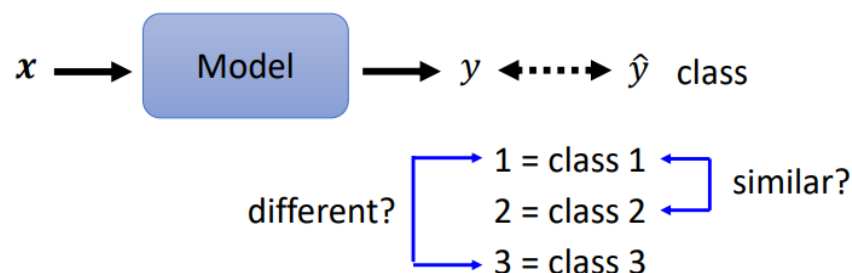
2.1 Classification with softmax

2.2 Loss of Classification

3. To Learn More

1. Classification as Regression

class 1 是編號1, class2是編號 2, class 3是編號 3,。希望模型的輸出 y 可以跟 class 的編號越接近越好



問題：

3 個 class 若分別設為1, 2, 3, 背後隱含 class 1 跟 class 2 比較相關, class 1 跟 class 3 比較不相關

- class 之間確實有相關性：假設根據身高體重預測是幾年級的小學生, 一年級真的跟二年級比較接近, 一年級真的跟三年級比較沒有關係
- class 之間沒有相關性：1, 2, 3 的假設可能會造成模型失準

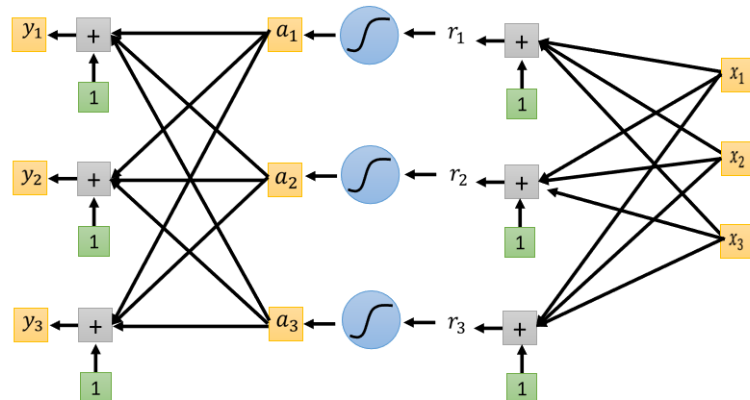
2. Class as one-hot vector

寫成向量形式, 任兩個 class 的距離都相同

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

Class 1 Class 2 Class 3

產生多個數值



2.1 Classification with softmax

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

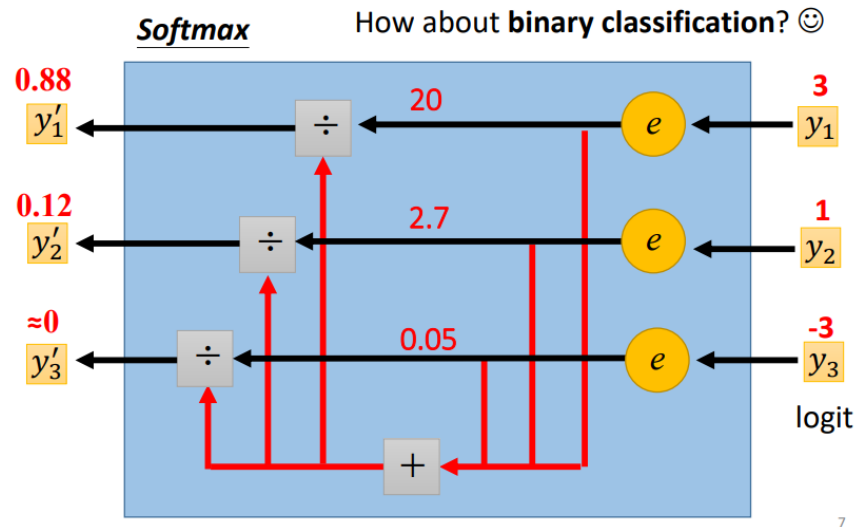
feature

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

0 or 1 Make all values between 0 and 1 Can have any value

當目標只有 0 跟 1，而 y 有任何值，可使用 **softmax** 先把它 **normalize** 到 0 到 1 之間，這樣才好跟 label 計算相似度

Soft-max
$$y'_i = \frac{\exp(y_i)}{\sum_j \exp(y_j)}$$
 $\blacksquare 1 > y'_i > 0$
 $\blacksquare \sum_i y'_i = 1$



經過計算後：

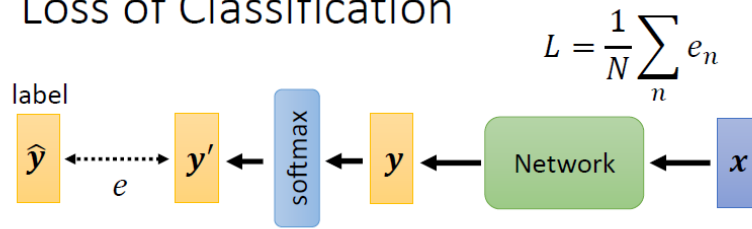
- 輸出值變成 0 到 1 之間
- 輸出值的和為 1
- 原本大的值跟小的值的**差距更大**

softmax 的輸入，稱作 **Logit**

二分類問題使用 **Sigmoid** 與 **Softmax** 是等價的

2.2 Loss of Classification

Loss of Classification



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.

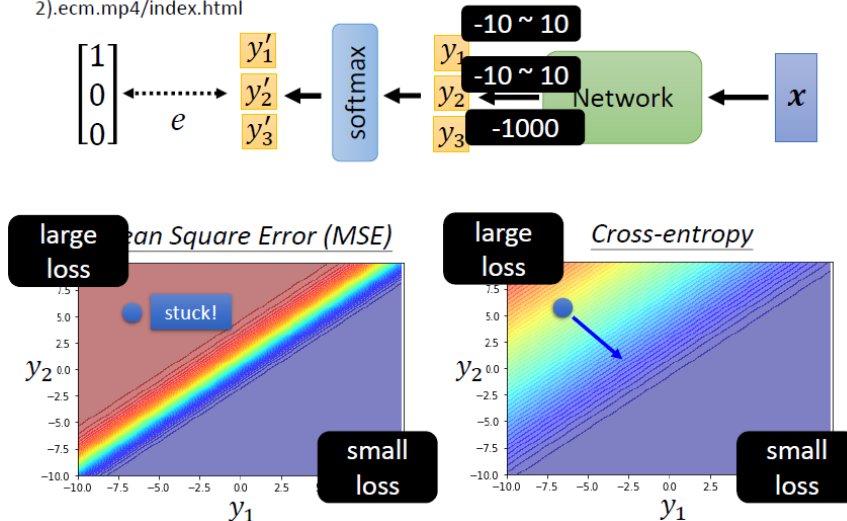
8

優化目標：減小 \hat{y} 和 y' 之間的差距 e

不同的損失函數：MSE, Cross-entropy, ...

選擇 **cross-entropy**，因為比 **MSE** 更加適用於分類問題！

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



Changing the loss function can change the difficulty of optimization.

9

從優化角度出發進行討論，使用 MSE 時，左上角的位置雖然 Loss 很大，但梯度平坦，難以優化；而 Cross-entropy 則更容易收斂 ⇒ **改變 Loss function**，也會影響訓練的過程

問題類型	最後一層激勵函數	損失函數

問題類型	最後一層激勵函數	損失函數
二分類問題	sigmoid	binary_crossentropy
多分類、單標籤問題	softmax	categorical_crossentropy
多分類、多標籤問題	sigmoid	binary_crossentropy
回歸到任一值	無	MSE/RMSE
回歸到 0~1 範圍內的值	sigmoid	MSE 或 binary_crossentropy

3. To Learn More

數學證明：

[http://speech.ee.ntu.edu.tw/~tlkagk/courses/MLDS_2015_2/Lecture/Deep_More_\(v2\).ecm.mp4/index.html](http://speech.ee.ntu.edu.tw/~tlkagk/courses/MLDS_2015_2/Lecture/Deep_More_(v2).ecm.mp4/index.html)



<https://youtu.be/fZAZUYeIMg>
(in Mandarin)



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